CRESST REPORT 847

MEASURING THE CAUSAL EFFECT OF THE NATIONAL MATH + SCIENCE INITIATIVE'S COLLEGE READINESS PROGRAM

AUGUST 2015

Richard S. Brown
Kilchan Choi

National Center for Research
on Evaluation, Standards, & Student Testing
UCLA | Graduate School of Education & Information Studies

Measuring the Causal Effect of the National Math + Science Initiative's College Readiness Program

CRESST Report 847

Richard S. Brown
West Coast Analytics

Kilchan Choi
CRESST/University of California, Los Angeles

August 2015

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Building, Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The findings and opinions expressed in this report are those of the authors.

To cite from this report, please use the following as your APA reference: Brown, R. S., & Choi, K. (2015). *Measuring the causal effect of the National Math + Science Initiative's College Readiness Program* (CRESST Report 847). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

TABLE OF CONTENTS

MEASURING THE CAUSAL EFFECT OF THE NATIONAL MATH + SCIENCE INITIATIVE'S COLLEGE READINESS PROGRAM

Richard S. Brown
West Coast Analytics

Kilchan Choi
CRESST/University of California, Los Angeles

Abstract

This study employs a potential outcomes modeling approach to estimate the causal effect of the National Math + Science Initiative's College Readiness Program on Advanced Placement test taking and qualifying score earning for three recent cohorts of schools. Results indicate substantial and significant increases in both AP test taking and qualifying score earning for all students. In addition, significant effects for AP test taking and qualifying score earning over baseline were found for female students and minority students when analyzed separately. This study provides evidence of the effectiveness of a College Readiness Program that is having a significant and important impact on preparing more students to succeed in math and science careers and improve the future of math and science education in this country.

Introduction

The National Math + Science Initiative (NMSI) was formed in 2007 for the purpose of improving the number of students prepared to enter into math and science careers with two teacher training programs (UTeach expansion and Laying the Foundation) and a college readiness program (formerly called APTIP or APIP). In this study, we will explore the effects of the NMSI's College Readiness Program by using a propensity score weighting, potential outcomes approach that enables us to make causal estimates in an observational, non-experimental study. West and Thoemmes state, "if all important covariates related to both treatment assignment and outcome have been measured and all propensity scores fall within the bounds of $0 < P(T) < 1$, then it is possible to achieve an unbiased estimate of the causal effect of T" (West & Thoemmes, 2010, p. 28).

The goal of the NMSI College Readiness Program (CRP) is to transform schools into centers of college readiness. After reviewing a school's application, NMSI's expert team analyzes all aspects of the school that would impact its students' STEM college readiness. From that analysis, teachers and administrators collaboratively create aggressive performance goals, resulting in an individualized school implementation plan to reach their targets. In order to meet their goals, schools must commit to opening up Advanced Placement to all students. Students who have never been considered or have never been given the opportunity to take more rigorous

classes that lead to AP are encouraged to enroll and given the resources to succeed. During the initial three-year engagement, NMSI increases teacher effectiveness and student achievement through training, teacher and student support, vertical teaming (meetings of middle and high school subject matter teachers for alignment across grades), open enrollment, and nominal monetary awards.

The program begins with NMSI's intensive summer teacher training for AP teachers, as well as non-AP teachers in Grades 3–11 from the high school and its feeder schools who will build the pipeline of students ready for AP courses. Since teacher training has limited effectiveness without additional support, NMSI AP teachers are assigned an expert mentor to provide coaching and assistance throughout the year. NMSI program schools also have access to in-depth, online content for both teachers and students to maximize their success.

The next phase of NMSI's program provides more time on task for students. Students access additional homework help through an online program in which teachers can track progress. Students attend three 6-hour Saturday study sessions taught by a master AP teacher—time that equates to three extra weeks of AP class time. The Saturday study sessions also provide professional development, as local teachers join their students to see how the best in their field help students tackle the most difficult parts of AP courses.

For the three years of NMSI program implementation, teachers continue to receive progressively more rigorous training and lessons; teachers and administrators continue to push further toward increasingly challenging goals; and both students and teachers receive nominal monetary awards for success. NMSI staff work with teachers and administrators throughout implementation to track progress toward their goals and troubleshoot where needed.

At the end of the three years, schools receive NMSI designation and agree to continue to set aggressive goals, provide training for any new teachers in their school, host Saturday study sessions, and report results. At no cost, these NMSI-designated schools have access to all NMSI resources, including pre-AP webinars, online discussion forums, and content upgrades.

A subset of NMSI program schools serve students from military families through a targeted military family initiative. This focus is to ensure that our military families have the best educational options available, that STEM talent near military bases is developed, and that future military recruits are STEM-capable. The initiative has already impacted 110 military-connected schools with plans to expand to a network of 200 military-connected schools.

Figure 1 shows a depiction of the logic model supporting NMSI's College Readiness Program. As discussed, teacher participation in professional development and mentoring, their access to rigorous materials and resources, and the use of incentives are designed to drive

increased knowledge and use of instructional strategies that NMSI considers to be effective, as well as increased content knowledge and increased effectiveness in the classroom. Those intermediate outcomes should then drive longer term outcomes for the school and for the teacher, including increases in AP course enrollment, increases in percentage of qualifying scores on AP exams, and increased number of teachers at the school qualified to teach AP courses.

For students, additional time on task; access to rigorous materials, resources, homework help and tutoring; awards for performance on the AP exam; and exposure to highly trained teachers all are designed to increase student engagement in the classroom, student preparation for the AP exam, and student motivation to perform well on that exam. These intermediate outcomes should further influence the rate of AP course enrollment (i.e., if taken prior to graduation, a student may enroll in additional AP courses) and percentage of qualifying AP scores. Finally, more students obtaining qualifying scores on exams and having positive experiences in these classes should further influence the number of students enrolling in postsecondary STEM courses and declaring STEM majors.

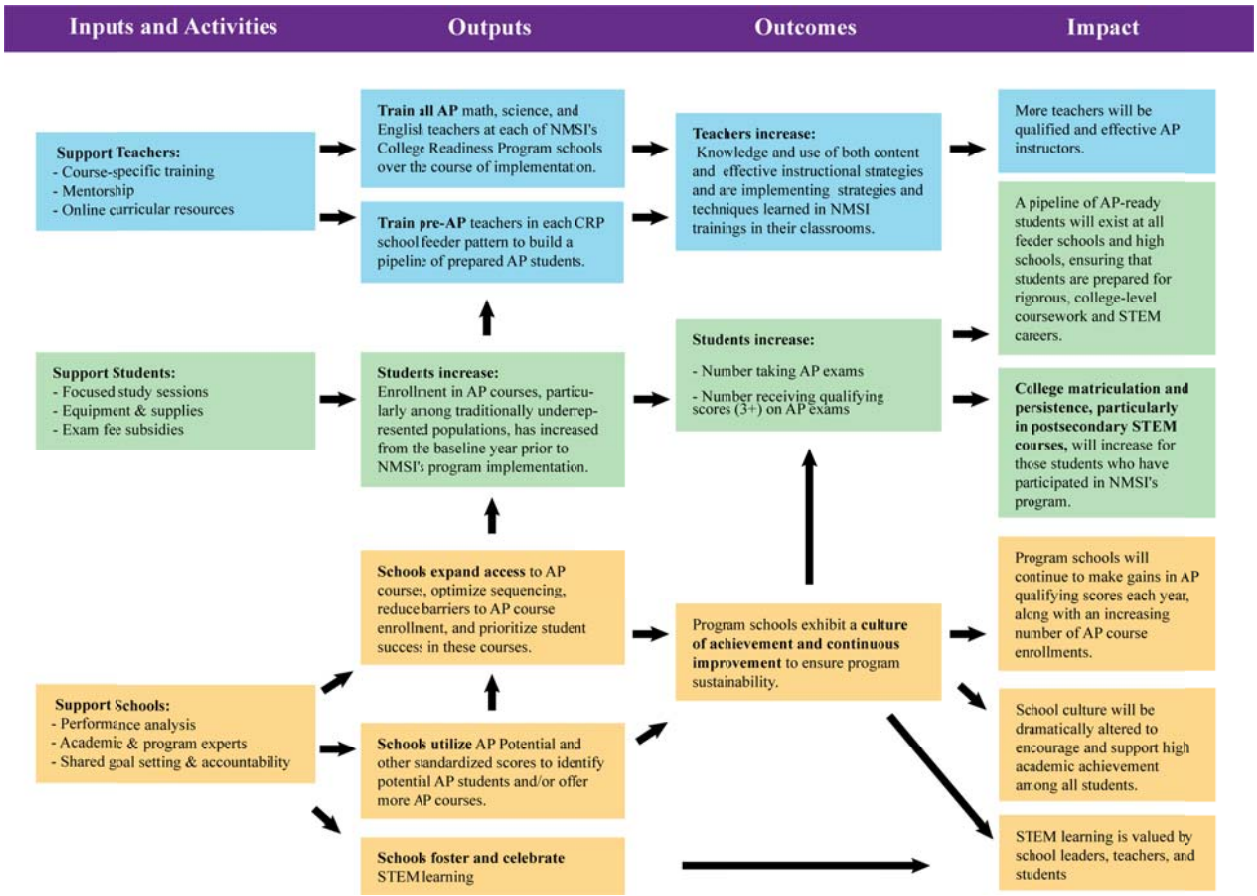


Figure 1. NMSI logic model.

Research has shown the NMSI program to be effective (Holtzman, 2010; Jackson 2010a, 2010b, 2012, 2014; Sherman & Song, 2014, 2015). Holtzman (2010) found that CRP had a positive and statistically significant first year impact on student enrollment in STEM-related AP courses. Likewise significant effects were found on students attaining qualifying scores of 3 or better on STEM-related AP tests. This study employed a Comparative Interrupted Time Series (CITS) design, and found positive effects among the 64 program schools and 128 matched schools. In each analysis investigated, CRP was associated with large and statistically significant increases in the percentages of students taking AP exams. Holtzman reported standardized effect sizes for percentage increase in the likelihood of students taking an AP test in excess of 1.0. Similarly, CRP implementation increased the percentage of students earning qualifying scores, with effect sizes up to 0.5.

Jackson (2010a, 2010b) examined the impact of the NMSI CRP on longer term outcomes in addition to secondary outcomes, such as post-secondary success using Texas data from an earlier incarnation of the program (called APTIP). In these studies, Jackson found positive program effects on AP course enrollment, SAT/ACT scores, and college matriculation (Jackson, 2010a) and on college matriculation, college GPAs, and college persistence (Jackson, 2010b). Jackson (2012, 2014) extended the 2010 studies by investigating the effect of the program on labor-market outcomes, such as wages. Using the same quasi-experimental difference-in-differences (DID) strategy, Jackson found a positive CRP effect on earnings, as well as a significant positive impact on college retention and college grade point average for students in schools implementing the NMSI program.

More recently, Sherman and Song (2014, 2015) analyzed data from two states, Colorado and Indiana, that implemented the NMSI College Readiness Program beginning with their first cohort of 20 schools in the 2012–2013 school year, and expanded to an additional 20 schools in their second cohort in the 2013–2014 academic year. They developed a matched comparison sample for treatment schools in each state. They showed that the NMSI CRP increased the likelihood of AP test taking in math and science significantly in the first year following program implementation in both locations. Specifically, they found that the percentage of students who took a STEM-related AP exam increased by 7.80 percentage points for the treatment schools, but deceased by 2.29 percentage points for the comparison schools over the same time period. They further found that these effects persisted into the second year for the first cohort of students and was repeated for the second cohort of students.

Taken together, this collection of studies provides evidence of the effectiveness of the NMSI College Readiness Program. Each of these studies demonstrated the positive impact of the program but, due to their designs and analytic approaches, failed to yield *causal* estimation of the

program's effect on student outcomes. This study extends and complements these previous investigations by applying a potential outcomes approach to estimate the causal effects of CRP on student outcomes.

Methods

This study employs a potential outcomes modeling approach (Rubin, 2005) to estimate the causal effect of program participation on first, second, and third year improvements over the baseline year in AP test taking and AP qualifying score earning in math and science AP subjects. The goal of the propensity score matching is to construct a sample of comparison schools that are similar to the treatment schools (Rosenbaum & Rubin, 1985) in terms of their likelihood of selection into treatment. This model has gained popularity in recent years and is frequently used to make causal estimates from observational studies. Rubin (2005) has argued, "the potential outcomes formulation of causal effects, whether in randomized experiments or in observational studies, has achieved widespread acceptance" (p. 329).

A propensity score is a scalar value that summarizes the likelihood for a unit to receive a treatment, often based on a large set of variables. In this study, we estimate the propensity score using a weighting approach applied in the Toolkit for Weighting and Analysis of Nonequivalent Groups ("twang") package written in the R programming language (Ridgeway, McCaffrey, Morral, Burgette, & Griffin, 2015).

Previous literature suggests that propensity score models should include all confounding variables, that is, variables that are related to the treatment assignment as well as to the outcome (Rubin, 2007; Rubin & Thomas, 1996; West & Thoemmes, 2010), or all variables that are related to the outcome (Rosenbaum, 2002). Stuart (2010) also argues that one should be generous in including predictors in the propensity score model, because the cost of omitting a variable that might predict the outcome is greater than the cost of including a variable that in fact did not predict the outcome (increase in bias versus slight increase in standard errors of propensity scores). In this study, baseline year AP assessment data provide ample information that may predict the outcomes of this study (i.e., number of students taking AP test and student performance on AP tests in STEM-related subjects). In addition, information such as the percentage of AP tests taken by minority and female students is used to balance the treatment and comparison schools. That is, four variables will be used to balance the treatment and control conditions: percentage of AP tests in the baseline year taken by females; percentage of AP tests in the baseline year taken by minority students; number of AP tests taken in the baseline year; and number of AP tests with a qualifying score of 3 or better in the baseline year.

The twang approach to propensity score estimation uses generalized boosted models (GBMs), a multivariate nonparametric regression technique, introduced in McCaffrey, Ridgeway, and Morral (2004). This approach is argued to allow for flexible, nonlinear relationships as well as a large number of variables, and shown to perform well under certain settings (see, e.g., Imai & Ratkovic, 2014). In the GBM approach, instead of matching, a weighting approach is used to estimate the treatment effect. One of the advantages of propensity score approaches is that once non-experimental data are used to "design an observational study" the study achieves balance between treatment and control groups as if it were based on an experimental study (Rubin, 2007). Then, the outcome analysis can proceed in the same way as the analysis that would have been done in an experimental study.

However, note that the effects we seek to obtain can either be the average effect of the treatment on the treated (ATT) or the average treatment effect (ATE). Generally, when we use *matching* strategies based on the estimated propensity scores, we estimate ATT instead of ATE, because we intentionally select and match control group schools that are like treatment schools. However, when we use *weighting* strategies (as is done with the twang package), depending on weights that are used, either ATT or ATE can be obtained. For this study, we estimated the effects of the CRP for both ATT and ATE; however we will focus our attention on the ATT results primarily. Results for ATE analysis are presented in Appendix A.

Data Sources

AP test data from a total of 287 treatment schools from the three most recent NMSI cohorts (108 in Cohort 4, 80 in Cohort 5, and 99 in Cohort 6) plus 10,097 non-treatment schools were analyzed for this study.

Results

The first step in reviewing the results is to check on the extent to which the propensity score weighting approach results in balance across the treatment and control groups in terms of the balancing variables. As mentioned earlier, several variables were used to balance the treatment and control samples. These included: the percentage of AP tests taken in the baseline year by females; the percentage of tests taken in the baseline year by minorities; the total number of tests taken in the baseline year; and the total number of qualifying scores earned at a school in the baseline year. Treatment and control groups for Cohort 4 were fairly balanced prior to weighting on the percentage of tests taken by female and minority students in the baseline year (58.9% vs. 56.2% for females; 24.7% vs. 22.1% for minority students). These minor differences were virtually eliminated through weighting (58.9% vs. 58.7% for females; 24.7% vs. 24.8 for minorities). However, substantial differences between treatment and control schools existed in

average number of tests taken in the baseline year (111.1 vs. 131.5) and in average number of qualifying scores earned (40.3 vs. 77.5). These differences were mitigated through the propensity weighting procedure. After propensity score weighting (ATT estimation), the treatment and control schools were comparable in terms of all four balancing variables (see Figure 2). Specifically, the average number of tests taken in the baseline year for the weighted samples was 111.1 and 108.5 for treatment and control respectively. Likewise, the average number of qualifying scores in the baseline year were balanced at 40.3 for the treatment schools and 44.7 for the control schools. Perfect balance is not to be expected. Austin cautions, "as with randomization, one should not expect that perfect balance will be achieved for all measured baseline variables between treated and untreated subjects in the matched sample" (Austin, 2008, p. 2040).

Similar balance was obtained in Cohorts 5 and 6 through the ATT propensity score weighting approach. For the weighted samples the percentage of baseline year tests taken by females (58.1 vs. 57.9 for Cohort 5; 56.8 vs. 56.8 for Cohort 6), percentage of baseline year tests taken by minorities (18.3 vs. 18.3 for Cohort 5; 30.3 vs. 30.3 for Cohort 6), average number of AP tests taken in the baseline year (128.0 vs. 123.6 for Cohort 5; 165.2 vs. 159.1 for Cohort 6), and average number of qualifying scores earned in the baseline year (54.7 vs. 56.0 for Cohort 5; 74.1 vs. 74.1 for Cohort 6) were comparable between treatment and control groups. For all three cohorts, all post-weighting balancing variables had a mean standardized difference less than 0.2 (see Figure 2, Figure 3, and Figure 4), which indicates good balance between the samples.

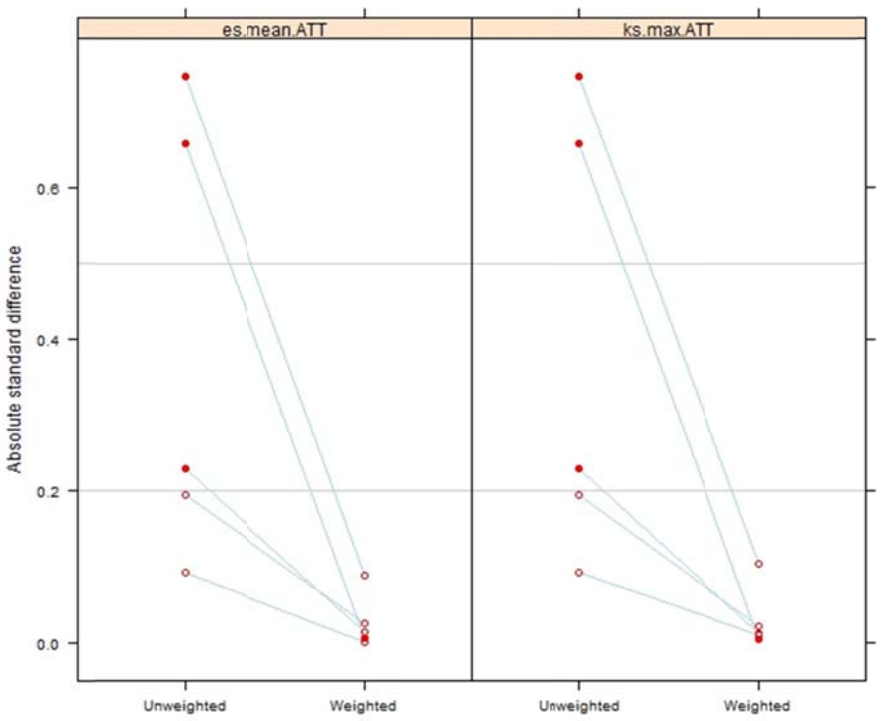


Figure 2. Balance plot for Cohort 4 ATT.

