The Advanced Placement Teacher Training Incentive Program (APTIP): Estimating the Impact of an Incentive and Training Program on AP Taking and Passing

Deborah J. Holtzman
American Institutes for Research

SUBMITTED FOR REVIEW TO
EDUCATIONAL EVALUATION AND POLICY ANALYSIS
MAY 2010

Abstract

This study examines the effects of a program whose goal is to increase student access to and success in AP courses in mathematics, science, and English. Sixty-four schools across six states who implemented the program in 2008-09 were matched with 128 other schools that closely resembled the program schools on the outcomes of interest in the three prior years. Results of a comparative interrupted time series analysis suggest that in its first year, the program increased schools' percentages of students taking AP exams. Implementation of the program was also associated with increases in schools' percentages of students scoring 3 or higher on the exams, possibly as a function of the increases in the percentages of students taking the exams.

Keywords: Advanced Placement, access, achievement, incentives

Introduction

An increasingly prominent goal in the education arena is to increase students' college readiness, enrollment, persistence, and graduation. Early in his administration, President Obama set as a goal for the U.S. to have the highest proportion of college graduates in the world by 2020 (White House, 2009).

The Advanced Placement (AP) program is one of the most widespread and well-known avenues by which high school students can be exposed to rigorous, college-preparatory coursework. In 2008-09, more than 30 different AP courses were being offered in high schools around the nation, in subjects as diverse as studio art, human geography, and chemistry. The culmination of each course is a college-level examination developed and scored on a five-point scale by higher education faculty and experienced AP teachers, with a score of 3 typically considered to be the cutoff for passing.

Research has indicated that AP exam scores of 3 and higher are predictive of college success and graduation. Dougherty, Mellor, & Jian (2006) found that "the percent of a school's students who take and pass AP exams is the best AP-related indicator of whether the school is preparing increasing percentages of its students to graduate from college" (p. 2). Geiser and Santelics (2004), similarly, found that AP exam scores are among the best predictors of college performance.

Some research suggests that taking AP courses is worthwhile even for students who score lower than 3 on the exams. Gonzalez, O'Connor, and Miles (2001) found that AP Calculus and Physics students scoring 1s or 2s displayed greater content mastery of advanced mathematics and physics than did non-AP students and even performed comparably to students from top-performing nations in those subject areas. More recently, Hargrove, Godin, and Dodd (2008) found that students who took AP courses and exams had higher four-year college graduation rates than did other students as long as they averaged 2 or above on the AP exams.

Begun in the 1950s, the AP program has grown steadily. Even just between 2004 and 2009, the percentage of high school graduates who took an AP exam at some point during high school grew from 19.9 percent to 26.5 percent (College Board, 2010). However, the composition of participants does not mirror that of the nation's student population. For instance, black students constituted 14.5 percent of the public school class of 2009, but only 8.2 percent of the AP examinee population (College Board, 2010). Although this gap is smaller than it used to be, concerns about such disparities have prompted a movement to increase the participation of members of traditionally underrepresented group in AP courses and exams, as well as to improve access to AP more generally.

One initiative that is part of this movement is the Advanced Placement Training and Incentive Program (APTIP) developed by the National Mathematics and Science Initiative (NMSI). Originated in Texas earlier in the decade, APTIP has recently expanded. The program was implemented in the 2008-09 school year in 67 schools in six states: Alabama, Arkansas, Connecticut, Kentucky, Massachusetts, and Virginia. Another 75 schools in these same six states joined the program in the 2009-10 year, and NMSI plans to continue adding schools into the future.

APTIP is designed to increase the participation rates and performance levels of students, particularly underserved students, taking AP courses in mathematics, science, and English. The program provides cash incentives for teachers and students that are based on the number of passing scores students receive on AP exams in the three subject areas. The cash incentives for students are $100 for each passing exam and support for exam fees; the cash incentives for teachers can be as high as several thousand dollars per AP course, with the exact amount dependent on the number of students who pass each exam.

The program also provides instructional support for teachers and students. Teachers participate in content-specific, summer workshops as well as school-based professional development (PD) meetings with lead teachers and other teachers in their AP content areas. The summer trainings focus on deepening teachers' content knowledge, while the school-based PD activities focus on improving teachers' instructional practices in a collaborative setting. Students participate in tutoring and study sessions outside of the regular school day. The study sessions typically take place in the spring on three Saturdays for each AP course, while regular after-school tutoring takes place throughout the school year.

In addition to providing financial incentives and instructional support, NMSI plays an active role in monitoring the implementation of the program. Schools must meet specific implementation milestones at fixed points during the school year in order to receive funding from NMSI and to continue to participate in APTIP. The implementation milestones are organized into five categories, which NMSI calls the program's "elements of success": content-focused teacher training, master teacher mentoring and vertical teaming, student mentoring and increased time on task, teacher and student incentives, and open enrollment.

Two studies, both conducted by Jackson (2007, 2010), have examined the impact of the earlier, Texas version of the APTIP program.[1] The first study found positive effects on AP course enrollment, SAT/ACT scores, and college matriculation for students in schools participating in the program (Jackson, 2007). The second study focused on longer term outcomes of the program and found positive effects on college matriculation, college GPAs and college persistence (Jackson, 2010). The second study also found that the number of AP exams passed increased after the first year of the program and in subsequent years. Both quasi-experimental studies utilized a differences-in-differences (DID) regression approach and included matched comparison schools that wanted to implement the program.[2]

This study focuses on the current incarnation of APTIP and its effects on the school-level percentages of students taking and scoring 3 or higher on AP exams.

Research Questions

The research questions for this study were as follows:

1) To what extent was schools' implementation of APTIP associated with increased percentages of high school students *taking* AP exams, in comparison both with pre-implementation years and with similar schools who did not adopt the program?

2) To what extent was schools' implementation of APTIP associated with increased percentages of students *scoring 3 or higher* on AP exams, in comparison both with pre-implementation years and with similar schools who did not adopt the program?