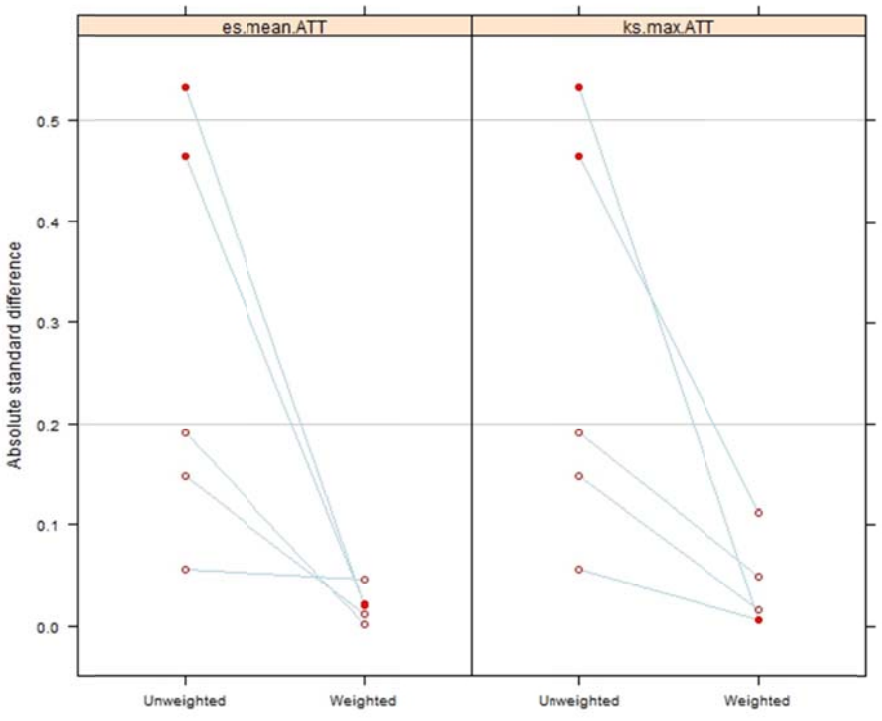


Figure 3. Balance plot for Cohort 5 ATT.

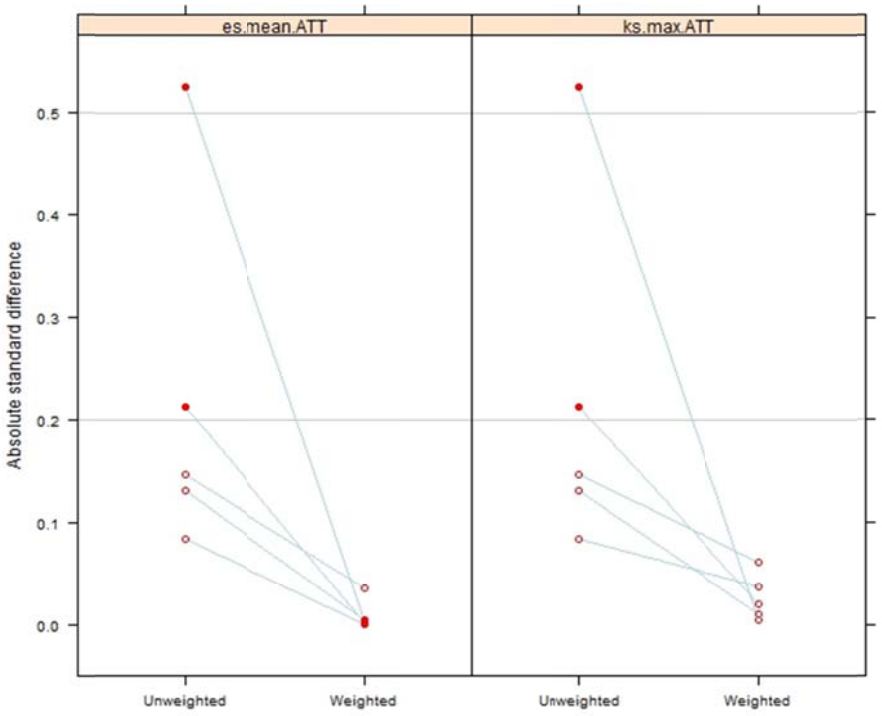


Figure 4. Balance plot for Cohort 6 ATT.

Treatment and control samples were less well balanced using the ATE propensity score estimation procedure (see Figure 5, Figure 6, and Figure 7). Specifically, differences in average number of qualifying scores in the baseline year persisted between the treatment and control groups with the ATE procedure for Cohorts 4 and 5. All of the balance tables for each cohort and both estimating procedures are presented in Appendix B. Given the failure to balance adequately with the ATE procedure, we will focus our attention on the causal estimates from the ATT procedure throughout this report.

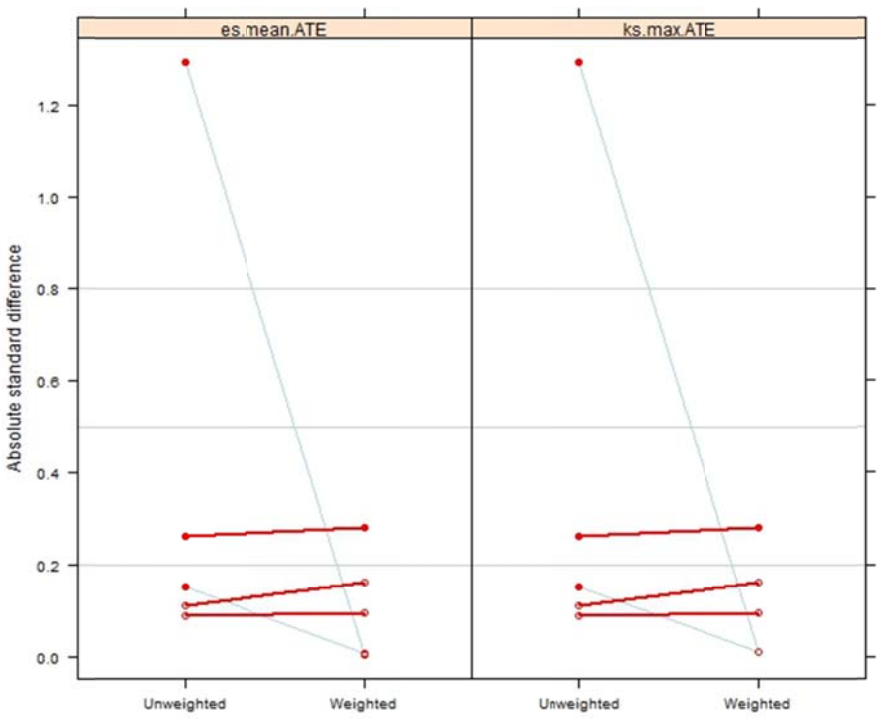


Figure 5. Balance plot for Cohort 4 ATE.

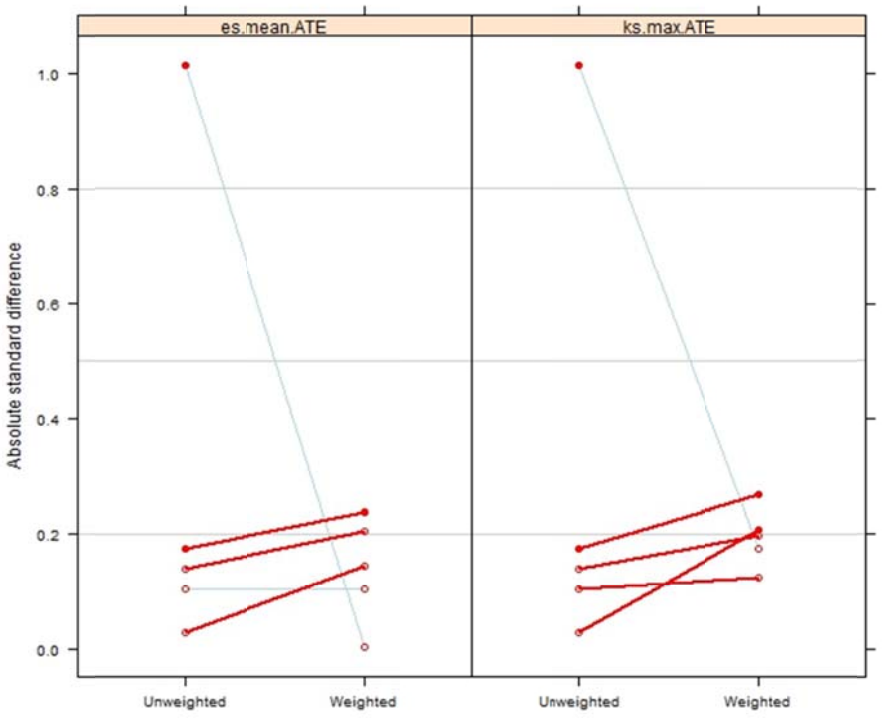


Figure 6. Balance plot for Cohort 5 ATE.

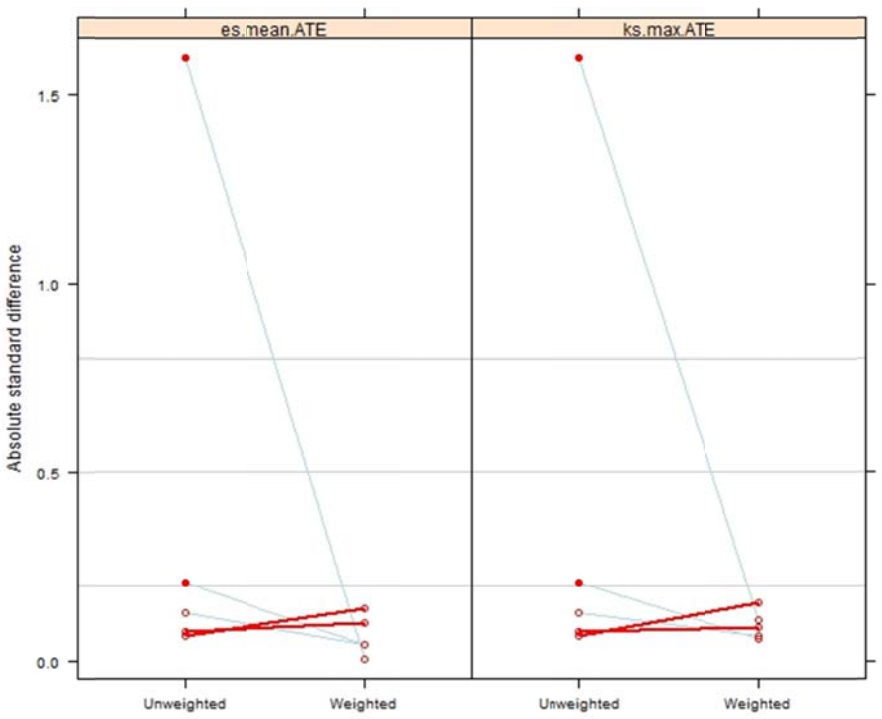


Figure 7. Balance plot for Cohort 6 ATE.

The results of the logistic regressions for the average treatment on the treated (ATT) effect are shown in Table 1, Table 2, Table 3, Table 4, Table 5, and Table 6. Table 1, Table 2, and Table 3 show the impact of the CRP on average school increases in AP test taking, while Table 4, Table 5, and Table 6 show the impact of the CRP on average school increases in earning a qualifying score of 3 or better on AP tests. Similar analyses were conducted for average treatment effects (ATE), the results of which are provided in the Appendix. As indicated in Table 1, the average increase in AP test taking in math, science, English courses rose dramatically for all students over the baseline year in the first, second, and third year following program implementation.

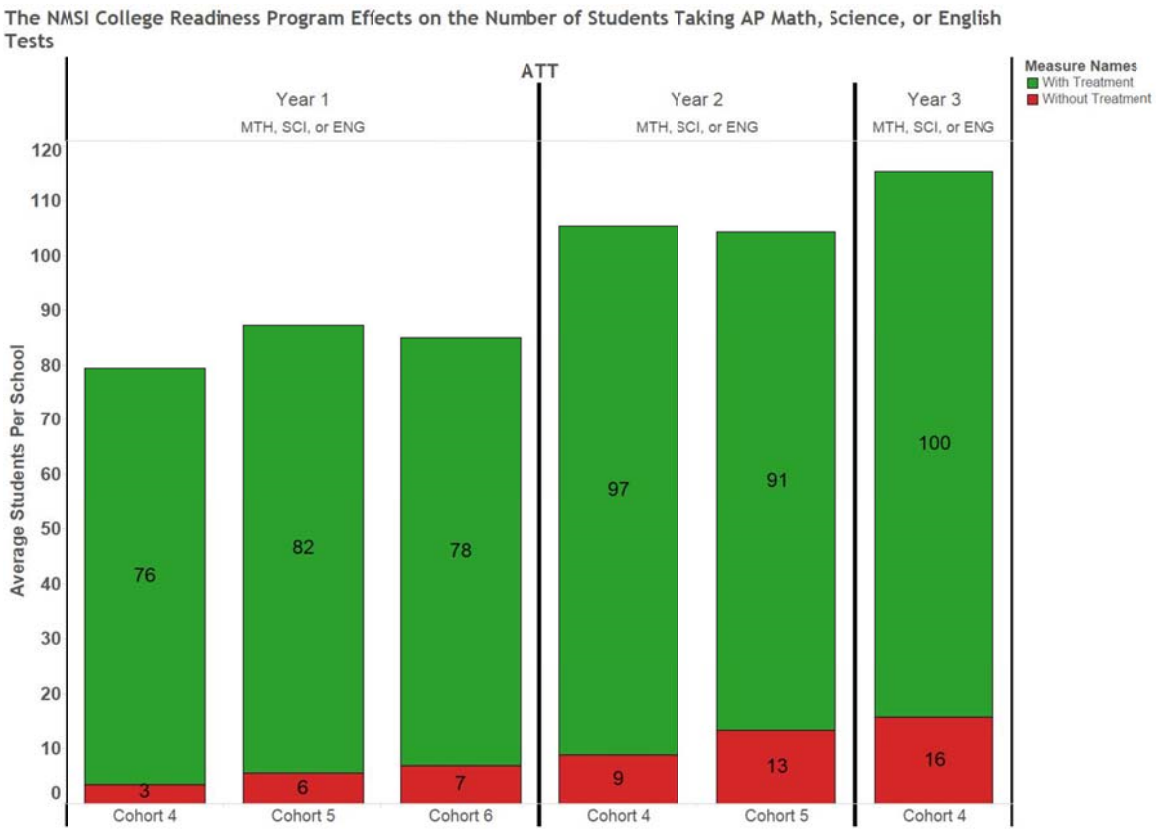


Figure 8. Effect of CRP on AP tests taken.

On average, participation in the CRP caused an average increase of almost 76 additional AP tests taken in the first year for Cohort 4, an additional 81 tests taken for Cohort 5, and an additional 78 tests taken for Cohort 6 (see Figure 8 above). For Cohorts 4 and 5, these results were enhanced in the second year of the program, with average increases over the baseline year of 96 tests for Cohort 4, and 91 tests for Cohort 5. The third year of implementation for Cohort 4 showed an average effect of 100 additional AP tests per school. All of the estimates are highly significant statistically, with standardized effect sizes at or above 1.0 (Cohen's *d*), indicating a very strong causal effect of the program on increases in student AP test taking in math, science, and English courses.

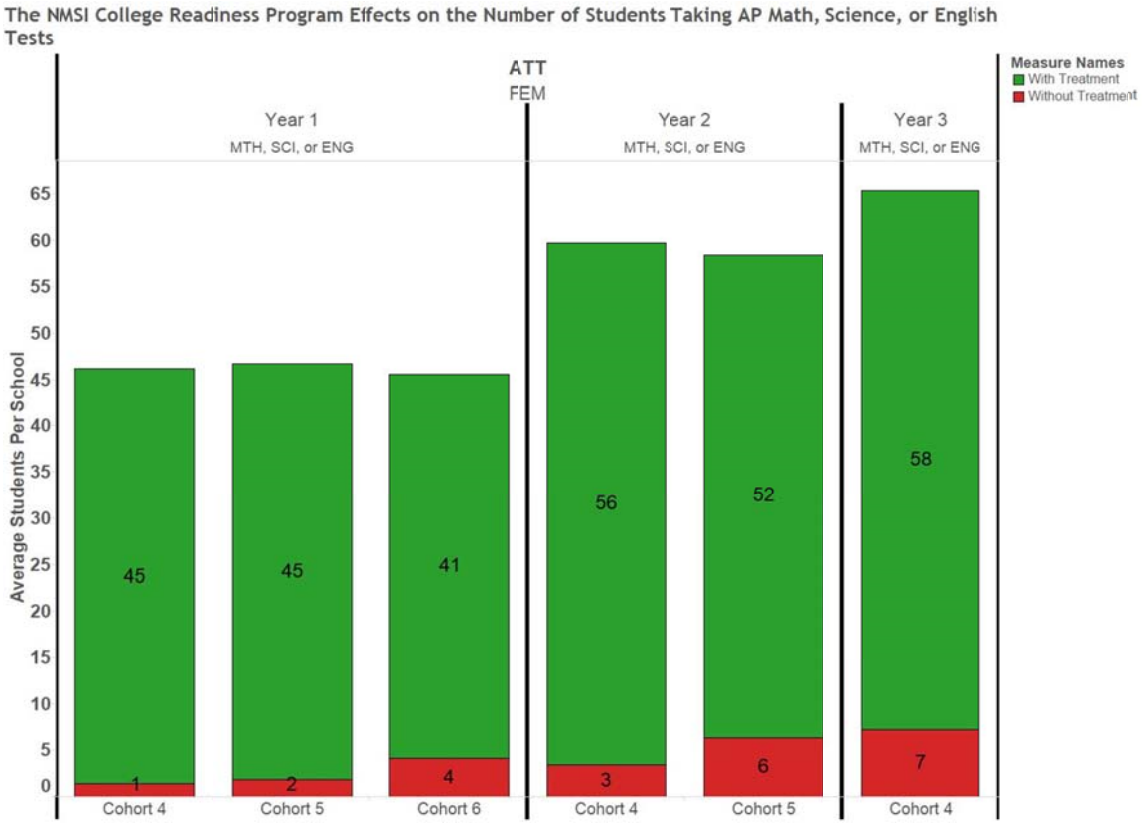


Figure 9. Effect of CRP on AP tests taken by female students.

In addition, the program was effective at increasing AP test taking for female and minority students when analyzed separately. For females, the increase over baseline year AP testing was substantial. Program participation enhanced the increase in the average number of AP tests taken by females by more than 40 tests in the first year, and by more than 50 tests in the second year of implementation. The third year of program implementation saw an effect of an average 58 additional AP tests taken by females per school (see Figure 9 above). As with the overall student population, the standardized values for these effects are substantial, all near or above 1.0 (Cohen's *d*).

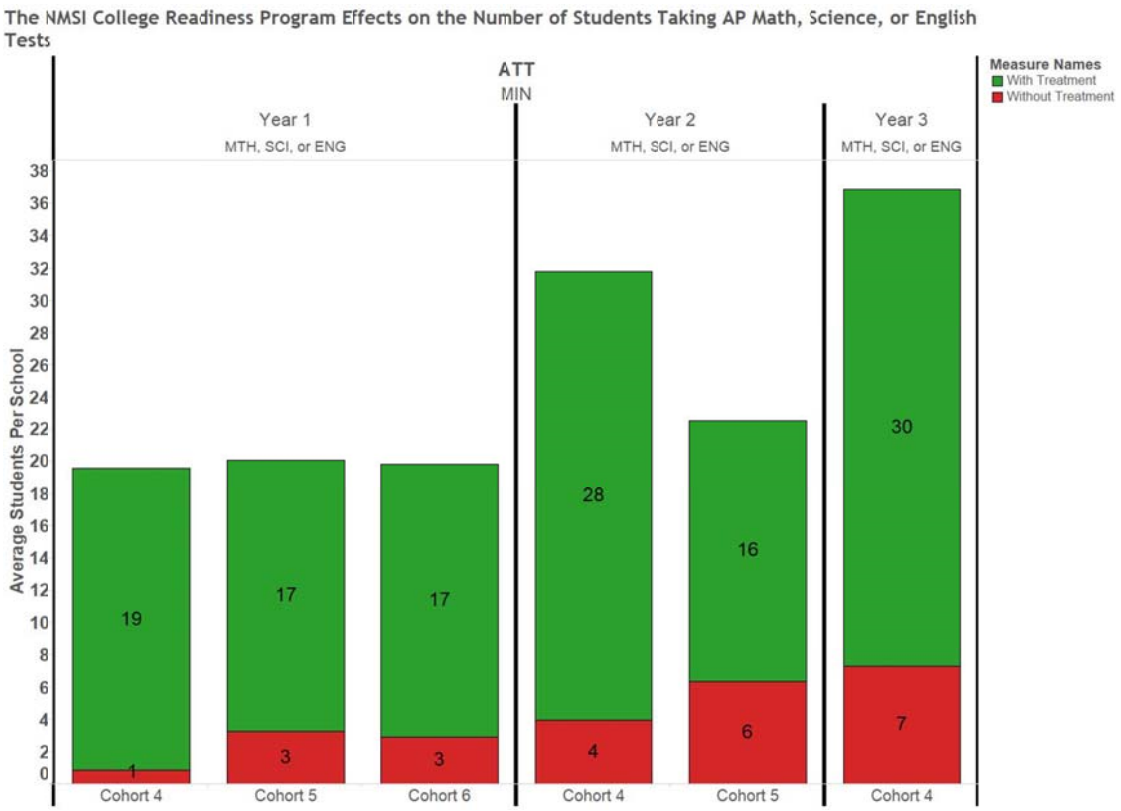


Figure 10. Effect of CRP on AP tests taken by minority students.

Although the standardized effect size estimates were smaller when viewing minority student test taking increases relative to increases for all students or for female students only, they are nonetheless highly significant and substantial. For example, in the first year following implementation, participation in the College Readiness Program increased the number of AP tests taken by minority students an average of over 18 per school for Cohort 4, and over 17 per school in Cohorts 5 and 6. Without the program, these schools would have only increased the number of tests taken by minority students an average of 1 to 3 tests per school (see Figure 10).

Table 1

ATT Estimates for Test Increases—Math, Science, and English

	Estimate	*t* value	*p* value <	Cohen's *d*
All students				
Cohort 4				
Year 1	75.910	11.569	0.000	1.119
Year 2	96.518	11.757	0.000	1.137
Year 3	99.546	11.500	0.000	1.112
Cohort 5				
Year 1	81.684	11.057	0.000	1.241
Year 2	91.061	10.791	0.000	1.211
Cohort 6				
Year 1	78.015	9.603	0.000	0.970
Female students				
Cohort 4				
Year 1	44.756	11.202	0.000	1.083
Year 2	56.275	12.105	0.000	1.171
Year 3	58.155	11.059	0.000	1.070
Cohort 5				
Year 1	44.7576	10.310	0.000	1.157
Year 2	52.1481	10.182	0.000	1.143
Cohort 6				
Year 1	41.297	9.229	0.000	0.932
Minority students				
Cohort 4				
Year 1	18.767	5.588	0.000	0.540
Year 2	27.863	7.016	0.000	0.679
Year 3	29.585	6.025	0.000	0.583
Cohort 5				
Year 1	16.849	5.894	0.000	0.662
Year 2	16.151	5.993	0.000	0.673
Cohort 6				
Year 1	16.946	4.423	0.000	0.447

As would be expected, similar results were found when looking at just math and science AP tests (Table 2) or just English AP tests (Table 3), as these subsamples fully comprise the overall sample. Nevertheless, these findings indicate that the program effects are generalizable across disciplines, as significant and substantial program effects were found for math and science (Table 2) and English (Table 3) separately.

Table 2

ATT Estimates for Test Increases—Math and Science

	Estimate	*t* value	*p* value <	Cohen's *d*
All students				
Cohort 4				
Year 1	38.517	10.710	0.000	1.036
Year 2	53.043	10.124	0.000	0.979
Year 3	55.282	10.230	0.000	0.990
Cohort 5				
Year 1	44.589	10.229	0.000	1.148
Year 2	51.170	9.210	0.000	1.034
Cohort 6				
Year 1	44.780	8.34	0.000	0.842
Female students				
Cohort 4				
Year 1	21.839	10.233	0.000	0.990
Year 2	29.316	10.593	0.000	1.025
Year 3	31.216	9.972	0.000	0.965
Cohort 5				
Year 1	22.049	9.311	0.000	1.045
Year 2	27.448	8.981	0.000	1.008
Cohort 6				
Year 1	21.133	7.688	0.000	0.776
Minority students				
Cohort 4				
Year 1	9.553	5.809	0.000	0.562
Year 2	14.628	6.072	0.000	0.587
Year 3	16.813	5.907	0.000	0.571
Cohort 5				
Year 1	8.725	5.850	0.000	0.657
Year 2	8.348	5.819	0.000	0.653
Cohort 6				
Year 1	8.952	3.783	0.000	0.382

Table 3

ATT Estimates for Test Increases—English

	Estimate	*t* value	*p* value <	Cohen's *d*
All students				
Cohort 4				
Year 1	35.938	7.923	0.000	0.766
Year 2	42.963	9.352	0.000	0.905
Year 3	42.483	8.684	0.000	0.840
Cohort 5				
Year 1	36.352	8.795	0.000	0.987
Year 2	39.063	8.755	0.000	0.983
Cohort 6				
Year 1	31.688	7.449	0.000	0.752
Female students				
Cohort 4				
Year 1	22.014	7.497	0.000	0.725
Year 2	26.569	8.925	0.000	0.863
Year 3	25.808	8.107	0.000	0.784
Cohort 5				
Year 1	22.131	8.684	0.000	0.975
Year 2	24.218	8.300	0.000	0.932
Cohort 6				
Year 1	19.096	7.606	0.000	0.768
Minority students				
Cohort 4				
Year 1	8.859	4.118	0.000	0.398
Year 2	27.863	7.016	0.000	0.679
Year 3	29.584	6.025	0.000	0.583
Cohort 5				
Year 1	8.649	4.677	0.000	0.525
Year 2	8.221	4.561	0.000	0.512
Cohort 6				
Year 1	8.009	3.442	0.000	0.348

In addition to increasing the number of students taking AP tests, the NMSI College Readiness Program increased the number of qualifying scores earned by students in STEM-related AP disciplines. Figure 11 shows that participation in CRP resulted in an additional 23 qualifying scores, on average, for Cohort 4 schools following the first year of implementation. This jumped to an average of 26 additional tests in Year 2, and up to an average of 27 additional tests following the third year of implementation. Cohort 5 results were even better, with an average first year effect of 30 additional qualifying scores and a second year effect of an average of 32 additional qualifying scores. Cohort 6 saw a one year effect of an average of 24 additional qualifying scores per school. Taken together, these findings indicate that more than 15,500 additional qualifying scores were earned as a result of participation in the NMSI CRP in these three cohorts of schools in the last three years. These gains are fairly equally distributed between math and science courses and English Courses.

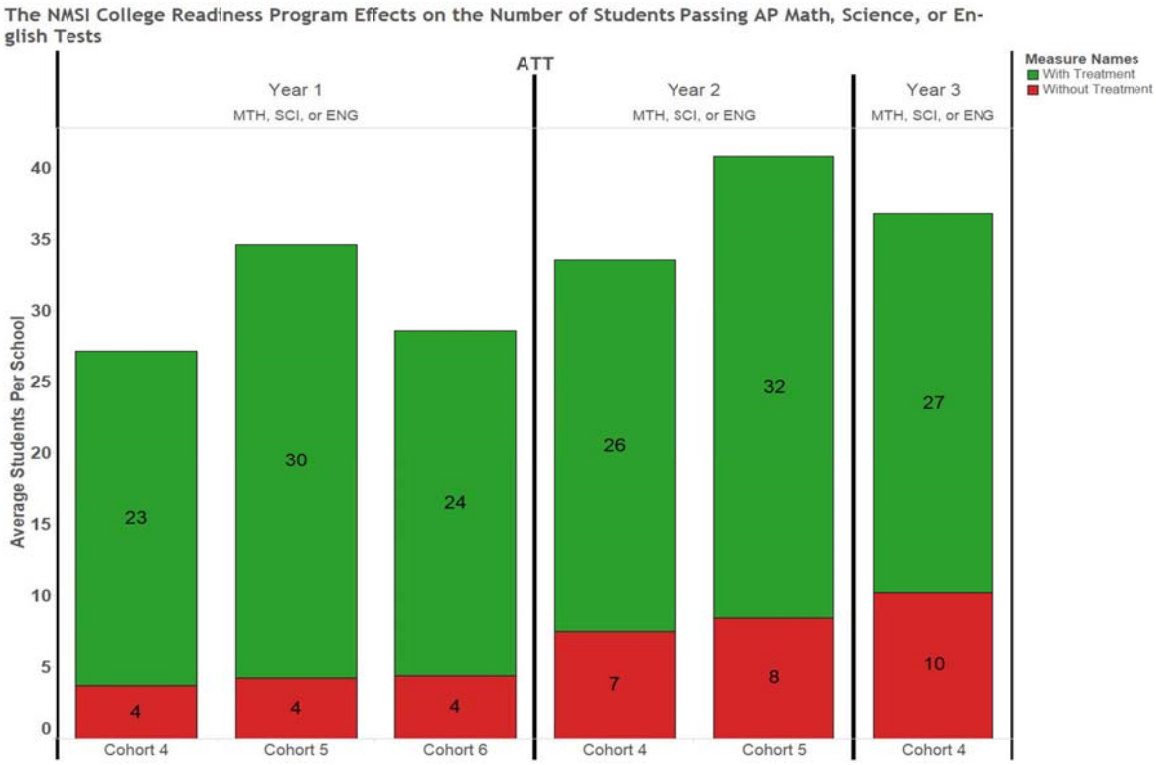


Figure 11. Effect of CRP on qualifying scores earned.

Table 4

ATT Estimates for Qualifying Score Increases—Math, Science, and English

	Estimate	*t* value	*p* value <	Cohen's *d*
All students				
Cohort 4				
Year 1	23.479	8.462	0.000	0.819
Year 2	26.129	7.445	0.000	0.720
Year 3	26.658	7.121	0.000	0.689
Cohort 5				
Year 1	30.428	8.103	0.000	0.910
Year 2	32.394	7.209	0.000	0.809
Cohort 6				
Year 1	24.205	6.733	0.000	0.680
Female students				
Cohort 4				
Year 1	13.160	8.483	0.000	0.820
Year 2	14.764	7.644	0.000	0.739
Year 3	15.125	7.415	0.000	0.717
Cohort 5				
Year 1	15.308	7.464	0.000	0.838
Year 2	17.655	6.288	0.000	0.706
Cohort 6				
Year 1	12.095	6.387	0.000	0.645
Minority students				
Cohort 4				
Year 1	3.029	4.617	0.000	0.447
Year 2	3.506	3.546	0.000	0.343
Year 3	4.110	3.654	0.000	0.353
Cohort 5				
Year 1	3.157	4.196	0.000	0.471
Year 2	3.598	3.808	0.000	0.427
Cohort 6				
Year 1	1.688	1.689	**0.091 (NS)**	0.171

It is worth noting that although the general pattern of substantial and significant causal effects of the NMSI CRP on student testing outcomes is apparent, there is much variation in the size of the effect across the variety of comparisons (see Figure 12), and in a few particular instances, no significant effect of the program was found. For Cohort 6, no significant program effect on the number of qualifying scores earned by minority students was found when looking at math, science, and English tests together, just math and science tests, or just English tests. Of the 108 estimates of average treatment on the treated (ATT) effects, these three were the only ones to not reach an acceptable level of statistical significance; 105 of the 108 comparisons did show a strong, significant effect (97.2%). One possible source of explanation may involve the fact that Cohort 6 schools had a much higher average number of qualifying scores earned in the baseline year (roughly 70) compared to Cohort 5 (55 to 56) and Cohort 4 (40 to 44).

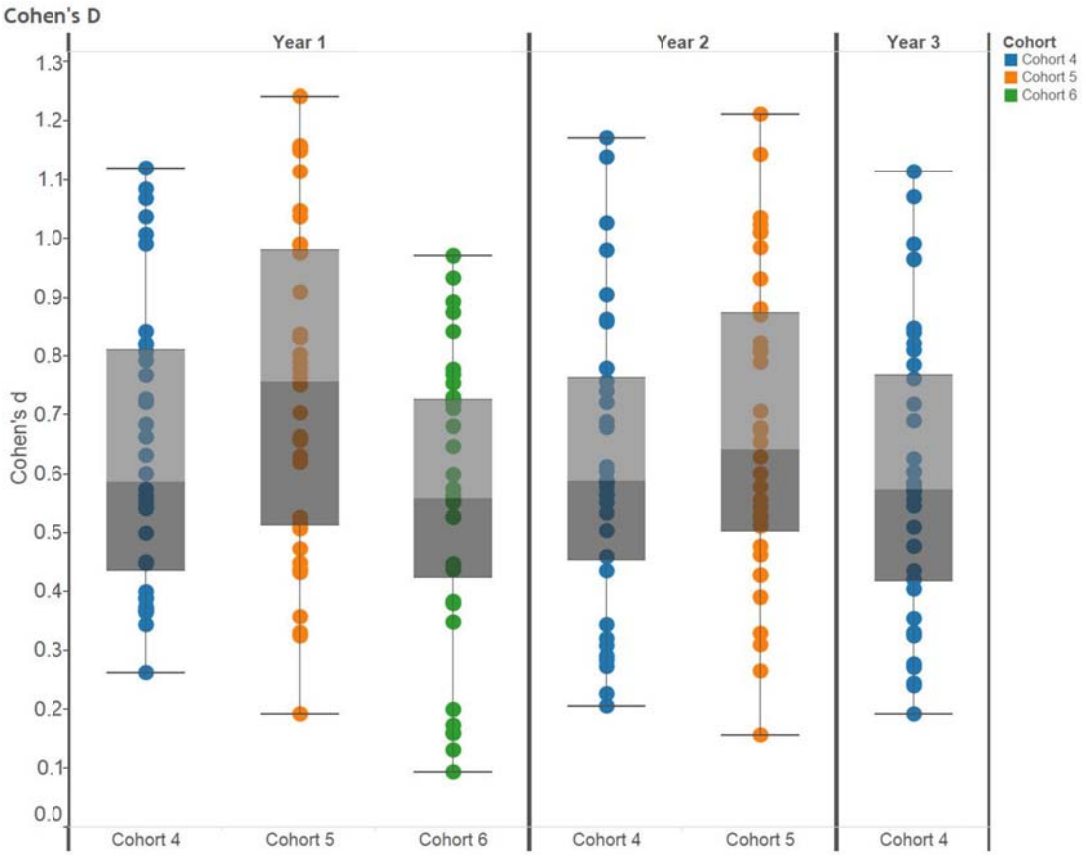


Figure 12. Variation in standardized program effect sizes (216 analyses).

Further investigation of the sources of variation in the effect size estimation is warranted. Clearly, which estimand (ATT or ATE) is being estimated has an effect on the size of the standardized effect estimate. As Figure 13 shows, standardized effect estimates are larger for ATT estimation than for ATE estimation. It also appears the average treatment on the treated

(ATT) impact of the CRP is stronger for increasing AP test taking in math, science, and English courses (mean effect size of .941) than it is for increasing qualifying score attainment (mean effect size of .628) in these courses.

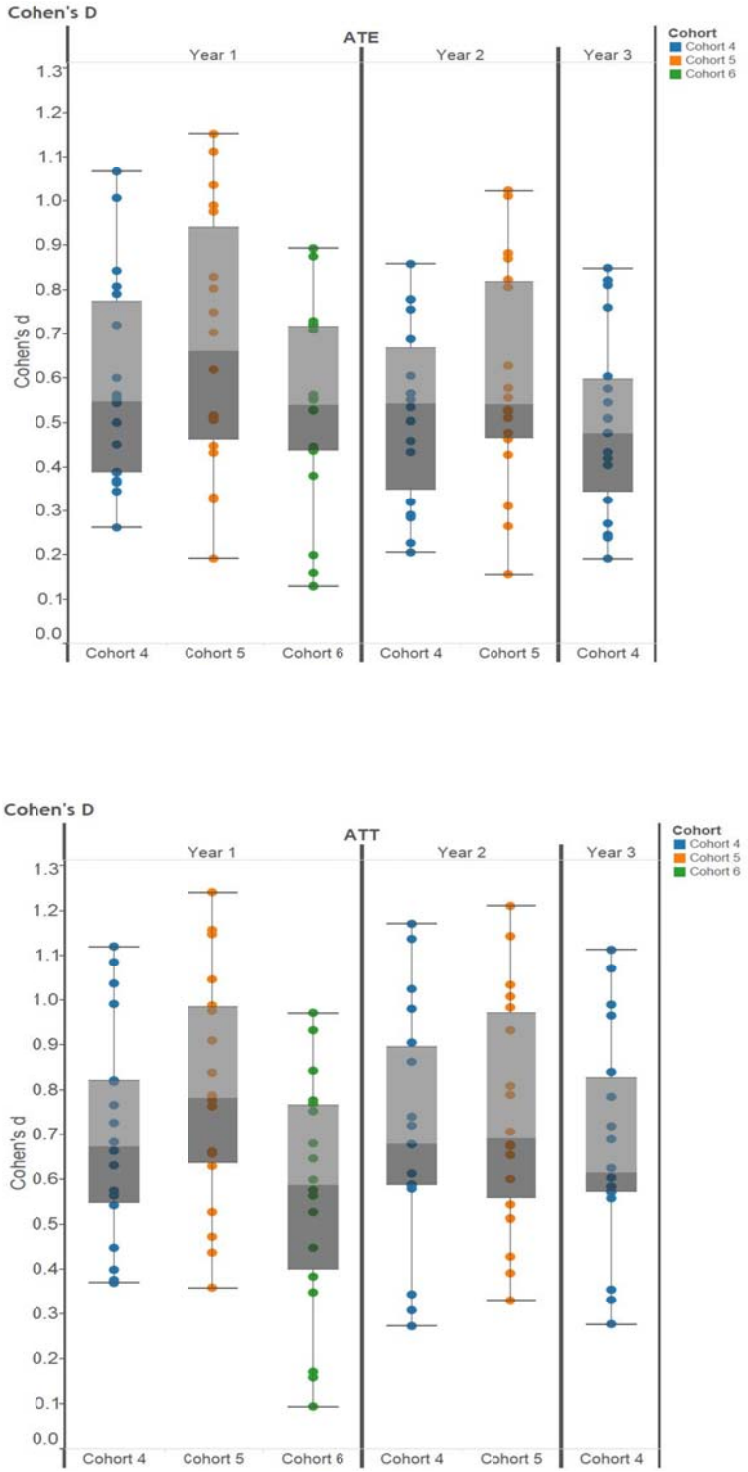


Figure 13. Variation in standardized program effect sizes.