Because APTIP targets mathematics, science, and English, I examine percentages (a) taking and (b) scoring 3 or higher on each of the following:

- Any AP mathematics exam[3]
- Any AP science exam[4]
- Any AP English exam[5]
- Any AP exam in mathematics, science, or English
- Any AP exam in any subject area

Accordingly, there are 10 outcomes of interest: percentages taking and percentages scoring 3 or higher in each of these 5 subject areas/combinations.

I use a regression-based method to select the comparison schools, and then analyze impact using a comparative interrupted time series design. These will be described in greater depth in the sections that follow.

Method

Data

The outcomes of interest are percentages of students at schools. These percentages were calculated from two separate sources of data.

The numerators for the percentages are the counts of students at each school taking (or scoring 3 or higher on) AP exams. These counts were calculated from data provided by to AIR by NMSI; these data consisted of student-level AP exam data for *every school* with any AP data, including non-APTIP schools, in each of the six states in each year from 2006 through 2009.[6] I aggregated these data to the school level and calculated each school's count of students on each of the 10 outcomes of interest in each of the four years.

The denominators for the percentages are the numbers of 11^{th} and 12^{th} grade students enrolled at each school.[7] For the 2005-06, 2006-07, and 2007-08 years, these enrollment data were obtained from the Common Core of Data (CCD), again on every school in each of the six states. For the 2008-09 year, which was not yet available from CCD at the time of data collection, I obtained the enrollment data from the websites of each of the six states' department of education. I also obtained from CCD some additional data about the schools, including school type, setting, years open, grade span, and demographic composition.

At the time of the data collection and analysis (February 2010), the 2009-10 school year was still underway. Thus, no data were available for 2009-10. As a result, I was unable to include the second cohort of APTIP schools in the analysis. Only the 67 schools in the 2008-09 year are included, and only one year of post-implementation outcome data (along with three pre-implementation years) could be examined. In other words, I examine the effects of the program in its first year of implementation only.

Characteristics of the 2008-09 APTIP Cohort Schools

I began by examining the APTIP schools on some very basic descriptive characteristics available in the CCD data. In particular, I observed that:

- All were classified as "regular" schools (not special education, vocational, or alternative schools)
- None were charter schools
- All except one had been open since 2005-06
- All except one had a 12th grade in every year since 2005-06

The one "exception" school, which opened in 2006-07, did not have an 11th grade until 2008-09. This was the year of APTIP implementation, so this school effectively had no pre-implementation data on AP taking and passing. Because of this, I excluded this school from the analysis, leaving 66 APTIP schools.

Selection of the Comparison Group

Because the APTIP schools were not selected randomly, I wanted to identify a set of comparison schools that were similar to the APTIP schools on observable characteristics, most importantly on the outcomes of interest (AP taking and passing rates) in the pre-implementation period.

The first step, however, was to filter out of the potential comparison pool any schools that did not match the APTIP schools on the basic characteristics noted above (regular, noncharter schools open since 2005-06 with a 12^{th} grade in every year). I also removed the 2009-10 APTIP implementers from the potential comparison pool so as to allow for the possibility of continued analysis in future years. Within each state, I also made a few further restrictions on grade span to match the APTIP schools in that state.[8] Table 1 shows the numbers of APTIP schools and potential comparison schools in each state, following these restrictions.

Table 1 here

Although every school had denominator (enrollment) data, there were 95 schools that lacked numerator (AP) data. I assumed that these schools did not participate in AP at all and filled in 0s as the numerators, yielding percentages of 0.[9]

Having thus identified the pool of potential comparison schools, the next step was to select particular schools from the pool.

One common method for identifying and selecting comparison schools is to determine schools' propensity for adopting the treatment, which involves modeling whether schools did or did not adopt the program as a function of a variety of school characteristics, and then finding non-treatment schools that were similar to the treatment schools in their probability (or propensity) of adopting the treatment, even though they did not actually adopt it. However, to be robust, this method requires a fairly large number of treatment schools. Because of the desirability of matching treatment schools with comparison schools within their own state, I ruled out the propensity score approach as a suitable matching method.

Instead, I used a different regression-based approach that takes advantage of the longitudinal nature of the design, specifically, the availability of multiple years of pre-implementation data. This is a two-stage method. In the first stage, the outcome for the year immediately *preceding* implementation of a program is regressed on the outcome for the two previous years, controlling for school characteristics such as enrollment size, percent minority, and urbanicity. The schools included in the regression include schools that later adopted the program and all other schools in the potential comparison group; no distinctions are made at this point between the treatment and non-treatment schools. Regression equations of the following form are estimated using ordinary least squares:

$$(1) \quad Y_{t=2008} = \beta_0 + \beta_1 Y_{t=2007} + \beta_2 Y_{t=2006} + \beta_3 Urban + \beta_4 TownRural + \beta_5 African\ American_{t=2007} + \beta_6 Hispanic_{t=2007} + \beta_7 Enrollment_{t=2007} + \varepsilon_{t=2008}$$

where

$Y_{t=2008}$ = outcome in year 2008 (spring)
$Y_{t=2007}$ = outcome in year 2007 (spring)

$Y_{t=2006}$ = outcome in year 2006 (spring)

$Urban$ = dummy variable equal to 1 if school is classified as being in a city

$TownRural$ = dummy variable equal to 1 if school is classified as being in a town or rural area

$AfricanAmerican_{t=2007}$ = percentage of school enrollment that was African American in year 2007

$Hispanic_{t=2007}$ = percentage of school enrollment that was Hispanic in year 2007

$Enrollment_{t=2007}$ = number of grades 11 & 12 students enrolled in year 2007

$\varepsilon_{t=2008}$ = random error term.