Table 5

ATT Estimates for Qualifying Score Increases—Math and Science

	Estimate	*t* value	*p* value <	Cohen's *d*
All students				
Cohort 4				
Year 1	11.220	6.517	0.000	0.630
Year 2	15.144	5.974	0.000	0.578
Year 3	14.728	5.754	0.000	0.557
Cohort 5				
Year 1	17.702	7.020	0.000	0.788
Year 2	20.330	7.028	0.000	0.789
Cohort 6				
Year 1	15.594	5.919	0.000	0.598
Female students				
Cohort 4				
Year 1	5.295	5.922	0.000	0.573
Year 2	7.234	6.057	0.000	0.586
Year 3	7.233	5.896	0.000	0.570
Cohort 5				
Year 1	7.420	5.603	0.000	0.629
Year 2	9.743	6.032	0.000	0.677
Cohort 6				
Year 1	6.416	5.210	0.000	0.526
Minority students				
Cohort 4				
Year 1	1.546	3.811	0.000	0.369
Year 2	2.195	2.816	0.000	0.272
Year 3	2.321	2.854	0.000	0.276
Cohort 5				
Year 1	1.609	3.182	0.000	0.357
Year 2	2.152	3.472	0.000	0.390
Cohort 6				
Year 1	1.303	1.566	**0.117 (NS)**	0.158

Table 6

ATT Estimates for Qualifying Score Increases—English

	Estimate	*t* value	*p* value <	Cohen's *d*
All students				
Cohort 4				
Year 1	12.378	7.058	0.000	0.683
Year 2	10.992	6.324	0.000	0.612
Year 3	11.561	6.453	0.000	0.624
Cohort 5				
Year 1	12.779	6.797	0.000	0.763
Year 2	12.150	5.338	0.000	0.599
Cohort 6				
Year 1	8.730	5.567	0.000	0.562
Female students				
Cohort 4				
Year 1	8.029	6.842	0.000	0.662
Year 2	7.673	6.066	0.000	0.587
Year 3	7.732	6.232	0.000	0.603
Cohort 5				
Year 1	7.845	6.923	0.000	0.777
Year 2	7.916	4.830	0.000	0.542
Cohort 6				
Year 1	5.712	5.688	0.000	0.574
Minority students				
Cohort 4				
Year 1	1.511	3.867	0.000	0.374
Year 2	1.307	3.182	0.001	0.308
Year 3	1.723	3.420	0.001	0.331
Cohort 5				
Year 1	1.615	3.893	0.000	0.437
Year 2	1.539	2.933	0.004	0.329
Cohort 6				
Year 1	0.395	0.916	**0.360 (NS)**	0.093

In sum, the results of this study indicate substantial and significant increases in both AP test taking and qualifying score earning for all students. In addition, significant first year effects for AP test taking and qualifying score earning were found for female students and minority students when analyzed separately. Average effect sizes (Cohen's *d*) for first year increases over both average treatment on treated (ATT) and average treatment effects for all students (ATE), all subgroups of students, both outcomes, and all disciplines was .64, showing a substantial positive causal impact (a total of 216 causal estimates). These first year effects persisted into the second year (average effect size of .64) but diminished slightly in the third year (average effect size of .59).

The effects are stronger when looking only at the average treatment on the treated (ATT) effects, where the average effect size for first year effects was 0.69 across all subsamples and subjects analyzed. This increased to 0.73 for average second year effects and returned to 0.68 for average third year effects. When looking just at the aggregated student samples with ATT estimation, the average standardized first year effects for increased test taking in math, science, and English courses exceeds 1.1, and for increased qualifying score attainment it exceeds .80.

Discussion

The effects of this program may have more distal impacts on students' academic careers. Research shows that students who take AP courses have a greater likelihood of attending college (Mattern, Marini, & Shaw, 2013). Mattern et al. state, "the odds of enrolling in a four-year institution increased by 171% for students who took one AP exam compared with students who took no AP exams. The increase in odds was even higher for students who took more than one AP exam" (p. 5). Students participating in AP classes also earn better grades in college (Shaw, Marini, & Mattern, 2013), and have a greater likelihood of persisting in and graduating from college (Dougherty, Mellor, & Jian, 2006; Hargrove, Godin, & Dodd, 2008). In addition, students who earn qualifying scores on AP tests outperform matched non-AP students on many college outcome measures (Murphy & Dodd, 2009).

This work is significant because it demonstrates the use of propensity score potential outcomes modeling to observational data to yield meaningful and significant causal estimates of program effectiveness in contexts where randomized assignment to treatment condition is either infeasible or impractical. This study provides evidence of the effectiveness of a College Readiness Program that is having significant and important impacts on preparing more students to succeed in math and science careers and improve the future of math and science education in this country.

References

Austin, P. C. (2008). A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Statistics in Medicine, 27,* 2037-2049.

Dougherty, C., Mellor, L., & Jian, S. (2006). *The relationship between Advanced Placement and college graduation* (National Center for Educational Accountability: 2005 AP Study Series, Report 1). Austin, TX: National Center for Educational Accountability.

Hargrove, L., Godin, D., & Dodd, B. (2008). *College outcomes comparisons by AP and non-AP high school experiences* (College Board Research Report 2008-3). New York: The College Board.

Holtzman, D. (2010). *The Advanced Placement Teacher Training Incentive Program (APTIP): Estimating the impact of an incentive and training program on AP taking and passing.* Unpublished manuscript. Washington, DC: American Institutes for Research.

Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B, 53,* 597-610.

Jackson, C. K. (2010a). *The effects of an incentive-based high-school intervention on college outcomes* (NBER Working Paper 15722). Cambridge, MA: National Bureau of Economic Research.

Jackson, C. K. (2010b). A little now for a lot later: A look at a Texas Advanced Placement incentive program. *Journal of Human Resources, 45*(3), 591-639.

Jackson, K. (2012). *Do college prep programs improve long term outcomes?* (Working Paper Series WP-12-12). Evanston, IL: Institute for Policy Research, Northwestern University.

Jackson, K. (2014). Do college preparatory programs improve long-term outcomes? *Economic Inquiry, 52*(1), 72-99.

Mattern, K. D., Marini, J. P., & Shaw, E. J. (2013). *The relationship between AP Exam performance and college outcomes* (College Board Research Report 2009-4). New York: The College Board.

McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods, 9,* 403-425.

Murphy, D., & Dodd, B. (2009). A comparison of college performance of matched AP and non-AP student groups. (College Board Research Report No. 2009-6). New York: The College Board.

Ridgeway, G., McCaffrey, D., Morral, A., Burgette, L., & Griffin, B. (2015). *twang: Toolkit for weighting and analysis of nonequivalent groups.* Available at http://cran.r-project.org/web/packages/twang/twang.pdf.

Rosenbaum, P. R. (2002). *Observational studies* (2nd ed). Springer, New York.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39,* 33-38.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association, 100*(469), 322-331.

Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Statistical Medicine, 26,* 20-36.

Rubin, D. B., & Thomas, N. (1996). Matching using estimated propensity scores, relating theory to practice. *Biometrics, 52,* 249-264.

Shaw, E. J., Marini, J. P., & Mattern, K. D. (2013). Exploring the utility of Advanced Placement participation and performance in college admission decisions. *Educational and Psychological Measurement, 73,* 229-253.

Sherman, D., & Song, M. (2014). *Findings for first-year impacts of APTIP on student outcomes in Cohort 1 schools*. Washington, DC: American Institutes for Research.

Sherman, D., & Song, M. (2015). *Year 2 Findings for APTIP impacts on students' AP performance in Cohorts 1 and 2 schools*. Washington, DC: American Institutes for Research.

Stuart, E. A. (2010). Matching methods for causal inference: a review and a look forward. *Statistical Science, 25*(1), 1-21.

West, S. G., & Thoemmes, F. (2010). Campbell's and Rubin's perspectives on causal inference. *Psychological Methods, 15*(1), 18-37.

Appendix A:
ATE Analysis

Table A1

ATE Estimates for Test Increases—Math, Science, and English

	Estimate	*t* value	*p* value <	Cohen's *d*
All students				
Cohort 4				
Year 1	73.62	11.020	0.000	1.066
Year 2	94.38	8.035	0.000	0.777
Year 3	99.24	8.483	0.000	0.821
Cohort 5				
Year 1	77.44	10.270	0.000	1.153
Year 2	87.29	9.109	0.000	1.022
Cohort 6				
Year 1	82.89	8.834	0.000	0.892
Female students				
Cohort 4				
Year 1	43.04	10.400	0.000	1.006
Year 2	54.57	8.876	0.000	0.859
Year 3	57.97	8.770	0.000	0.848
Cohort 5				
Year 1	41.66	9.913	0.000	1.113
Year 2	50.60	9.005	0.000	1.011
Cohort 6				
Year 1	46.00	8.663	0.000	0.875
Minority students				
Cohort 4				
Year 1	15.19	4.649	0.000	0.450
Year 2	23.32	5.826	0.000	0.564
Year 3	26.90	4.918	0.000	0.476
Cohort 5				
Year 1	14.77	4.589	0.000	0.515
Year 2	12.63	4.692	0.000	0.527
Cohort 6				
Year 1	12.72	5.205	0.000	0.526

Table A2

ATE Estimates for Test Increases—Math and Science

	Estimate	*t* value	*p* value <	Cohen's *d*
All students				
Cohort 4				
Year 1	32.10	8.703	0.000	0.842
Year 2	47.81	6.248	0.000	0.604
Year 3	50.54	5.939	0.000	0.575
Cohort 5				
Year 1	39.61	8.689	0.000	0.975
Year 2	45.25	7.179	0.000	0.806
Cohort 6				
Year 1	45.27	7.117	0.000	0.719
Female students				
Cohort 4				
Year 1	18.40	8.340	0.000	0.807
Year 2	26.16	7.104	0.000	0.687
Year 3	28.78	6.227	0.000	0.602
Cohort 5				
Year 1	18.86	7.389	0.000	0.829
Year 2	25.28	7.323	0.000	0.822
Cohort 6				
Year 1	22.70	7.223	0.000	0.729
Minority students				
Cohort 4				
Year 1	6.89	5.157	0.000	0.499
Year 2	11.54	4.737	0.000	0.458
Year 3	14.48	4.341	0.000	0.420
Cohort 5				
Year 1	7.51	3.983	0.000	0.447
Year 2	5.43	3.803	0.000	0.427
Cohort 6				
Year 1	6.36	4.340	0.000	0.438

Table A3

ATE Estimates for Test Increases—English

	Estimate	*t* value	*p* value <	Cohen's *d*
All students				
Cohort 4				
Year 1	38.11	8.175	0.000	0.791
Year 2	46.69	8.049	0.000	0.779
Year 3	45.84	8.369	0.000	0.810
Cohort 5				
Year 1	35.70	8.816	0.000	0.990
Year 2	38.08	7.749	0.000	0.870
Cohort 6				
Year 1	34.85	7.021	0.000	0.709
Female students				
Cohort 4				
Year 1	22.70	7.438	0.000	0.720
Year 2	28.33	7.811	0.000	0.756
Year 3	27.29	7.855	0.000	0.760
Cohort 5				
Year 1	21.36	9.220	0.000	1.035
Year 2	23.06	7.849	0.000	0.881
Cohort 6				
Year 1	21.66	7.159	0.000	0.723
Minority students				
Cohort 4				
Year 1	7.11	4.019	0.000	0.389
Year 2	23.32	5.826	0.000	0.564
Year 3	26.90	4.918	0.000	0.476
Cohort 5				
Year 1	8.31	4.577	0.000	0.514
Year 2	7.76	4.115	0.000	0.462
Cohort 6				
Year 1	6.12	3.749	0.000	0.379

Table A4

ATE Estimates for Qualifying Score Increases—Math, Science, and English

	Estimate	*t* value	*p* value <	Cohen's *d*
All students				
Cohort 4				
Year 1	22.39	5.686	0.000	0.550
Year 2	24.81	4.484	0.000	0.434
Year 3	24.43	4.174	0.000	0.404
Cohort 5				
Year 1	25.62	6.256	0.000	0.702
Year 2	28.13	5.593	0.000	0.628
Cohort 6				
Year 1	25.35	5.476	0.000	0.553
Female students				
Cohort 4				
Year 1	12.37	5.802	0.000	0.561
Year 2	13.64	5.191	0.000	0.502
Year 3	13.32	4.484	0.000	0.434
Cohort 5				
Year 1	12.39	5.507	0.000	0.618
Year 2	15.47	4.946	0.000	0.555
Cohort 6				
Year 1	12.95	5.455	0.000	0.551
Minority students				
Cohort 4				
Year 1	2.67	3.804	0.000	0.368
Year 2	2.42	2.928	0.003	0.283
Year 3	3.79	2.798	0.005	0.271
Cohort 5				
Year 1	1.99	2.894	0.004	0.325
Year 2	1.83	2.358	0.018	0.265
Cohort 6				
Year 1	1.34	1.964	0.050	0.198

Table A5

ATE Estimates for Qualifying Score Increases—Math and Science

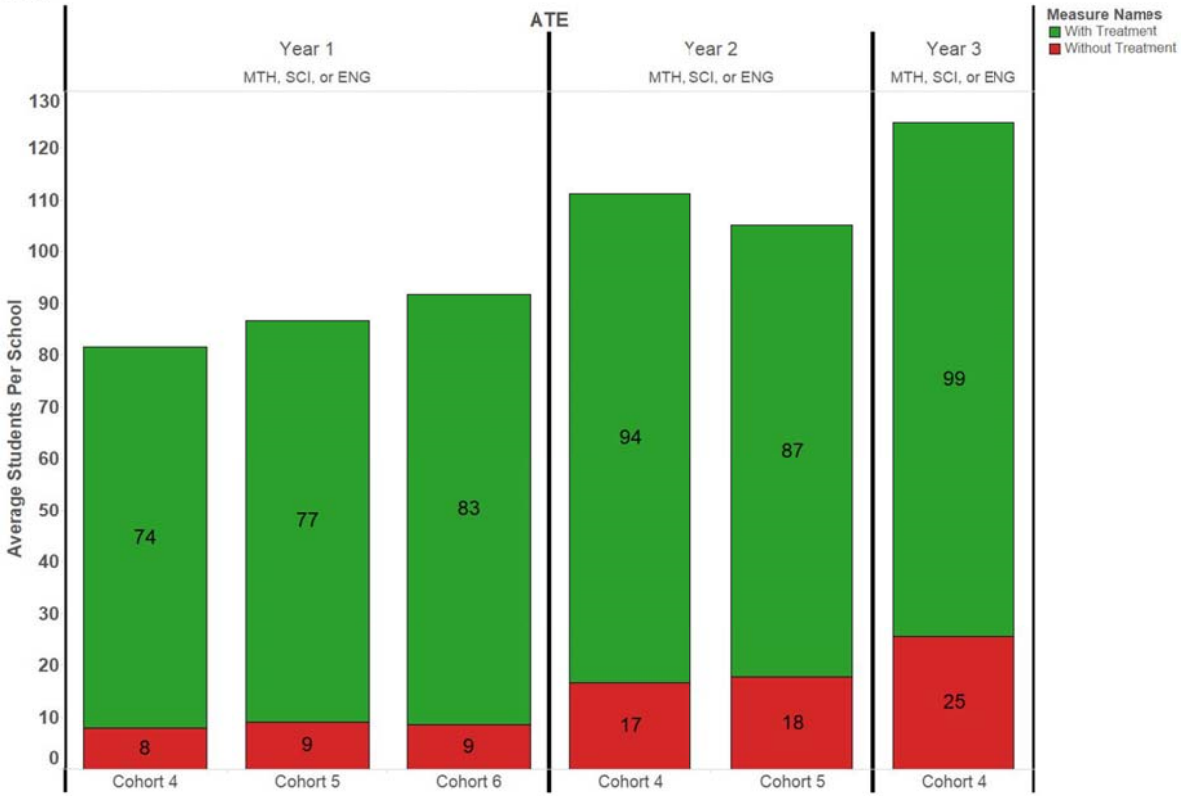
	Estimate	*t* value	*p* value <	Cohen's *d*
All students				
Cohort 4				
Year 1	9.09	3.763	0.000	0.364
Year 2	12.40	2.994	0.003	0.290
Year 3	10.68	2.523	0.012	0.244
Cohort 5				
Year 1	12.82	4.502	0.000	0.505
Year 2	15.31	4.670	0.000	0.524
Cohort 6				
Year 1	15.95	4.315	0.000	0.436
Female students				
Cohort 4				
Year 1	4.49	3.549	0.000	0.343
Year 2	5.78	3.299	0.001	0.319
Year 3	4.95	2.464	0.014	0.238
Cohort 5				
Year 1	4.59	2.939	0.003	0.330
Year 2	7.38	4.242	0.000	0.476
Cohort 6				
Year 1	6.35	4.407	0.000	0.445
Minority students				
Cohort 4				
Year 1	1.37	2.699	0.007	0.261
Year 2	1.51	2.112	0.035	0.204
Year 3	1.96	1.964	0.050	0.190
Cohort 5				
Year 1	0.81	1.699	0.095	0.191
Year 2	0.74	1.372	0.170	0.154
Cohort 6				
Year 1	0.88	1.568	0.117	0.158

Table A6

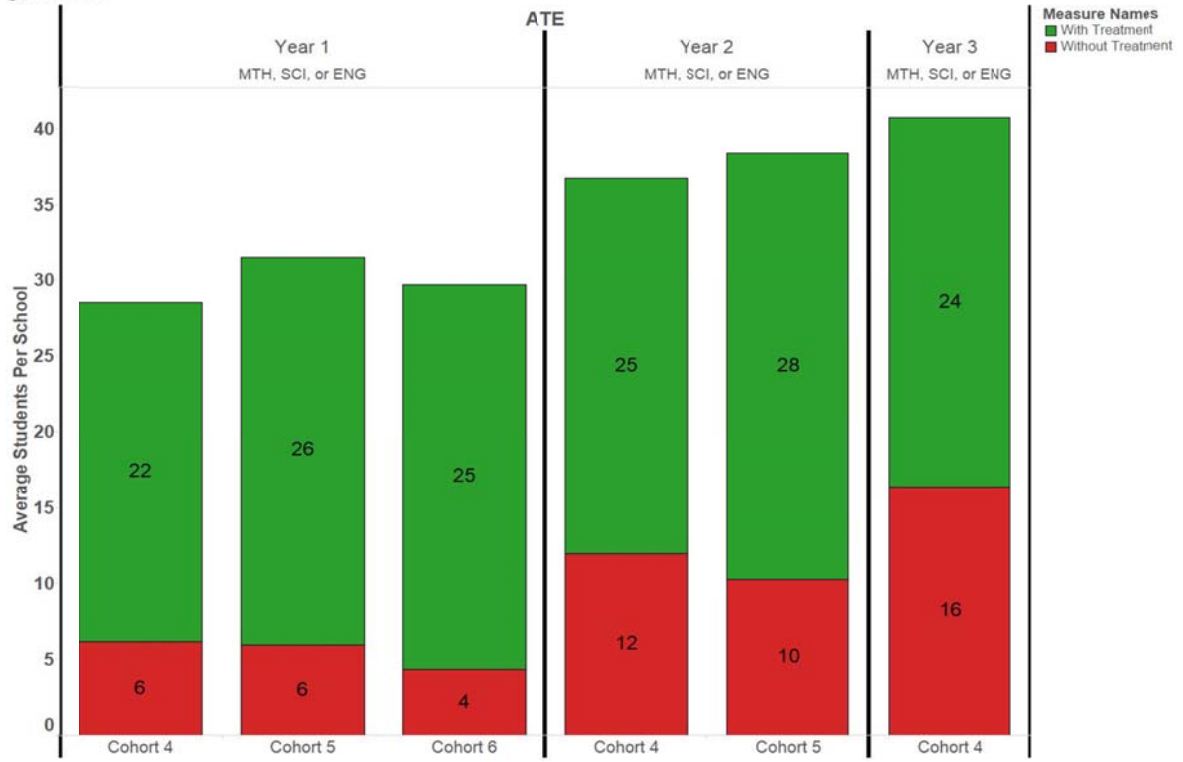
ATE Estimates for Qualifying Score Increases—English

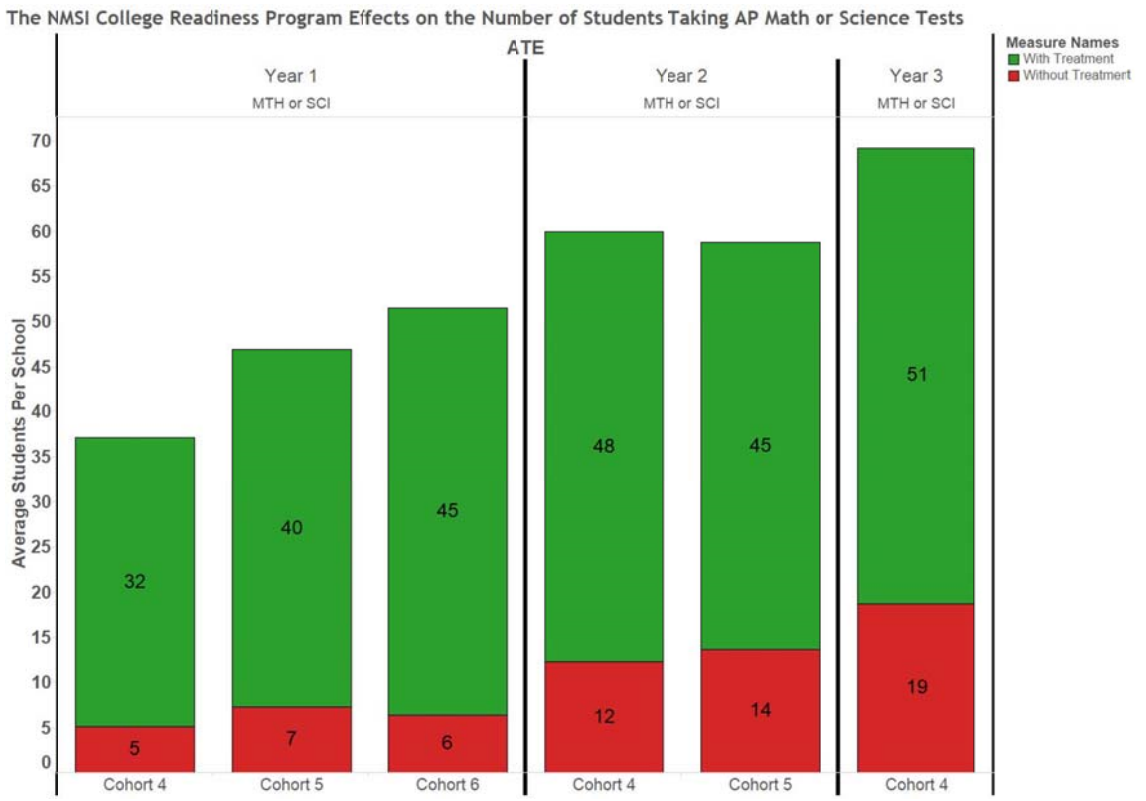
	Estimate	*t* value	*p* value <	Cohen's *d*
All students				
Cohort 4				
Year 1	13.685	6.195	0.000	0.599
Year 2	12.652	5.694	0.000	0.551
Year 3	13.135	5.628	0.000	0.544
Cohort 5				
Year 1	12.496	6.674	0.000	0.749
Year 2	12.003	5.138	0.000	0.577
Cohort 6				
Year 1	9.371	5.562	0.000	0.562
Female students				
Cohort 4				
Year 1	8.332	5.616	0.000	0.543
Year 2	8.314	5.512	0.000	0.533
Year 3	8.109	5.259	0.000	0.509
Cohort 5				
Year 1	7.525	7.146	0.000	0.802
Year 2	7.572	4.548	0.000	0.510
Cohort 6				
Year 1	6.69	5.219	0.000	0.527
Minority students				
Cohort 4				
Year 1	1.38	3.997	0.000	0.387
Year 2	0.90	2.324	0.020	0.225
Year 3	1.77	3.346	0.001	0.324
Cohort 5				
Year 1	1.29	3.847	0.000	0.432
Year 2	1.14	2.763	0.006	0.310
Cohort 6				
Year 1	0.45	1.279	0.201	0.129

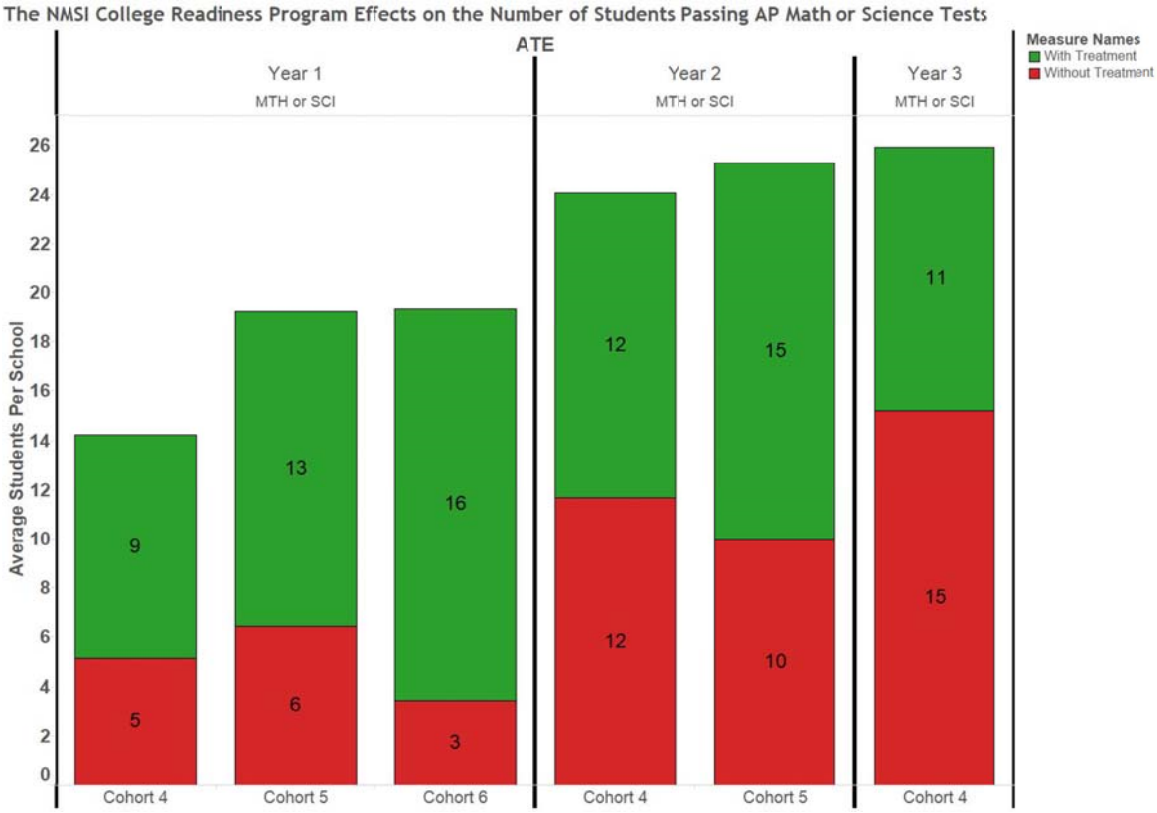
The NMSI College Readiness Program Effects on the Number of Students Taking AP Math, Science, or English Tests



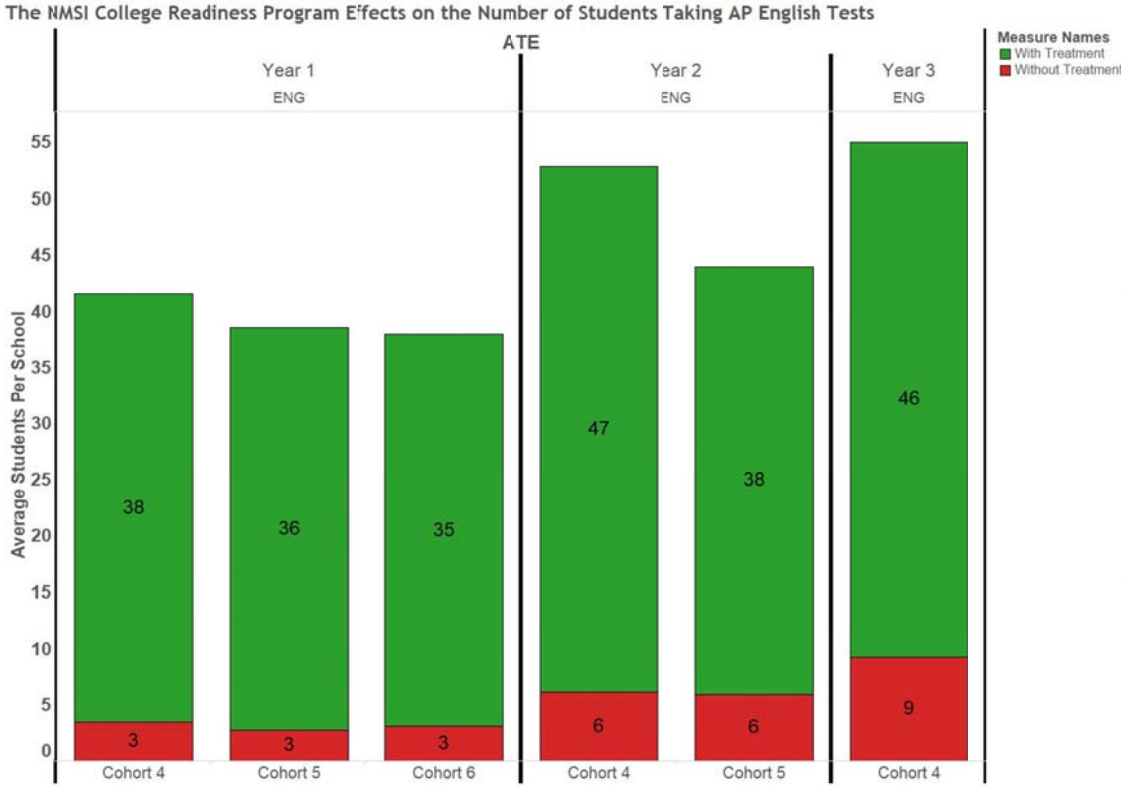
The NMSI College Readiness Program Effects on the Number of Students Passing AP Math, Science, or English Tests



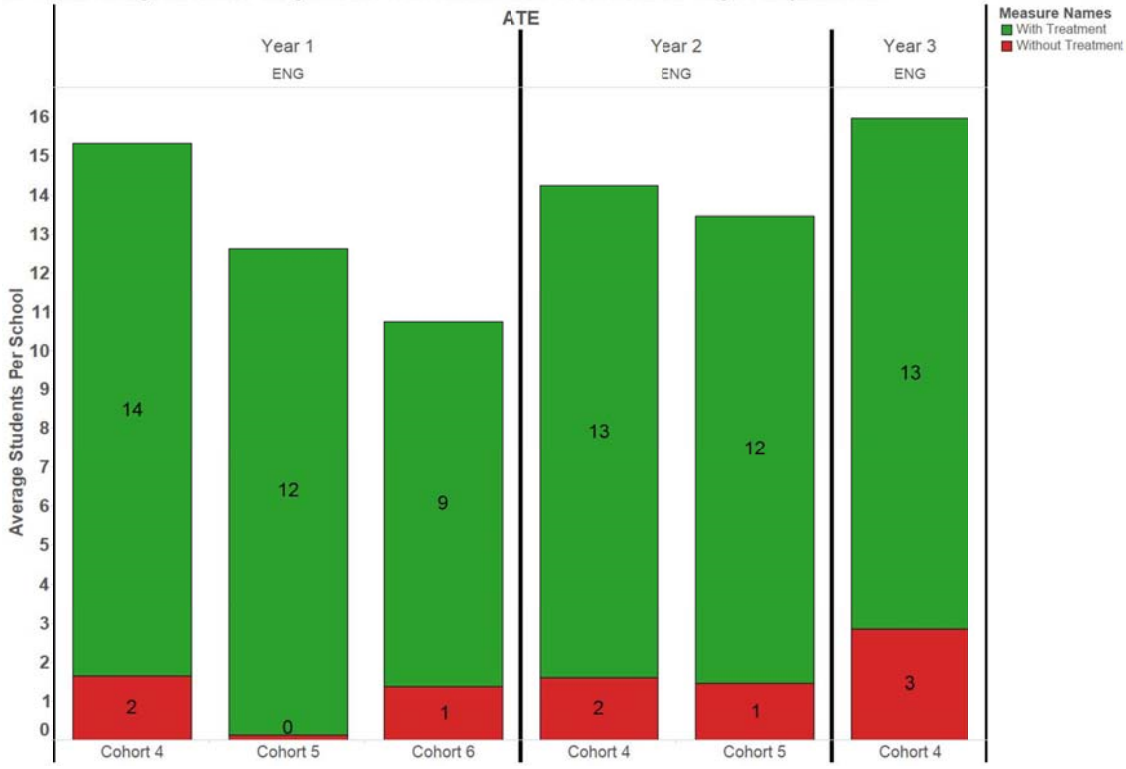

The NMSI College Readiness Program Effects on the Number of Students Taking AP Math or Science Tests
ATE
Year 1
MTH or SCI
Year 2
MTH or SCI
Year 3
MTH or SCI
Measure Names
With Treatment
Without Treatment
Average Students Per School
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
32
5
40
7
45
6
48
12
45
14
51
19
Cohort 4
Cohort 5
Cohort 6
Cohort 4
Cohort 5
Cohort 4


The NMSI College Readiness Program Effects on the Number of Students Passing AP Math or Science Tests
ATE
Year 1
MTH or SCI
Year 2
MTH or SCI
Year 3
MTH or SCI
Measure Names
With Treatment
Without Treatment
Average Students Per School
0
2
4
6
8
10
12
14
16
18
20
22
24
26
9
5
13
6
16
3
12
12
15
10
11
15
Cohort 4
Cohort 5
Cohort 6
Cohort 4
Cohort 5
Cohort 4

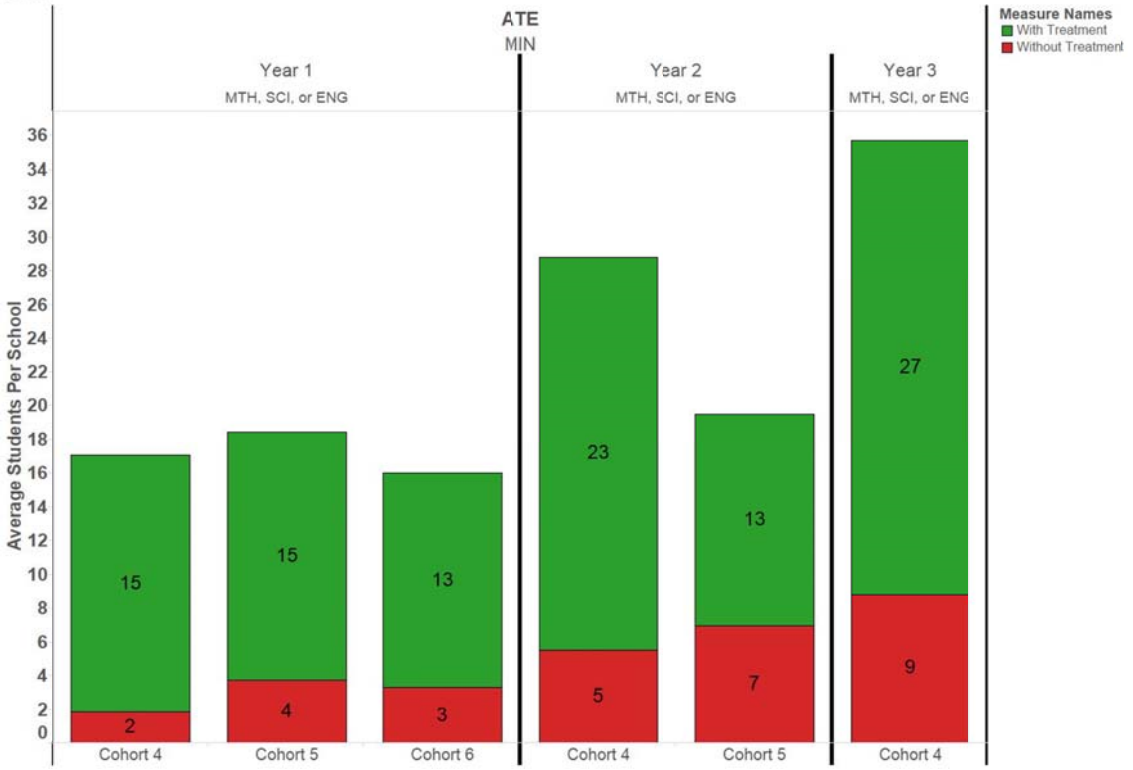
The NMSI College Readiness Program Effects on the Number of Students Taking AP English Tests