In the second stage, the parameters (intercept and β coefficients) established in stage 1 are used to predict the outcome in the first year of implementation (2009), using the outcomes for the two previous years (2008 and 2007) and the control variables. Note that schools' actual 2009 outcomes are not used. Again, all treatment and all potential comparison schools are included, with no distinctions between the two groups.

$$(2)\ \hat{Y}_{t=2009} = \hat{\beta}_0 + \hat{\beta}_1 Y_{t=2008} + \hat{\beta}_2 Y_{t=2007} + \hat{\beta}_3 Urban + \hat{\beta}_4 TownRural + \hat{\beta}_5 African\ American_{t=2008} + \hat{\beta}_6 Hispanic_{t=2008} + \hat{\beta}_7 Enrollment_{t=2008}$$

We then examine the predicted outcome values to identify comparison schools that were *predicted* to have performed similarly to how APTIP schools were *predicted* to have performed in the first year of implementation. In other words, we use the equation described in (2) to predict implementation-year outcome values for each school, and then match to each APTIP school the schools that have the closest predicted outcome in the first post-implementation year. The logic behind this process is that we want our comparison schools for each APTIP school to have implementation-year outcomes similar to the outcomes APTIP schools would have had if they had not adopted APTIP.

Specifically, we rank all of the schools on their predicted values, locate each APTIP school, and then select its nearest-above and nearest-below neighbors. Selecting two comparison schools per APTIP school provides more power for the statistical analysis, and the selection of the nearest-above and nearest-below neighbors (as opposed to, say, the two nearest neighbors, regardless of whether they are above or below) helps enhance the balance, or similarity, between the comparison and APTIP schools on the pre-implementation outcomes.[10] If the lowest- or highest-ranked school is an APTIP school, then we conclude that there is not a suitable pair of matches for that school, and accordingly we omit it from all remaining analyses.

In my application of this approach, a complication was caused by the fact that I had multiple different outcome measures I wanted to analyze, but I did not want the APTIP schools to be matched with a different set of comparison schools for each different outcome; rather, I wanted to have a set of comparison schools that would remain constant across all of the various analyses. I resolved this issue by combining the 10 outcome measures for each year into a single composite measure. I did this by standardizing each of the measures (within year, within state), and then, for each school, averaging together the standardized values.[11] These composite indexes were highly reliable, with alphas ranging from .93 to .99 depending on the year and the state. (Of the 3 years x 6 states = 18 alphas, all were .95 or above except the three for Arkansas.)

I then proceeded to employ the regression approach, as described above. The regressions and selection were done separately within each of the six states. The stage-1 regression results are provided in Appendix A. In each of the six regressions, the one-year prior composite index score (2007) was far and away the most powerful predictor, although the two-years prior score (2006) was also highly significant for four of the six states.

Results of the Match

In only one state did the issue arise that suitable matches for APTIP schools could not be found: in Alabama, the two schools with the highest predicted index values were both APTIP schools.[12] Thus, these schools were discarded, leaving a total sample of 64 APTIP schools and 128 comparison schools.

Table 2 presents the average percentages for the two groups of schools on each of the 10 outcome measures in each of the three pre-implementation years. The means were very similar; none of the differences were significant at p<.05, and only one was significant at p<.10.

Table 2 here

I also compared the two groups of schools on school setting, enrollment size, and demographics. (See Tables 3 and 4.) There were no significant differences between the APTIP schools and the comparison schools on percent of schools located in a city, on percent of schools located in a town or rural area, or on enrollment size. However, there were significant differences on demographic composition, with the APTIP schools averaging higher percentages of black and Hispanic students. Although this was not ideal, adjustments to make the two groups more equivalent on composition would have been at the expense of similarity of the outcome measures, which I deemed more important.

Tables 3 and 4 here

With the comparison group selected, I then proceeded to conduct the actual impact analysis.

Impact Analysis Method

The method used for the analysis was a comparative interrupted time series (CITS) design. In this method, treatment schools are analyzed in relation to their own performance prior to implementation and also in relation to comparison schools that never implemented the program. The design makes use of four years of data, from 2006 through 2009: the first three years preceded the implementation of APTIP (the "interruption") for the initial cohort of schools, while 2009, the final year, was a post-implementation year. The use of the comparison group helps guard against the possibility that some other new program or change in the state context is the real explanation behind any observed improvement.

The analysis examines patterns in outcomes before and after implementation to determine increases in percentages for treatment schools, and over the same time period for the comparison schools. It then calculates the difference between the schools in their growth; this difference is the estimate of the benefit of adopting APTIP after one year.

Using as an example an analysis examining the percentage of students taking any AP exam in mathematics, the equation for the model is as follows:

$$Percentage_{st} = \pi_0 + \pi_1 Yr2007_{st} + \pi_2 Yr2008_{st} + \pi_3 Yr2009_{st} + \pi_4 APTIP_{st} + \theta_s + \upsilon_{st}$$

where:

$Percentage_{st}$ = percentage of students taking any mathematics AP exam for school *s* in year *t*.

$Yr2007_{st}$ = dummy variable equal to 1 for year 2007 (spring), 0 otherwise. $Yr2008_{st}$ and $Yr2009_{st}$ are defined similarly. The reference year in this formulation is 2006.

$APTIP_{st}$ = dummy variable equal to 1 if school *s* is an APTIP school in year *t*.

θ_s = fixed effect for school *s*.

υ_s = random error term for school *s* in year *t*, independently and identically distributed across years.

The terms $Yr2007_{st}$ through $Yr2009_{st}$ are fixed effects that represent systematic variation in the percentages by year across schools in the sample.

The key term in the model is the indicator variable for $APTIP_{st}$. The coefficient π_4 provides an estimate of whether the APTIP schools had higher percentages than would be expected (based on their prior percentages and on the percentages in the comparison schools) in the initial year of implementation of APTIP.

I conducted these regressions for each of the 10 outcomes using Stata 10's xtreg, fe command, specifying the option for robust standard errors.

In addition to running the regressions, I also constructed time-series graphs for each outcome. These provide a clear picture of how the APTIP schools are performing over time and in relation to the comparison schools, based on their actual (non-adjusted) percentages, averaged across schools.

Results

<u>Percentages of Students Taking AP Exams</u>

The fixed-effects regressions show that in all five of the subject areas/combinations, implementation of APTIP was associated with large and statistically significant increases in the percentages of students *taking* the AP exams. (See Table 5.) Perhaps most notably, implementation of APTIP was associated with a 12-point increase in the percentage of students taking at least one mathematics, science, or English AP exam. In each of the three pre-implementation years, the standard deviation on this percentage (across all 192 APTIP and non-APTIP schools) was about 10 percentage points, so a 12-point increase is more than a full-standard deviation increase.

The positive and often-significant coefficients on the year variables indicate that across all of the schools (non-APTIP as well as APTIP), there was a general upward trend in the percentages of students taking the AP exams.

One interesting result is that the APTIP effect on percentage taking at least one math, science, or English exam (12.39) is greater than the APTIP effect on the percentage taking at least one AP exam in any subject area (11.38). This might suggest that the program, while increasing the probability that students taking math, science, or English AP exams, slightly decreases the probability that students take exams in other subject areas. Since the incentives only apply to math, science, and English courses, this is perhaps not surprising. Presumably, students feel they can only take so many AP courses or exams.

Table 5 here

Because of the inclusion of the school fixed effects, the regressions help factor out individual school-level influences on the outcomes, but even based on the actual, unadjusted percentages, the effect of APTIP implementation is evident from the graphs in Figure 1. These graphs show the similarity between the APTIP and comparison schools in the three-year pre-period, and then the sharp uptick for the APTIP schools following implementation, represented by the vertical line between 2008 and 2009. Figure 1e does, however, suggest that the APTIP schools already had slightly higher percentages of students taking *any* AP exam than did the comparison schools in the year immediately prior to implementation. This could have been a planning-year effect, as the schools already knew that they would be adopting the program in fall 2008 by the time the AP exams were administered in spring 2008.

Figure 1 here

Percentages of Students Scoring 3 or Higher on AP Exams

For examining the impact of APTIP on the percentages of students scoring 3 or higher on AP exams, I conducted two sets of analyses. Both include the APTIP variable, the year effects, and the school fixed effects, as per the equation presented above; the second set also adds in a control for the percentage *taking* the AP exam.

The results for the first set are presented in Table 6. Again there are strongly significant positive results on the APTIP variable, indicating that implementation of APTIP is associated with increases in the percentages of students scoring well on the AP exams.[13] These increases are also reflected in the actual percentages, as shown in the graphs of Figure 2.

Table 6 here

Figure 2 here

Adding in a control for the percentage of students taking the exams allows us to get a sense of the extent to which the increases in percentages of students scoring well on the exam may be a function of the increases in percentages of students *taking* the exams. For instance, imagine a school with 500 eleventh and twelfth grade students in both 2008 and 2009. Suppose that in 2008, 100 of the 500 students took an AP exam, of whom 40 scored 3 or higher; in 2009, suppose that 150 of the 500 students took an AP exam, of whom 50 scored 3 or higher. It seems reasonable to think that the increase in the number of students scoring 3 or higher, from 40 to 50, may have been due to the larger increase in the number of students taking the exam, from 100 to 150. Controlling for the percentage of students taking the exam helps us see how, or whether, the percentage scoring 3 or higher would change if the percentage of students taking were held constant. The results of this analysis are presented in Table 7.

Indeed, the percentage of students taking the exams is positively, significantly associated with the percentage of students scoring 3 or higher on the exams, and when this is taken into account, the APTIP effect is much smaller than it was in Table 6. In fact, it is effectively zero for mathematics and science; it is still positive, but not significantly so, for English, for any math/science/English, and for any AP exam.

By no means does this suggest that APTIP is not working—it merely suggests that the major mechanism by which the program helps increase the percentages of students passing AP exams is its success in increasing the percentages of students taking AP exams. Put simply, more students taking means more students passing. Given the large increases in students taking the exams demonstrated earlier, it might in fact not be completely surprising to see actual *decreases* in percentages of students passing; this could happen if, for instance, teachers of AP courses are not prepared to handle large influxes of students into their courses and instruction suffers, possibly at the expense of students who formerly would have performed well on the exams. Yet this does not seem to happen here. Moreover, even if the "new takers" (that is, students who would not have taken the AP courses or exams in the absence of the program) do not pass the exams at the same rates as the "unconditional takers" (that is, students who would have taken the courses or exams regardless of the program), their exposure to the AP course material could well be seen as a positive effect in and of itself.

Conclusion

In summary, APTIP appears to be having a large positive effect on the percentages of students taking AP exams, particularly in mathematics, science, and English within the first year of implementation. There are also increases in the percentages of students scoring 3 or higher on the exams, although to a large extent, these are function of the increases in the percentages taking the exams. Even so, the program appears to be quite successful.

Further research would, of course, be valuable. Examination of effects beyond the first year of implementation will be critical to see whether the program can sustain its positive effects. Will the percentages of students taking AP exams continue to rise? If so, what will happen to the percentages scoring 3 or higher? There may be a "saturation point" beyond which it will be difficult to increase the percentages of students scoring 3 or higher, even if the percentages taking continue to rise. There may even be a "tipping point" where the percentages scoring 3 or higher could begin to decline if, say, AP courses start enrolling more students than can be supported given available resources or if course quality slips.

Incorporating the second cohort of APTIP schools would also be of benefit. The two cohorts could be combined into a single statistical analysis that estimates first-year effects, second-year effects, and so on into the future.