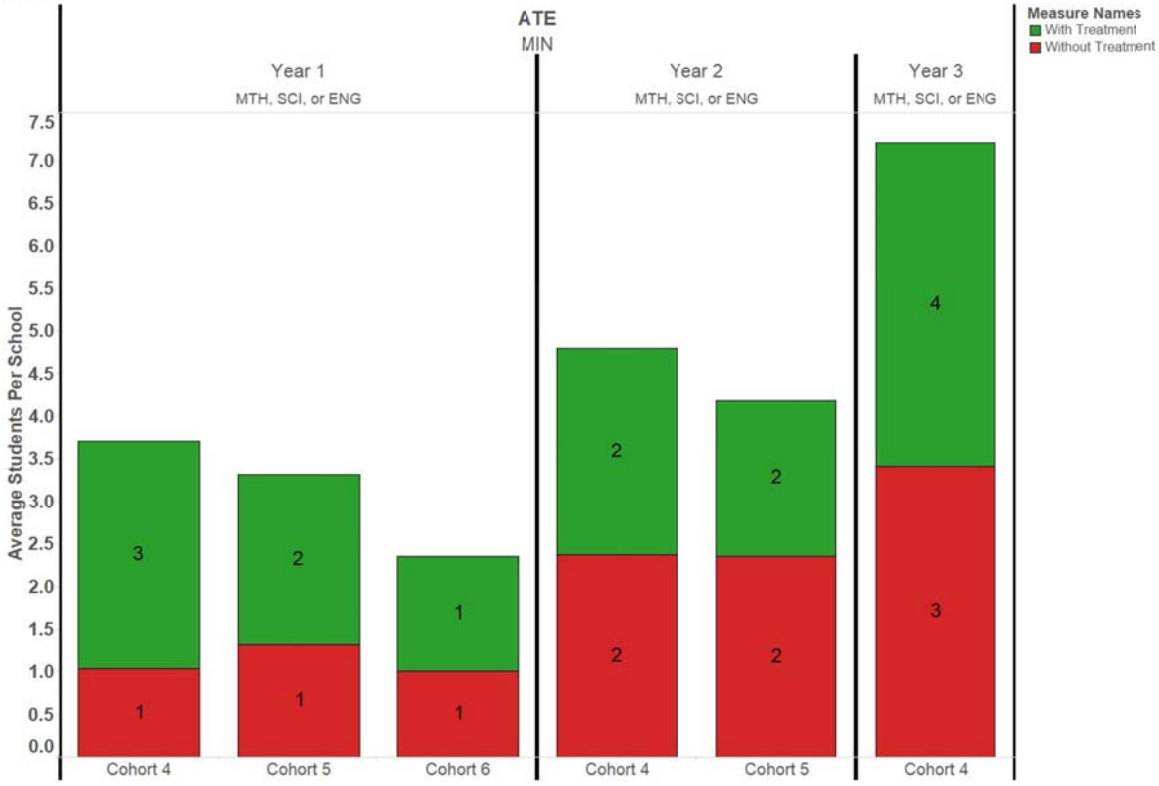
The NMSI College Readiness Program Effects on the Number of Students Passing AP English Tests



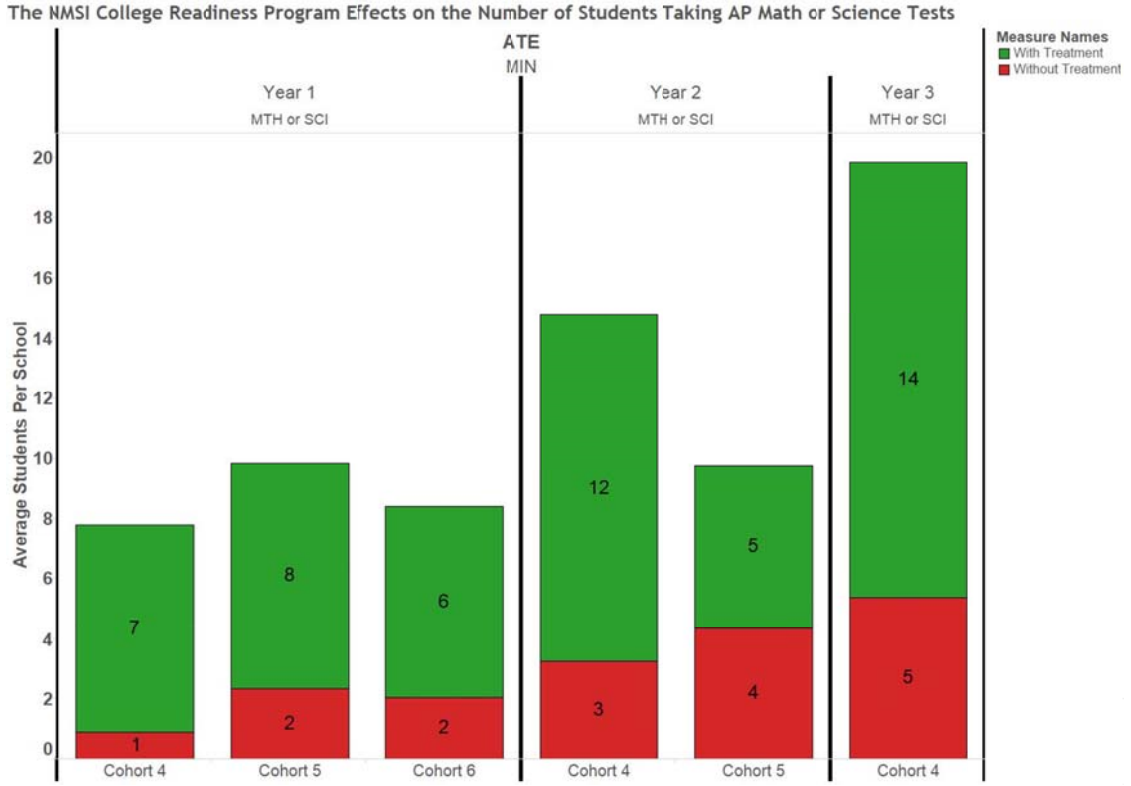
The NMSI College Readiness Program Effects on the Number of Students Taking AP Math, Science, or English Tests



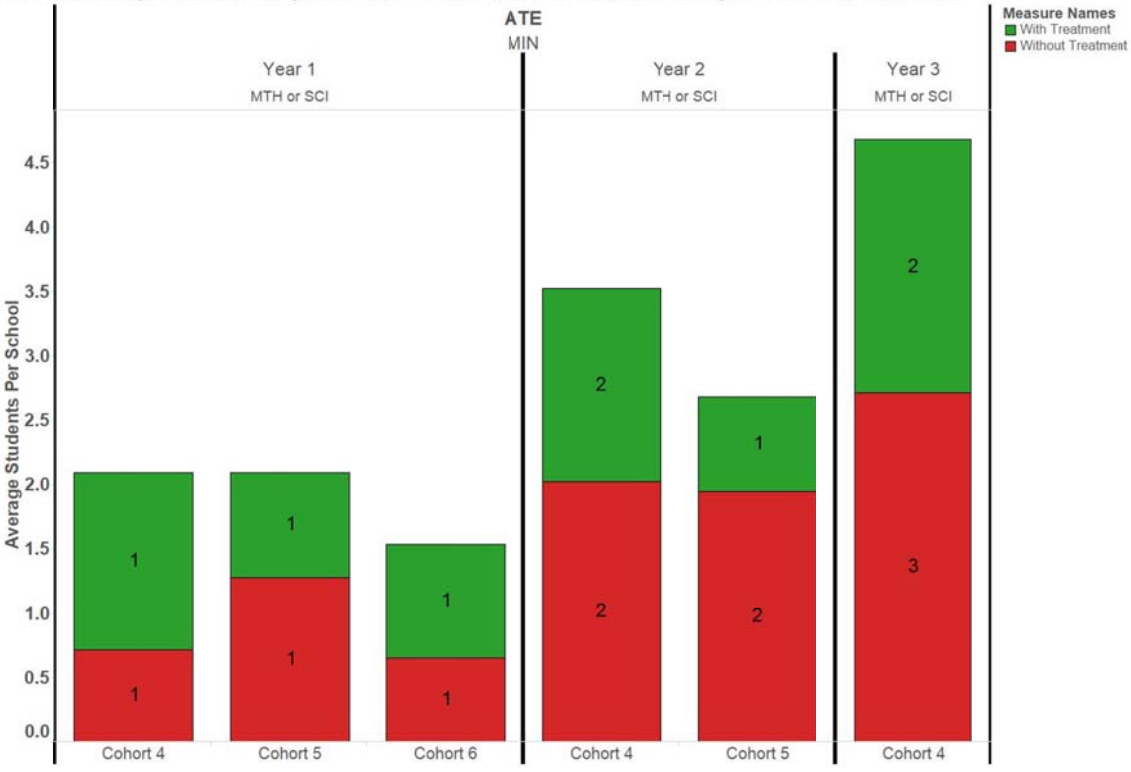
The NMSI College Readiness Program Effects on the Number of Students Passing AP Math, Science, or English Tests



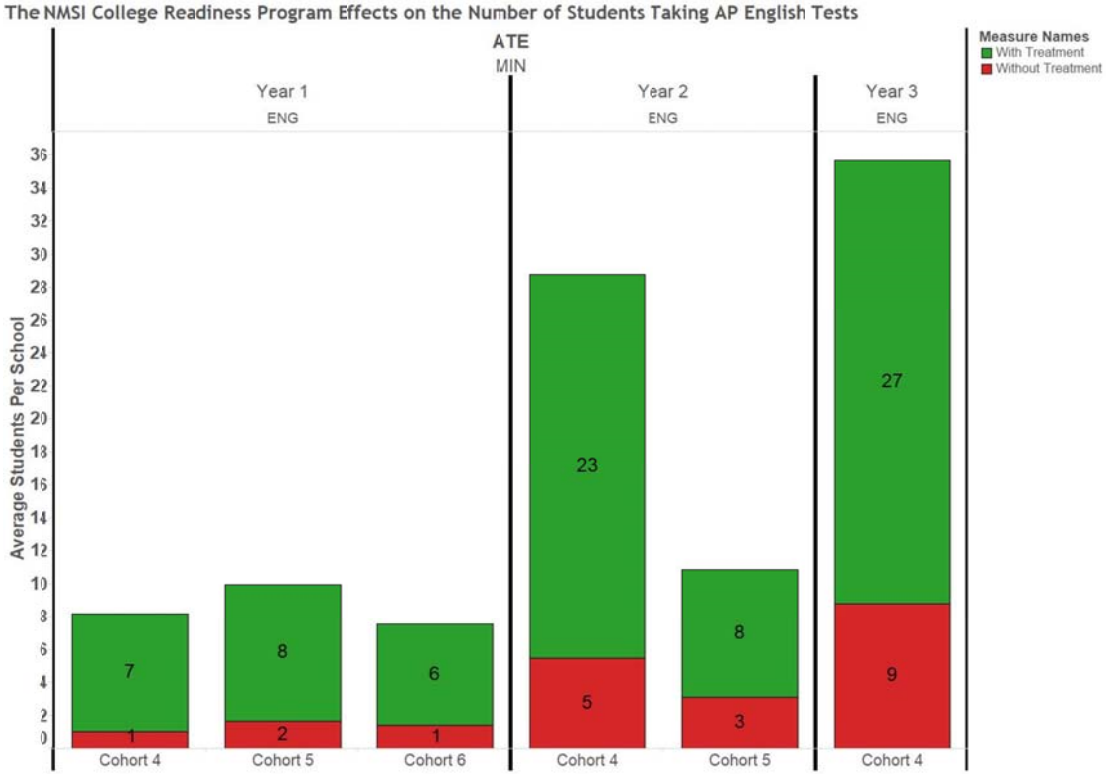
The NMSI College Readiness Program Effects on the Number of Students Taking AP Math or Science Tests



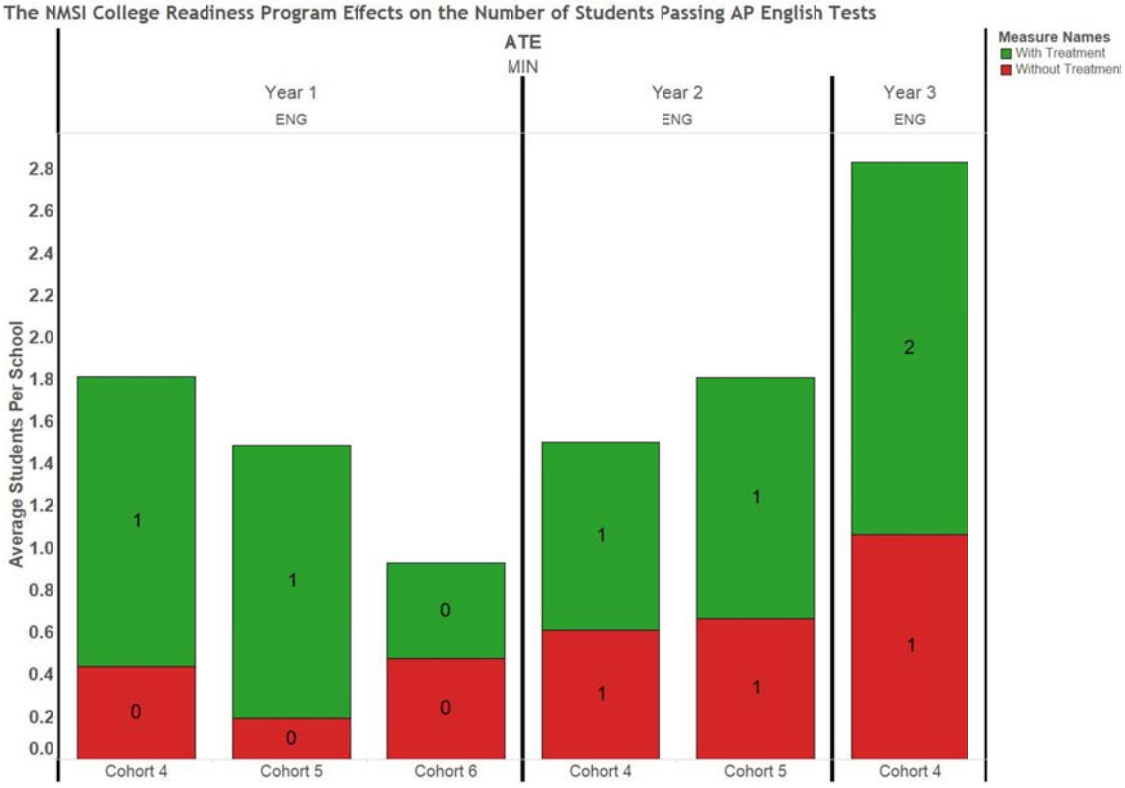
The NMSI College Readiness Program Effects on the Number of Students Passing AP Math or Science Tests



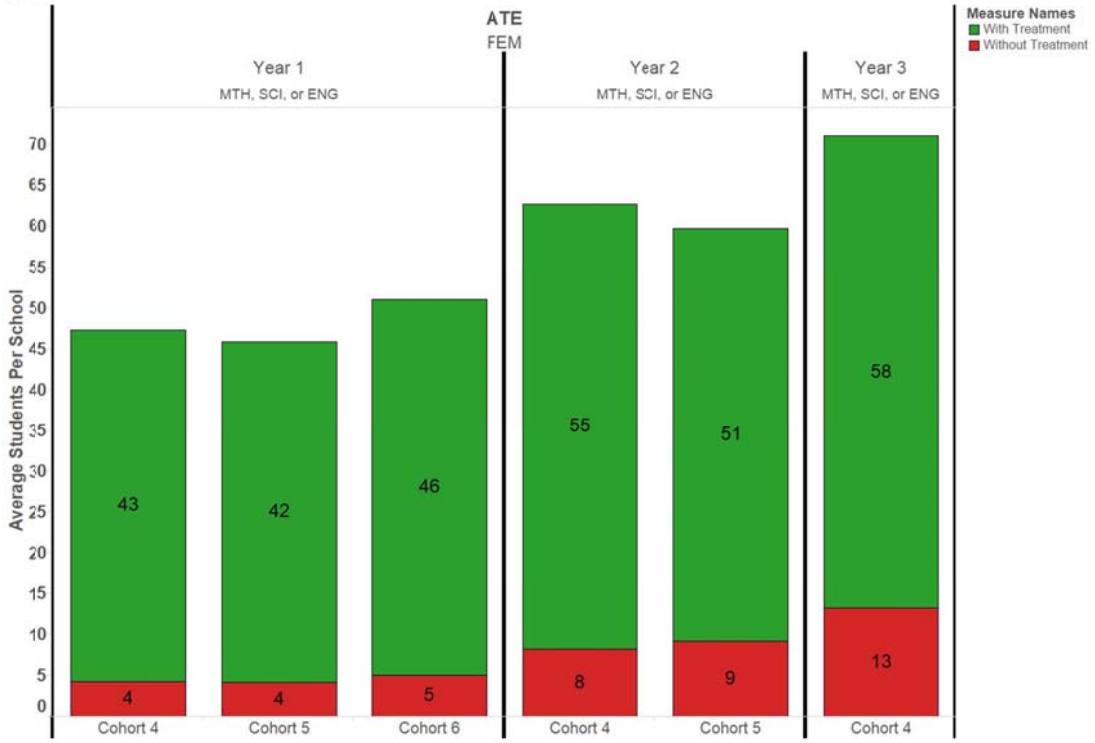
The NMSI College Readiness Program Effects on the Number of Students Taking AP English Tests