Another line of research might be to examine effects on student subgroups, such as traditionally underrepresented minority students or female students. Such an analysis would be similar to the one described here, except that the taking/passing counts and the enrollment totals would include only students in the subgroup under study. However, a different set of comparison schools might be needed, since in the analysis reported here, schools were not matched based on subgroup outcomes and might not be sufficiently similar in the pre-implementation years to warrant a robust CITS analysis.

Finally, different outcomes might be examined. For instance, instead of looking at percentages of students taking at least one AP exam, we might look at percentages taking at least 3 AP exams; instead of looking at percentages of students scoring 3 or higher, we might look at percentages scoring 4 or 5, or maybe even 2 or above. Some non-AP outcomes would also be of interest, such as high school graduation rates, college going rates, and college completion rates (though data on the latter two may be difficult to obtain in most states). One would hope that whatever success APTIP might be having in AP-related outcomes would also be reflected more broadly in other important indicators of student success and predictors of life chances.

Notes

1. The Texas version of APTIP is called the Advanced Placement Incentive Program (APIP). Both APTIP and APIP share the same features—financial incentives for teachers and students, instructional support, “elements of success,” etc.—so the two programs are essentially the same.

2. In order to participate in the program, schools must secure partial funding. The matched comparison schools in Jackson’s studies could not implement the program right away because they lacked funding. This created a staggered roll out of the program among schools that all wanted the program, but only some had the funding to start right away.

3. Calculus AB, Calculus BC, Computer Science A, Computer Science AB, and Statistics

4. Biology, Chemistry, Environmental Science, Physics B, Physics C: Mechanics, and Physics C: Electricity and Magnetism

5. English Literature and English Language

6. Although these data were student-level data, there was no constant identifier across years.

7. I use 11^{th} and 12^{th} grade enrollment instead of total school enrollment because most AP takers are in those grades, and to be comparable with existing analyses that use 11^{th} and 12^{th} grade enrollments as the base.

8. For instance, all 14 of the APTIP schools in Virginia were grades 9-12 schools. I restricted the pool of potential comparison schools similarly.

9. Most (75) of these schools were rural schools, so the assumption that they did not have an AP program seemed reasonable.

10. Because some APTIP schools may share the same neighbors, the comparison schools for each APTIP school are selected serially in a random order.

11. The standardization helps address the issue that as percentages, some of the measures are nested within one another. For instance, percent of students taking any math exam is nested within the percent of students taking any math, science, or English exam; similarly, percent of students scoring 3 or higher on any math exam is nested within the percent taking any math exam. Standardizing puts each measure on a more uniform scale, albeit one that is relative to the performance of all schools in the group.

12. Both of these schools had percentages exceeding 100 on some of the outcomes, even in the years prior to the implementation of APTIP. Percentages exceeding 100 occur because of 10^{th} graders taking AP exams; recall that the denominator consists of 11^{th} and 12^{th} grade enrollment.

13. Recall that the denominator on these percentages is the total school enrollment in grades 11 & 12. It is *not* the number of AP exam takers. An alternate approach might be to look at passing rates among students taking the exams (i.e., the numbers of students scoring 3 or higher divided

by the number of students taking the exam). However, schools that have no one taking the exam would have to be omitted from such an analysis, at least for the year(s) in which that was the case. As a result, such an analysis would be biased against APTIP because many of the participating schools are starting their AP programs from scratch and would be excluded because they lack pre-implementation data. Thus, I elected not to take this approach.

References

College Board. (2010). The 6th annual AP report to the nation. Retrieved 4/15/10 from http://www.collegeboard.com/html/aprtn/pdf/ap_report_to_the_nation.pdf.

Dougherty, C., Mellor, L., & Jian, S. (2006). The relationship between Advanced Placement and college graduation." Austin, TX: National Center for Educational Accountability. Retrieved 4/15/10 from http://www.nc4ea.org/files/relationship_between_ap_and_college_graduation_02-09-06.pdf.

Hargrove, L., Godin, D., & Dodd, B. (2008). College outcomes comparisons by AP and non-AP high school experiences. Research Report No. 2008-3. New York: The College Board.

Jackson, C. K. (2007). A little now for a lot later: An look at a Texas Advanced Placement incentive program. Retrieved 3/30/2010 from www.ilr.cornell.edu/cheri/workingPapers/upload/cheri_wp107.pdf.

Jackson, C. K. (2010). A stitch in time: The effects of a novel incentive-based high-school intervention on college outcomes. National Bureau of Economic Research Working Paper 15722. Retrieved3/30/10 from http://www.nber.org/papers/w15722.

Geiser, S., & Santelices, V. (2004). *The role of Advanced Placement and honors courses in college admissions.* Berkeley, CA: Center for Studies in Higher Education, University of California, Berkeley.

Gonzalez, E., O'Connor, K., & Miles, J. (2000). How well do Advanced Placement students perform on the TIMSS Advanced Mathematics and Physics Tests? Chestnut Hill, MA: The International Study Center, Lynch School of Education, Boston College.

White House. (2009). Remarks of President Barack Obama – As Prepared for Delivery Address to Joint Session of Congress**,** Tuesday, February 24th, 2009. Retrieved 4/15/10 from http://www.whitehouse.gov/the_press_office/remarks-of-president-barack-obama-address-to-joint-session-of-congress/.

Tables

Table 1: Number of APTIP schools and potential comparison schools, by state

State	APTIP	Non-APTIP	Total
Alabama	12	221	233
Arkansas	10	126	136
Connecticut	9	117	126
Kentucky	12	141	153
Massachusetts	9	229	238
Virginia	14	235	249
Total	66	1069	1135

Table 2: Outcome Measures in Pre-Implementation Years, APTIP and non-APTIP schools, Means (Standard deviations), and Mean Differences

	2008			2007			2006		
Percentage of students…	Non-APTIP	APTIP	Diff.	Non-APTIP	APTIP	Diff.	Non-APTIP	APTIP	Diff.
Taking any AP exam	19.47 (13.57)	22.37 (16.23)	-2.90	18.18 (13.85)	18.81 (16.04)	-0.64	17.50 (13.06)	18.12 (17.43)	-0.63
Taking any math, science, or English AP exam	12.95 (9.73)	13.91 (11.12)	-0.95	12.02 (9.96)	11.60 (10.77)	0.41	11.43 (9.04)	11.07 (10.60)	0.36
Taking any AP math exam	4.25 (4.32)	3.96 (4.22)	0.29	4.20 (4.25)	3.74 (4.16)	0.46	4.01 (3.97)	3.47 (3.69)	0.54
Taking any AP science exam	4.18 (4.63)	4.74 (5.20)	-0.56	4.19 (4.97)	4.18 (4.62)	0.01	3.66 (4.49)	3.98 (5.18)	-0.31
Taking any AP English exam	8.60 (8.49)	9.65 (9.92)	-1.06	7.66 (8.53)	7.51 (8.69)	0.16	7.38 (7.79)	7.38 (8.81)	0.01
Scoring ≥3 on any AP exam	9.60 (9.07)	9.21 (8.20)	0.39	9.21 (8.99)	8.89 (8.65)	0.32	8.92 (8.42)	8.35 (8.08)	0.57
Scoring ≥3 on any math, science, or English AP exam	6.46 (6.15)	6.07 (6.22)	0.40	5.91 (6.03)	5.47 (5.89)	0.44	5.71 (5.59)	5.03 (5.37)	0.68
Scoring ≥3 on any AP math exam	2.27 (3.12)	1.69 (2.27)	0.58	2.09 (2.88)	1.43 (1.97)	0.66	2.05 (2.76)	1.37 (1.72)	0.68^{+}
Scoring ≥3 on any AP science exam	1.74 (3.00)	1.49 (2.13)	0.25	1.96 (3.25)	1.65 (2.64)	0.31	1.77 (2.95)	1.46 (2.41)	0.31
Scoring ≥3 on any AP English exam	4.18 (4.25)	4.43 (5.04)	-0.25	3.60 (4.01)	3.75 (4.50)	-0.15	3.45 (3.67)	3.43 (4.24)	0.02

$^{+}p<.10$; $^{*}p<.05$; $^{**}p<.01$; $^{***}p<.001$
Non-APTIP N=128; APTIP N=64

Table 3: School Setting, APTIP and non-APTIP schools, Means (Standard deviations), and Mean Differences

Percentage of schools…	Non-APTIP	APTIP	Diff
In a city	16.41	21.88	-5.47
	(37.18)	(41.67)	
In a town or rural area	53.91	56.25	-2.34
	(50.04)	(50.00)	

+p<.10; *p<.05; **p<.01; ***p<.001
Non-APTIP N=128; APTIP N=64

Table 4: School Characteristics, APTIP and non-APTIP schools, Means (Standard deviations), and Mean Differences

	2008			2007			2006		
	Non-APTIP	APTIP	Diff	Non-APTIP	APTIP	Diff	Non-APTIP	APTIP	Diff
Enrollment in grades 11 & 12	479.56	539.28	-59.72	466.83	535.59	-68.77	455.41	516.09	-60.69
	(292.00)	(269.93)		(283.88)	(274.79)		(276.63)	(279.77)	
Percent black	16.10	26.58	-10.48**	15.96	25.97	-10.00**	15.73	25.09	-9.36**
	(23.29)	(25.59)		(23.30)	(24.96)		(23.18)	(23.91)	
Percent Hispanic	5.27	10.47	-5.21**	4.94	9.97	-5.03**	4.52	9.29	-4.77**
	(7.99)	(17.54)		(7.79)	(16.82)		(7.40)	(15.95)	

+p<.10; *p<.05; **p<.01; ***p<.001
Non-APTIP N=128; APTIP N=64

Table 5: Percentages Taking AP Exams, Coefficients (Robust SE)

Variable	Math	Science	English	Any MSE	Any AP
Yr2007	0.22	0.42*	0.23	0.57⁺	0.69⁺
	(0.15)	(0.19)	(0.27)	(0.29)	(0.36)
Yr2008	0.33	0.60**	1.57***	1.96***	2.73***
	(0.21)	(0.22)	(0.32)	(0.37)	(0.50)
Yr2009	0.49*	0.48⁺	1.57***	1.81***	2.81***
	(0.25)	(0.28)	(0.44)	(0.48)	(0.59)
APTIP	2.83***	5.88***	9.28***	12.39***	11.38***
	(0.42)	(0.97)	(1.42)	(1.72)	(1.64)
Constant	3.83***	3.77***	7.38***	11.31***	17.71***
	(0.12)	(0.16)	(0.23)	(0.27)	(0.31)
sigma_u	3.85	4.32	8.50	9.77	14.19
sigma_e	1.95	2.91	4.28	5.03	5.57
rho	0.80	0.69	0.80	0.79	0.87

⁺p<.10; *p<.05; **p<.01; ***p<.001

Table 6: Percentages Scoring 3 or Higher on AP Exams, Coefficients (Robust SE)

Variable	Math	Science	English	Any MSE	Any AP
Yr2007	0.05	0.19*	0.21	0.29⁺	0.37*
	(0.08)	(0.08)	(0.13)	(0.15)	(0.18)
Yr2008	0.26*	-0.01	0.81***	0.85***	0.74**
	(0.10)	(0.09)	(0.17)	(0.19)	(0.23)
Yr2009	0.31*	-0.05	1.03***	1.11***	1.37***
	(0.15)	(0.12)	(0.22)	(0.24)	(0.30)
APTIP	0.72**	0.91***	2.44***	3.01***	2.82***
	(0.21)	(0.25)	(0.40)	(0.43)	(0.51)
Constant	1.82***	1.67***	3.45***	5.48***	8.73***
	(0.07)	(0.06)	(0.11)	(0.13)	(0.15)
sigma_u	2.62	2.73	4.15	5.87	8.58
sigma_e	0.97	0.99	1.64	1.79	2.19
rho	0.88	0.88	0.87	0.92	0.94