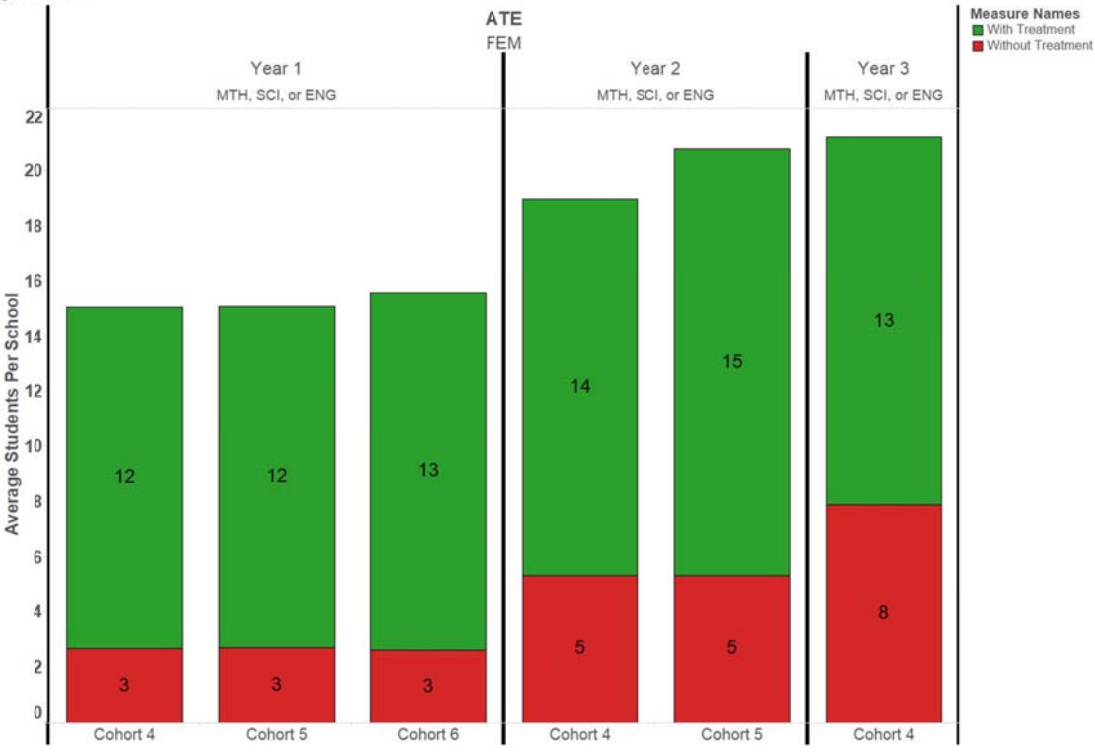
The NMSI College Readiness Program Effects on the Number of Students Passing AP English Tests



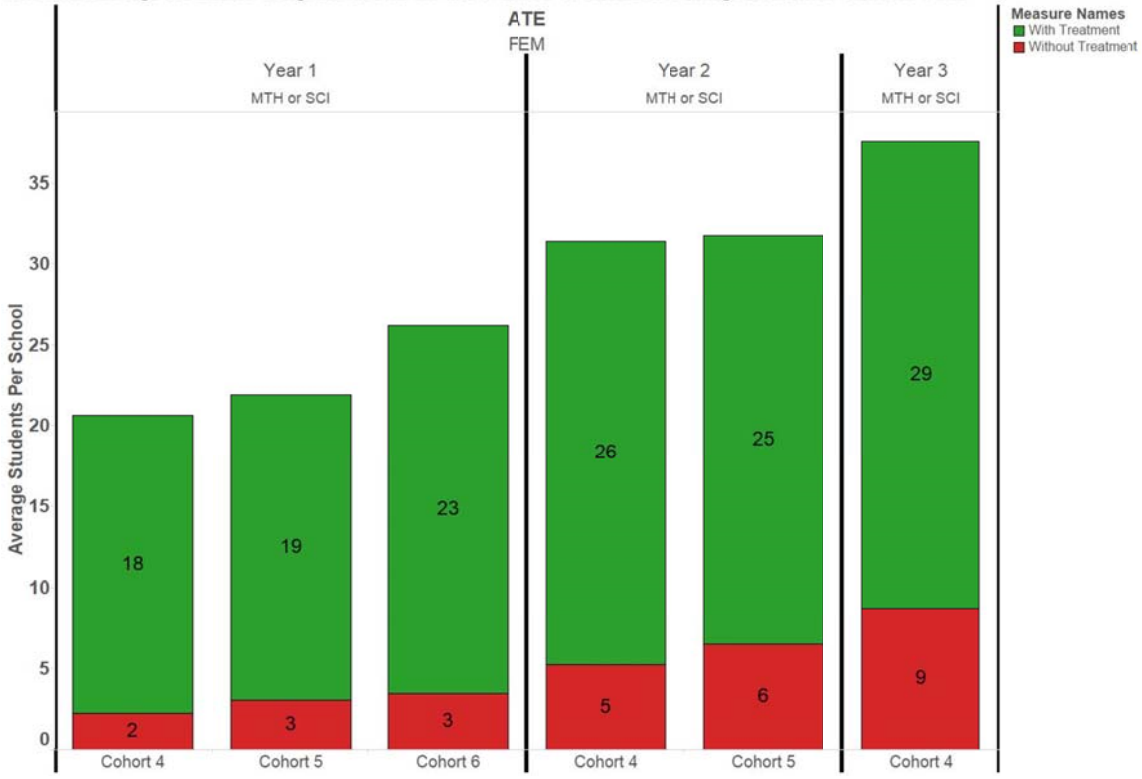
The NMSI College Readiness Program Effects on the Number of Students Taking AP Math, Science, or English Tests



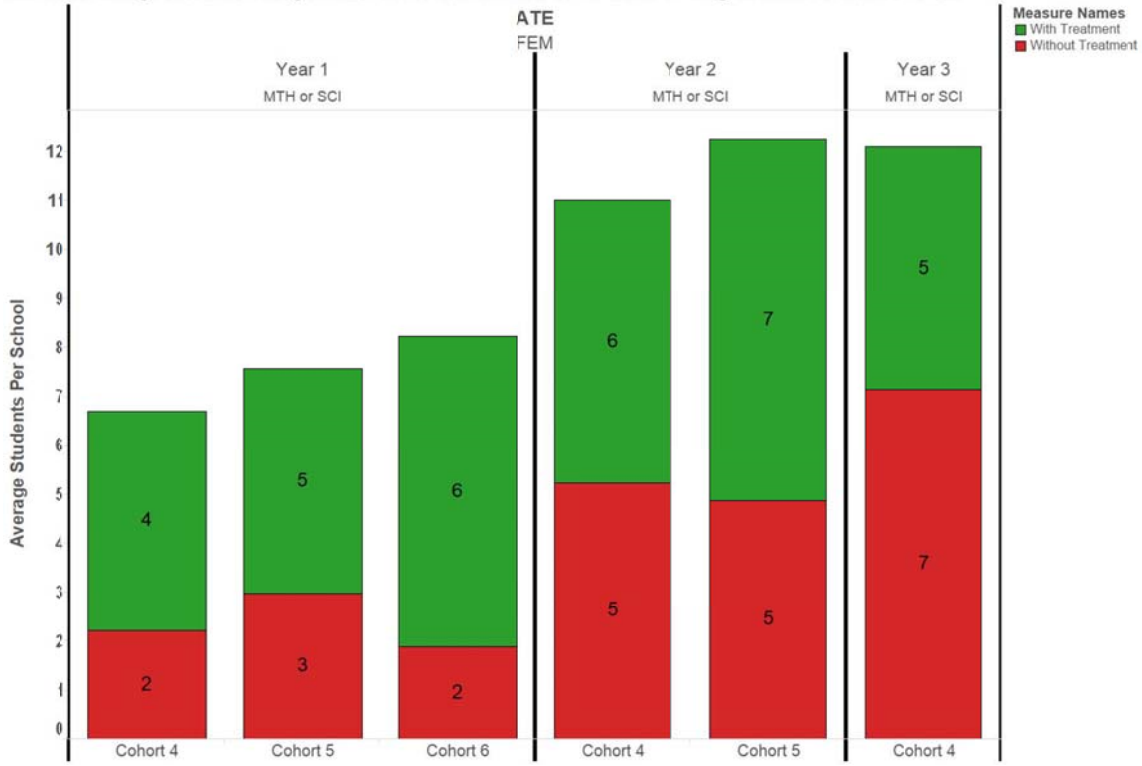
The NMSI College Readiness Program Effects on the Number of Students Passing AP Math, Science, or English Tests



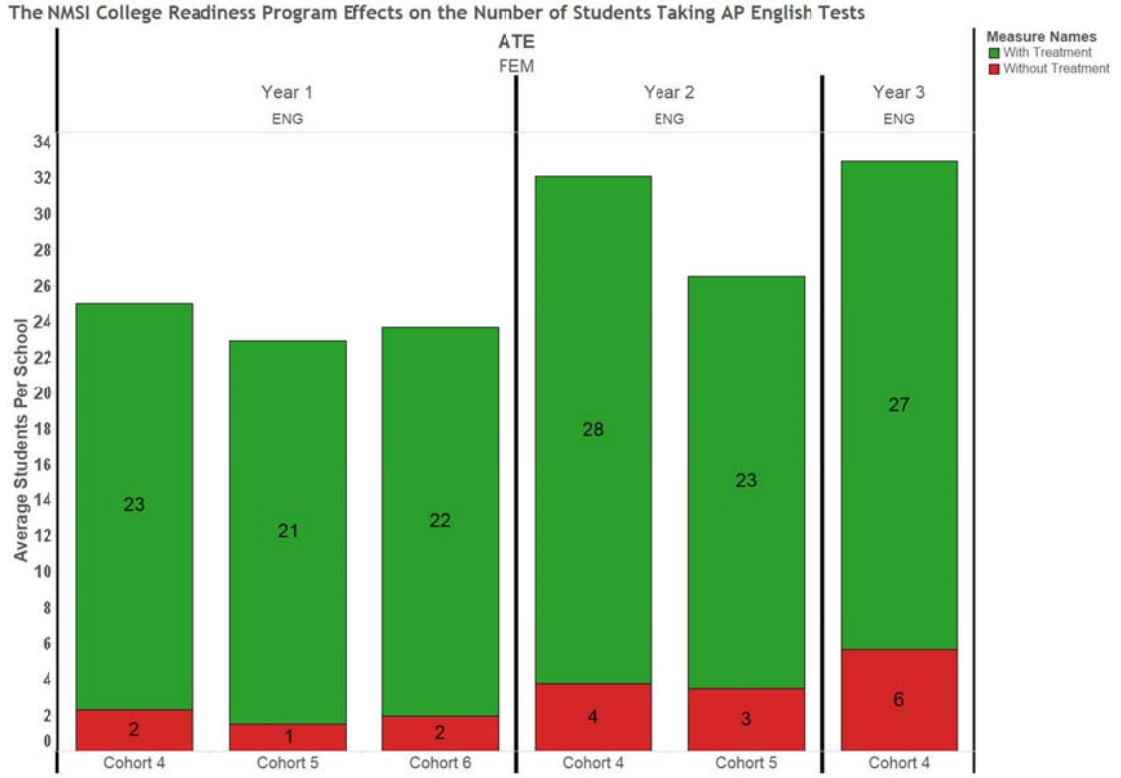
The NMSI College Readiness Program Effects on the Number of Students Taking AP Math or Science Tests



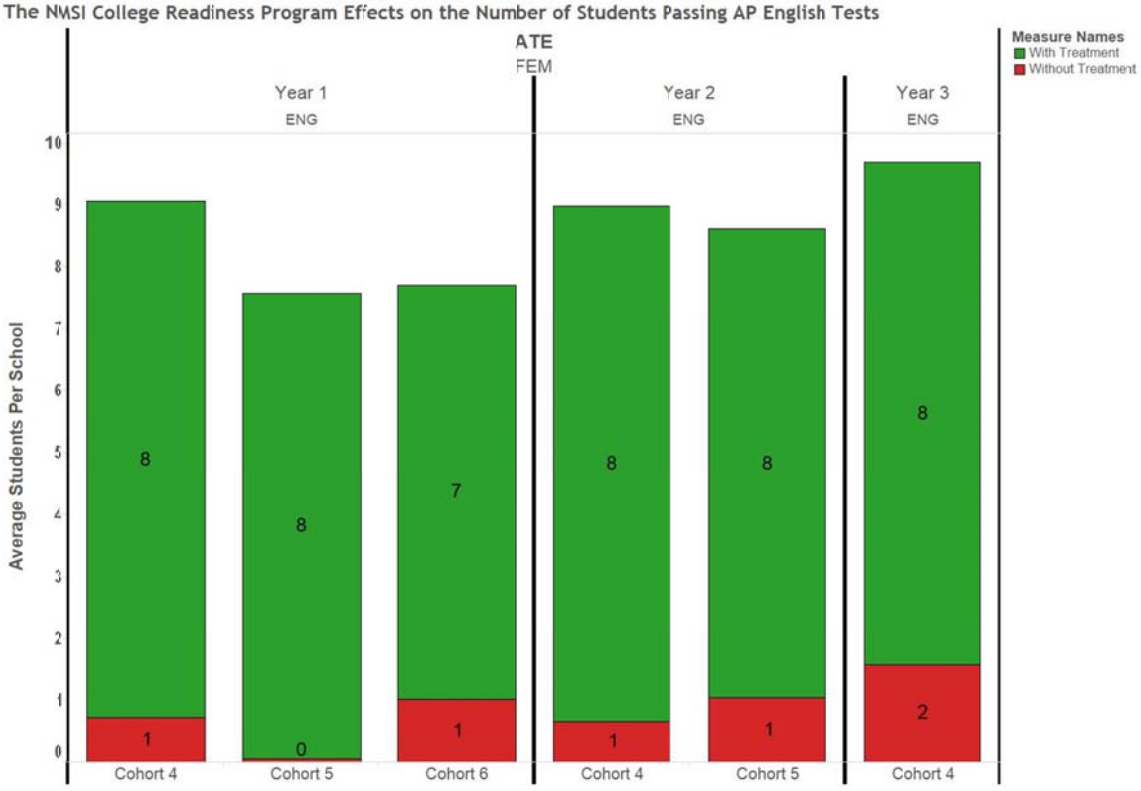
The NMSI College Readiness Program Effects on the Number of Students Passing AP Math or Science Tests



The NMSI College Readiness Program Effects on the Number of Students Taking AP English Tests



The NMSI College Readiness Program Effects on the Number of Students Passing AP English Tests



Appendix B:
Balance Tables

Balance Table—Cohort 4 ATT

```
> bal.table(ps.coh4_y2y)
$unw
                                 tx.mn   tx.sd   ct.mn   ct.sd std.eff.sz   stat     p    ks ks.pval
mse_femaledemo_2011              0.589   0.120   0.562   0.180      0.229  2.268 0.023 0.142   0.034
mse_femaledemo_2011:<NA>         0.065   0.246   0.384   0.486     -0.657 -5.609 0.000 0.319   0.000
mse_minoritydemo_2011            0.247   0.287   0.221   0.298      0.092  0.925 0.355 0.109   0.174
mse_minoritydemo_2011:<NA>       0.065   0.246   0.384   0.486     -0.657 -5.609 0.000 0.319   0.000
mse_allstudents_pass_2011       40.267  50.049  77.545 142.676     -0.745 -7.064 0.000 0.145   0.027
mse_allstudents_pass_2011:<NA>   0.065   0.246   0.384   0.486     -0.657 -5.609 0.000 0.319   0.000
mse_allstudents_total_2011     111.099 105.105 131.522 184.774     -0.194 -1.914 0.056 0.232   0.000
mse_allstudents_total_2011:<NA>  0.065   0.246   0.384   0.486     -0.657 -5.609 0.000 0.319   0.000

$es.mean.ATT
                                 tx.mn   tx.sd   ct.mn   ct.sd std.eff.sz   stat     p    ks ks.pval
mse_femaledemo_2011              0.589   0.120   0.587   0.124      0.015  0.148 0.882 0.032   1.000
mse_femaledemo_2011:<NA>         0.065   0.246   0.068   0.252     -0.007 -35.539 0.000 0.003   0.895
mse_minoritydemo_2011            0.247   0.287   0.248   0.284     -0.001 -0.007 0.995 0.031   1.000
mse_minoritydemo_2011:<NA>       0.065   0.246   0.068   0.252     -0.007 -35.539 0.000 0.003   0.895
mse_allstudents_pass_2011       40.267  50.049  44.693  71.680     -0.088 -0.880 0.379 0.025   1.000
mse_allstudents_pass_2011:<NA>   0.065   0.246   0.068   0.252     -0.007 -35.539 0.000 0.003   0.895
mse_allstudents_total_2011     111.099 105.105 108.479 113.070      0.025  0.247 0.805 0.039   0.997
mse_allstudents_total_2011:<NA>  0.065   0.246   0.068   0.252     -0.007 -35.539 0.000 0.003   0.895

$ks.max.ATT
                                 tx.mn   tx.sd   ct.mn   ct.sd std.eff.sz   stat     p    ks ks.pval
mse_femaledemo_2011              0.589   0.120   0.588   0.125      0.011  0.113 0.910 0.032   1.000
mse_femaledemo_2011:<NA>         0.065   0.246   0.067   0.251     -0.005 -35.539 0.000 0.003   0.917
mse_minoritydemo_2011            0.247   0.287   0.250   0.286     -0.010 -0.103 0.918 0.033   1.000
mse_minoritydemo_2011:<NA>       0.065   0.246   0.067   0.251     -0.005 -35.539 0.000 0.003   0.917
mse_allstudents_pass_2011       40.267  50.049  45.483  73.876     -0.104 -1.038 0.299 0.027   1.000
mse_allstudents_pass_2011:<NA>   0.065   0.246   0.067   0.251     -0.005 -35.539 0.000 0.003   0.917
mse_allstudents_total_2011     111.099 105.105 108.818 114.769      0.022  0.216 0.829 0.039   0.997
mse_allstudents_total_2011:<NA>  0.065   0.246   0.067   0.251     -0.005 -35.539 0.000 0.003   0.917
```

Balance Table—Cohort 4 ATE

```
> bal.table(ps.NMSI4.y2y)
$unw
                                tx.mn   tx.sd   ct.mn   ct.sd std.eff.sz   stat     p    ks ks.pval
mse_femaledemo_2011             0.589   0.120   0.562   0.180      0.152  2.268 0.023 0.142   0.034
mse_femaledemo_2011:<NA>        0.065   0.246   0.384   0.486     -1.295 -5.609 0.000 0.319   0.000
mse_minoritydemo_2011           0.247   0.287   0.221   0.298      0.089  0.925 0.355 0.109   0.174
mse_minoritydemo_2011:<NA>      0.065   0.246   0.384   0.486     -1.295 -5.609 0.000 0.319   0.000
mse_allstudents_pass_2011      40.267  50.049  77.545 142.676     -0.263 -7.064 0.000 0.145   0.027
mse_allstudents_pass_2011:<NA>  0.065   0.246   0.384   0.486     -1.295 -5.609 0.000 0.319   0.000
mse_allstudents_total_2011    111.099 105.105 131.522 184.774     -0.111 -1.914 0.056 0.232   0.000
mse_allstudents_total_2011:<NA> 0.065   0.246   0.384   0.486     -1.295 -5.609 0.000 0.319   0.000

$es.mean.ATE
                                tx.mn   tx.sd   ct.mn   ct.sd std.eff.sz   stat     p    ks ks.pval
mse_femaledemo_2011             0.563   0.150   0.562   0.180      0.006  0.044 0.965 0.086   0.647
mse_femaledemo_2011:<NA>        0.381   0.486   0.381   0.486      0.001 -0.664 0.506 0.000   0.997
mse_minoritydemo_2011           0.194   0.274   0.221   0.298     -0.094 -0.875 0.381 0.082   0.695
mse_minoritydemo_2011:<NA>      0.381   0.486   0.381   0.486      0.001 -0.664 0.506 0.000   0.997
mse_allstudents_pass_2011      44.362  63.763  77.223 142.171     -0.282 -3.547 0.000 0.101   0.442
mse_allstudents_pass_2011:<NA>  0.381   0.486   0.381   0.486      0.001 -0.664 0.506 0.000   0.997
mse_allstudents_total_2011    105.128 124.431 131.321 184.190     -0.162 -1.469 0.142 0.089   0.607
mse_allstudents_total_2011:<NA> 0.381   0.486   0.381   0.486      0.001 -0.664 0.506 0.000   0.997

$ks.max.ATE
                                tx.mn   tx.sd   ct.mn   ct.sd std.eff.sz   stat     p    ks ks.pval
mse_femaledemo_2011             0.564   0.149   0.562   0.180      0.009  0.066 0.947 0.087   0.628
mse_femaledemo_2011:<NA>        0.376   0.484   0.381   0.486     -0.009 -0.830 0.406 0.004   0.964
mse_minoritydemo_2011           0.195   0.274   0.221   0.298     -0.093 -0.864 0.387 0.082   0.705
mse_minoritydemo_2011:<NA>      0.376   0.484   0.381   0.486     -0.009 -0.830 0.406 0.004   0.964
mse_allstudents_pass_2011      44.424  63.743  77.227 142.175     -0.282 -3.550 0.000 0.100   0.449
mse_allstudents_pass_2011:<NA>  0.376   0.484   0.381   0.486     -0.009 -0.830 0.406 0.004   0.964
mse_allstudents_total_2011    105.245 124.241 131.324 184.194     -0.161 -1.471 0.141 0.090   0.593
mse_allstudents_total_2011:<NA> 0.376   0.484   0.381   0.486     -0.009 -0.830 0.406 0.004   0.964
```