⁺p<.10; *p<.05; **p<.01; ***p<.001

Table 7: Percentages Scoring 3 or Higher on AP Exams Controlling for Percentages Taking the Exams, Coefficients (Robust SE)

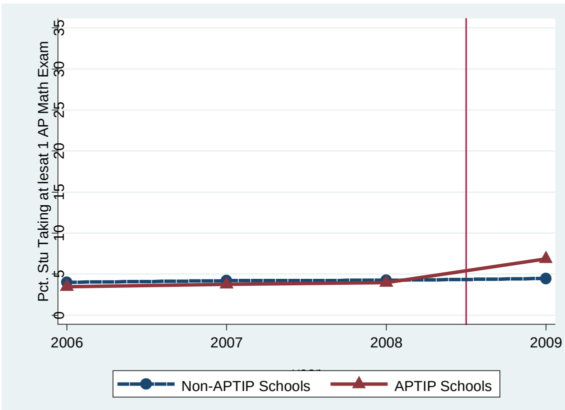
Variable	Math	Science	English	Any MSE	Any AP
Yr2007	-0.01	0.13+	0.15	0.18	0.22
	(0.06)	(0.07)	(0.10)	(0.12)	(0.16)
Yr2008	0.17*	-0.10	0.45***	0.47**	0.14
	(0.08)	(0.08)	(0.13)	(0.16)	(0.17)
Yr2009	0.18	-0.11	0.66***	0.75***	0.75**
	(0.12)	(0.10)	(0.18)	(0.20)	(0.24)
APTIP	-0.05	0.08	0.27	0.58	0.32
	(0.22)	(0.24)	(0.37)	(0.42)	(0.44)
Percent	0.27***	0.14***	0.23***	0.20***	0.22***
Taking	(0.05)	(0.03)	(0.04)	(0.04)	(0.03)
Constant	0.78***	1.13***	1.72***	3.26***	4.84***
	(0.20)	(0.15)	(0.36)	(0.48)	(0.52)
sigma_u	1.89	2.25	2.97	4.59	6.42
sigma_e	0.81	0.90	1.30	1.49	1.81
rho	0.84	0.86	0.84	0.90	0.93

+p<.10; *p<.05; **p<.01; ***p<.001

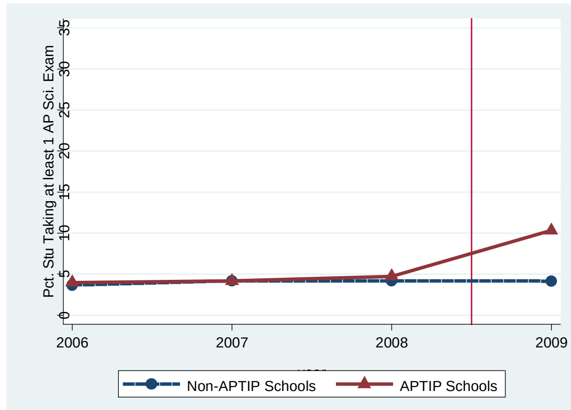
Figures

Figure 1: Percentages Taking AP Exams, APTIP Average (n=64) vs. non-APTIP Average (n=128)

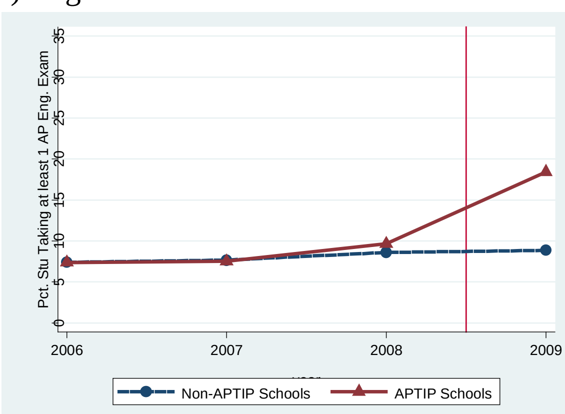
a) Math



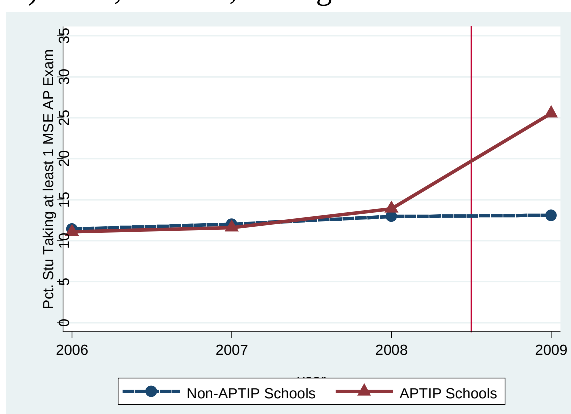
b) Science



c) English



d) Math, Science, or English



e) Any AP Subject

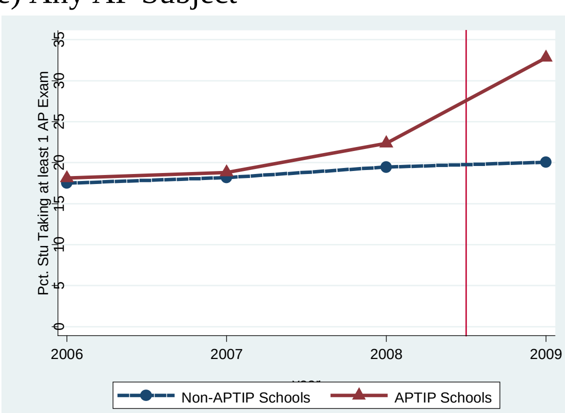
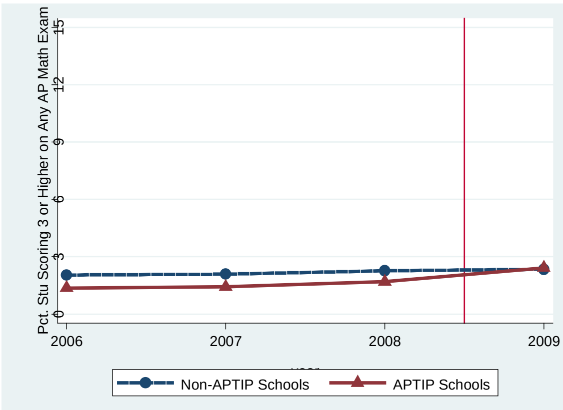
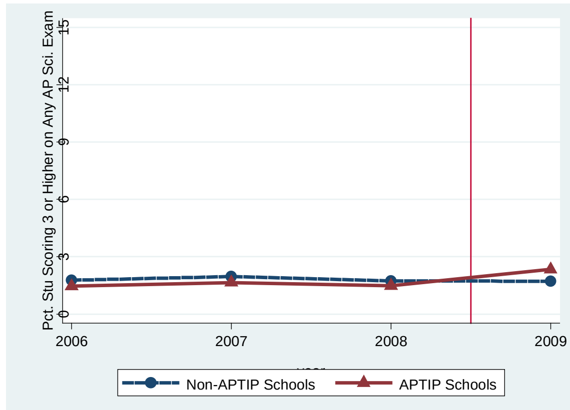


Figure 2: Percentages Scoring 3 or Higher on AP Exams, APTIP Average (n=64) vs. non-APTIP Average (n=128)

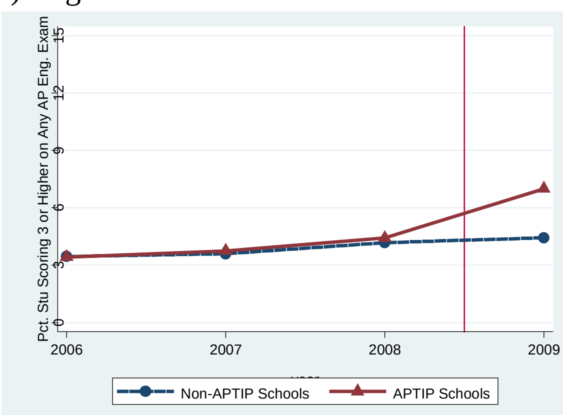
a) Math



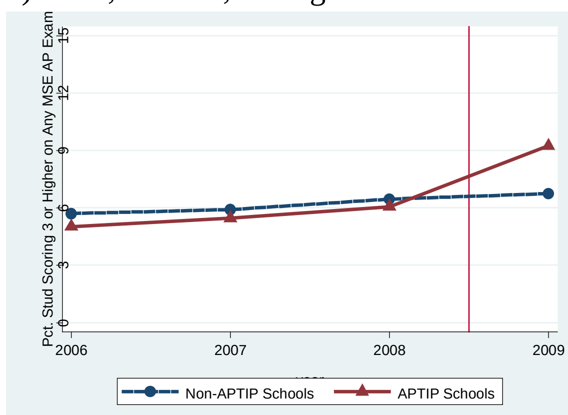
b) Science



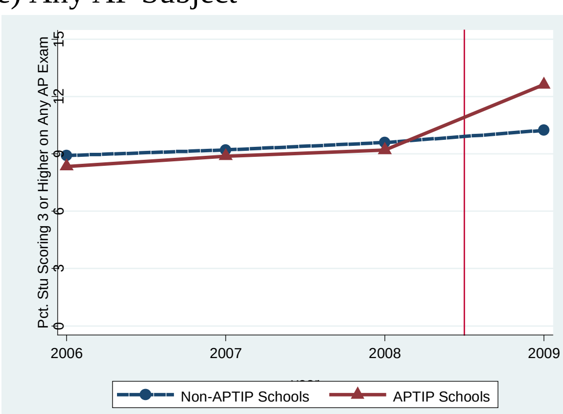
c) English



d) Math, Science, or English



e) Any AP Subject



Appendix A: Comparison Group Selection, Stage 1 Regression Results

Table A1: Comparison Group Selection Stage 1 Regression Results (Prediction of 2008 Composite Index Value), Coefficients (SE)

Variable	Alabama	Arkansas	Connecticut	Kentucky	Massachusetts	Virginia
2007 composite index	0.917***	0.831***	0.735***	0.600***	0.723***	0.749***
	(0.06)	(0.08)	(0.08)	(0.09)	(0.06)	(0.06)
2006 composite index	0.057	0.032	0.234**	0.352***	0.219***	0.214***
	(0.06)	(0.08)	(0.08)	(0.09)	(0.06)	(0.06)
City	-0.048	0.228*	0.158*	-0.067	0.039	0.030
	(0.06)	(0.11)	(0.07)	(0.11)	(0.07)	(0.03)
Town/rural	-0.072	0.005	0.047	-0.042	-0.053	0.021
	(0.05)	(0.10)	(0.05)	(0.08)	(0.04)	(0.03)
2006-07 percentage black	0.000	-0.003**	-0.003+	-0.004	0.001	0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
2006-07 percentage Hispanic	0.001	0.001	-0.001	0.031*	-0.002	-0.001
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
2006-07 enrollment in grades 11 and 12	-0.000+	0.000	0.000	0.000	0.000	0.000***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.099+	0.002	0.004	0.008	0.041	-0.117**
	(0.06)	(0.12)	(0.05)	(0.10)	(0.04)	(0.04)
R-squared	0.951	0.873	0.943	0.908	0.913	0.974
R-squared, adjusted	0.950	0.866	0.940	0.904	0.910	0.973
N	233	136	126	153	238	249

+p<.10; *p<.05; **p<.01; ***p<.001