Balance Table—Cohort 5 ATT

```
> bal.table(ps.coh5_y2y)
$unw
                                tx.mn  tx.sd   ct.mn   ct.sd std.eff.sz   stat     p    ks ks.pval
mse_femaledemo_2012             0.581  0.132   0.561   0.187      0.148  1.276 0.202 0.173   0.022
mse_femaledemo_2012:<NA>        0.062  0.242   0.309   0.462     -0.534 -4.111 0.000 0.246   0.000
mse_minoritydemo_2012           0.183  0.217   0.224   0.300     -0.191 -1.649 0.099 0.139   0.105
mse_minoritydemo_2012:<NA>      0.062  0.242   0.309   0.462     -0.534 -4.111 0.000 0.246   0.000
mse_allstudents_pass_2012      54.747 55.602  80.520 148.263     -0.464 -3.892 0.000 0.199   0.005
mse_allstudents_pass_2012:<NA>  0.062  0.242   0.309   0.462     -0.534 -4.111 0.000 0.246   0.000
mse_allstudents_total_2012    127.960 95.924 133.273 189.158     -0.055 -0.473 0.636 0.259   0.000
mse_allstudents_total_2012:<NA> 0.062  0.242   0.309   0.462     -0.534 -4.111 0.000 0.246   0.000

$es.mean.ATT
                                tx.mn  tx.sd   ct.mn   ct.sd std.eff.sz   stat     p    ks ks.pval
mse_femaledemo_2012             0.581  0.132   0.579   0.125      0.011  0.097 0.922 0.026   1.000
mse_femaledemo_2012:<NA>        0.062  0.242   0.071   0.258     -0.019 -29.075 0.000 0.009   0.757
mse_minoritydemo_2012           0.183  0.217   0.183   0.220      0.000  -0.004 0.997 0.046   0.997
mse_minoritydemo_2012:<NA>      0.062  0.242   0.071   0.258     -0.019 -29.075 0.000 0.009   0.757
mse_allstudents_pass_2012      54.747 55.602  55.959  68.887     -0.022  -0.186 0.852 0.033   1.000
mse_allstudents_pass_2012:<NA>  0.062  0.242   0.071   0.258     -0.019 -29.075 0.000 0.009   0.757
mse_allstudents_total_2012    127.960 95.924 123.636 106.622      0.045   0.382 0.703 0.055   0.974
mse_allstudents_total_2012:<NA> 0.062  0.242   0.071   0.258     -0.019 -29.075 0.000 0.009   0.757

$ks.max.ATT
                                tx.mn  tx.sd   ct.mn   ct.sd std.eff.sz   stat     p    ks ks.pval
mse_femaledemo_2012             0.581  0.132   0.579   0.125      0.016  0.137 0.891 0.036   1.000
mse_femaledemo_2012:<NA>        0.062  0.242   0.065   0.247     -0.006 -29.075 0.000 0.003   0.921
mse_minoritydemo_2012           0.183  0.217   0.193   0.232     -0.049  -0.423 0.672 0.044   0.998
mse_minoritydemo_2012:<NA>      0.062  0.242   0.065   0.247     -0.006 -29.075 0.000 0.003   0.921
mse_allstudents_pass_2012      54.747 55.602  60.977  79.332     -0.112  -0.962 0.336 0.033   1.000
mse_allstudents_pass_2012:<NA>  0.062  0.242   0.065   0.247     -0.006 -29.075 0.000 0.003   0.921
mse_allstudents_total_2012    127.960 95.924 127.386 115.584      0.006   0.051 0.959 0.046   0.997
mse_allstudents_total_2012:<NA> 0.062  0.242   0.065   0.247     -0.006 -29.075 0.000 0.003   0.921
```

Balance Table—Cohort 5 ATE

```
> bal.table(ps.NMSI5.y2y)
$unw
                                 tx.mn  tx.sd   ct.mn   ct.sd std.eff.sz   stat     p    ks ks.pval
mse_femaledemo_2012              0.581  0.132   0.561   0.187       0.105  1.276 0.202 0.173   0.022
mse_femaledemo_2012:<NA>         0.062  0.242   0.309   0.462      -1.016 -4.111 0.000 0.246   0.000
mse_minoritydemo_2012            0.183  0.217   0.224   0.300      -0.139 -1.649 0.099 0.139   0.105
mse_minoritydemo_2012:<NA>       0.062  0.242   0.309   0.462      -1.016 -4.111 0.000 0.246   0.000
mse_allstudents_pass_2012       54.747 55.602  80.520 148.263      -0.175 -3.892 0.000 0.199   0.005
mse_allstudents_pass_2012:<NA>   0.062  0.242   0.309   0.462      -1.016 -4.111 0.000 0.246   0.000
mse_allstudents_total_2012     127.960 95.924 133.273 189.158      -0.028 -0.473 0.636 0.259   0.000
mse_allstudents_total_2012:<NA>  0.062  0.242   0.309   0.462      -1.016 -4.111 0.000 0.246   0.000

$es.mean.ATE
                                 tx.mn  tx.sd   ct.mn   ct.sd std.eff.sz   stat     p    ks ks.pval
mse_femaledemo_2012              0.581  0.190   0.562   0.187       0.102  0.532 0.594 0.101   0.634
mse_femaledemo_2012:<NA>         0.306  0.461   0.307   0.461      -0.001 -0.591 0.554 0.000   0.996
mse_minoritydemo_2012            0.167  0.244   0.224   0.300      -0.205 -1.624 0.104 0.107   0.553
mse_minoritydemo_2012:<NA>       0.306  0.461   0.307   0.461      -0.001 -0.591 0.554 0.000   0.996
mse_allstudents_pass_2012       51.607  64.092  80.422 147.945      -0.237 -2.970 0.003 0.136   0.267
mse_allstudents_pass_2012:<NA>   0.306  0.461   0.307   0.461      -0.001 -0.591 0.554 0.000   0.996
mse_allstudents_total_2012     110.386 101.840 133.265 188.789      -0.143 -1.574 0.115 0.135   0.279
mse_allstudents_total_2012:<NA>  0.306  0.461   0.307   0.461      -0.001 -0.591 0.554 0.000   0.996

$ks.max.ATE
                                 tx.mn  tx.sd   ct.mn   ct.sd std.eff.sz   stat     p    ks ks.pval
mse_femaledemo_2012              0.585  0.204   0.562   0.186       0.124  0.561 0.575 0.115   0.582
mse_femaledemo_2012:<NA>         0.392  0.488   0.307   0.461       0.175 -0.228 0.820 0.085   0.424
mse_minoritydemo_2012            0.168  0.258   0.224   0.300      -0.197 -1.376 0.169 0.115   0.582
mse_minoritydemo_2012:<NA>       0.392  0.488   0.307   0.461       0.175 -0.228 0.820 0.085   0.424
mse_allstudents_pass_2012       47.400 63.217  80.383 147.844      -0.267 -3.180 0.001 0.110   0.642
mse_allstudents_pass_2012:<NA>   0.392  0.488   0.307   0.461       0.175 -0.228 0.820 0.085   0.424
mse_allstudents_total_2012     100.117 95.957 133.242 188.676      -0.207 -2.245 0.025 0.105   0.694
mse_allstudents_total_2012:<NA>  0.392  0.488   0.307   0.461       0.175 -0.228 0.820 0.085   0.424
```

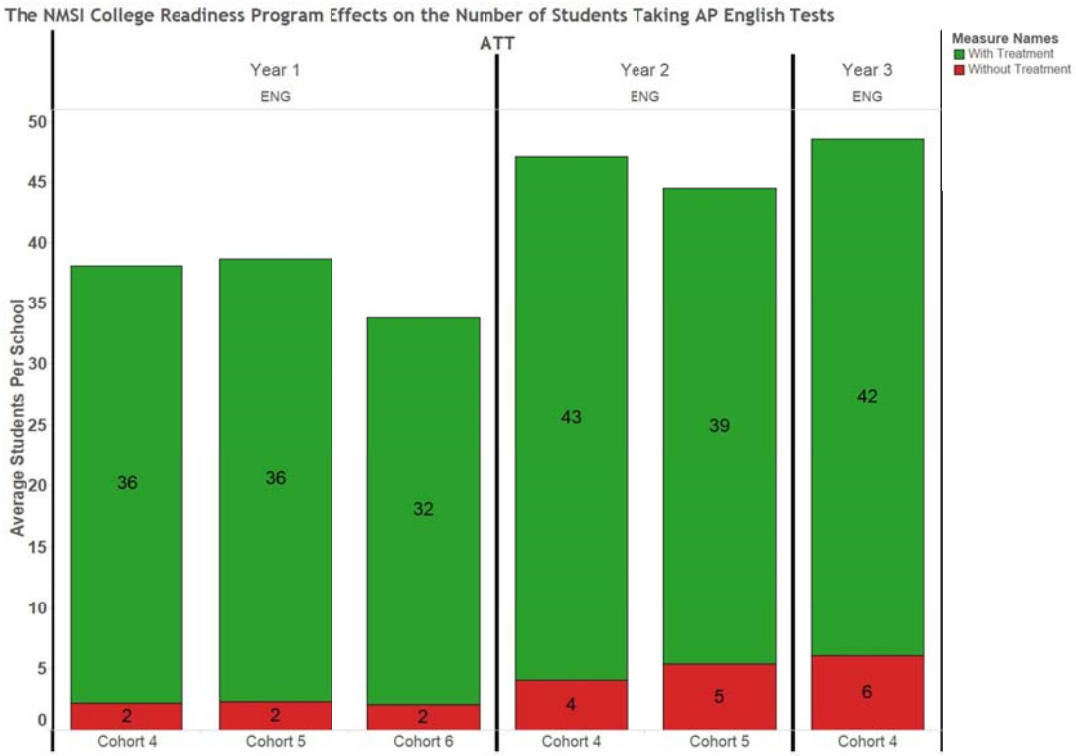
Balance Table—Cohort 6 ATT

```
> bal.table(ps.COH6)
$unw
                                 tx.mn  tx.sd   ct.mn   ct.sd std.eff.sz   stat     p    ks ks.pval
mse_allstudents_total_2013     165.186 173.264 139.857 198.888       0.146  1.435 0.151 0.258   0.000
mse_allstudents_total_2013:<NA>  0.020   0.141   0.246   0.430      -0.525 -3.860 0.000 0.225   0.000
mse_allstudents_pass_2013       74.072 119.267  84.089 155.170      -0.084 -0.822 0.411 0.121   0.110
mse_allstudents_pass_2013:<NA>   0.020   0.141   0.246   0.430      -0.525 -3.860 0.000 0.225   0.000
mse_minoritydemo_2013            0.303   0.297   0.239   0.305       0.214  2.103 0.036 0.257   0.000
mse_minoritydemo_2013:<NA>       0.020   0.141   0.246   0.430      -0.525 -3.860 0.000 0.225   0.000
mse_femaledemo_2013              0.568   0.107   0.554   0.186       0.130  1.265 0.206 0.116   0.142
mse_femaledemo_2013:<NA>         0.020   0.141   0.246   0.430      -0.525 -3.860 0.000 0.225   0.000

$es.mean.ATT
                                 tx.mn  tx.sd   ct.mn   ct.sd std.eff.sz   stat     p    ks ks.pval
mse_allstudents_total_2013     165.186 173.264 159.064 173.305       0.035  0.342 0.732 0.049   0.970
mse_allstudents_total_2013:<NA>  0.020   0.141   0.021   0.144      -0.003 -0.078 0.938 0.001   0.938
mse_allstudents_pass_2013       74.072 119.267  74.102 120.741       0.000 -0.002 0.998 0.030   1.000
mse_allstudents_pass_2013:<NA>   0.020   0.141   0.021   0.144      -0.003 -0.078 0.938 0.001   0.938
mse_minoritydemo_2013            0.303   0.297   0.303   0.303      -0.001 -0.012 0.990 0.042   0.995
mse_minoritydemo_2013:<NA>       0.020   0.141   0.021   0.144      -0.003 -0.078 0.938 0.001   0.938
mse_femaledemo_2013              0.568   0.107   0.568   0.124      -0.005 -0.050 0.960 0.039   0.998
mse_femaledemo_2013:<NA>         0.020   0.141   0.021   0.144      -0.003 -0.078 0.938 0.001   0.938

$ks.max.ATT
                                 tx.mn  tx.sd   ct.mn   ct.sd std.eff.sz   stat     p    ks ks.pval
mse_allstudents_total_2013     165.186 173.264 154.527 166.200       0.062  0.594 0.553 0.047   0.983
mse_allstudents_total_2013:<NA>  0.020   0.141   0.022   0.147      -0.005 -0.134 0.894 0.002   0.894
mse_allstudents_pass_2013       74.072 119.267  69.552 112.585       0.038  0.368 0.713 0.029   1.000
mse_allstudents_pass_2013:<NA>   0.020   0.141   0.022   0.147      -0.005 -0.134 0.894 0.002   0.894
mse_minoritydemo_2013            0.303   0.297   0.297   0.297       0.020  0.185 0.853 0.042   0.995
mse_minoritydemo_2013:<NA>       0.020   0.141   0.022   0.147      -0.005 -0.134 0.894 0.002   0.894
mse_femaledemo_2013              0.568   0.107   0.569   0.118      -0.010 -0.099 0.921 0.036   0.999
mse_femaledemo_2013:<NA>         0.020   0.141   0.022   0.147      -0.005 -0.134 0.894 0.002   0.894
```

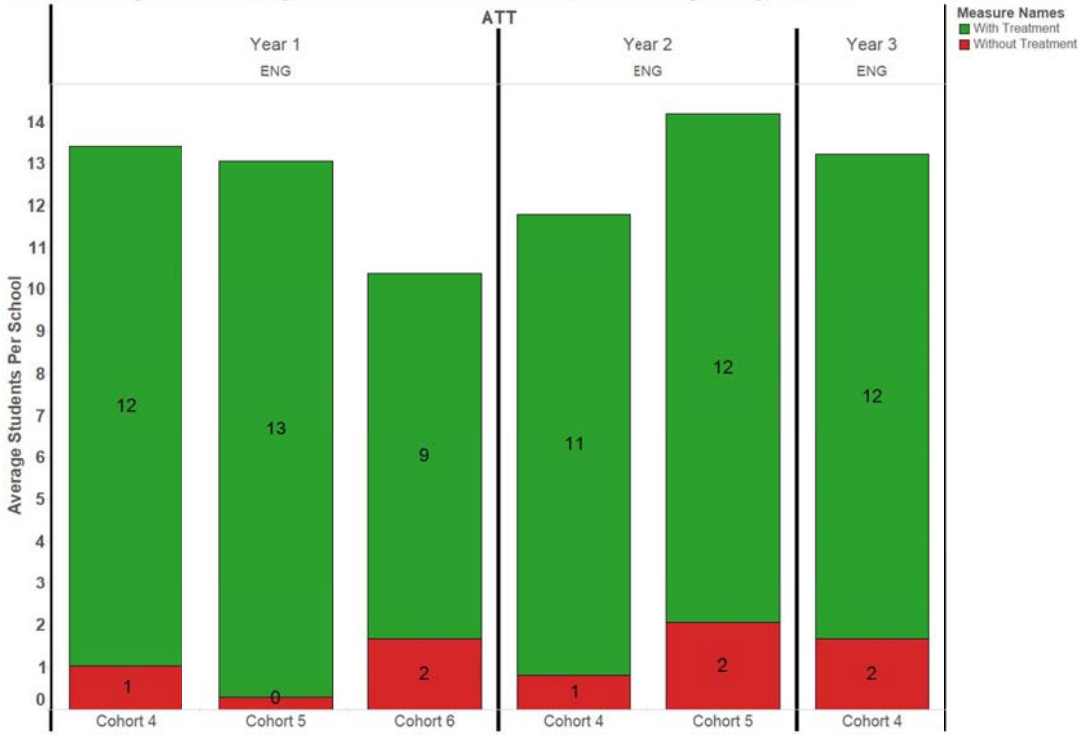
Balance Table—Cohort 6 ATE

```
>
> bal.table(ps.COH6ATE)
$unw
                                tx.mn   tx.sd   ct.mn   ct.sd std.eff.sz   stat     p    ks ks.pval
mse_allstudents_total_2013    165.186 173.264 139.857 198.888      0.128  1.435 0.151 0.258   0.000
mse_allstudents_total_2013:<NA>  0.020   0.141   0.246   0.430     -1.602 -3.860 0.000 0.225   0.000
mse_allstudents_pass_2013      74.072 119.267  84.089 155.170     -0.065 -0.822 0.411 0.121   0.110
mse_allstudents_pass_2013:<NA>   0.020   0.141   0.246   0.430     -1.602 -3.860 0.000 0.225   0.000
mse_minoritydemo_2013            0.303   0.297   0.239   0.305      0.208  2.103 0.036 0.257   0.000
mse_minoritydemo_2013:<NA>       0.020   0.141   0.246   0.430     -1.602 -3.860 0.000 0.225   0.000
mse_femaledemo_2013              0.568   0.107   0.554   0.186      0.075  1.265 0.206 0.116   0.142
mse_femaledemo_2013:<NA>         0.020   0.141   0.246   0.430     -1.602 -3.860 0.000 0.225   0.000

$es.mean.ATE
                                tx.mn   tx.sd   ct.mn   ct.sd std.eff.sz   stat     p    ks ks.pval
mse_allstudents_total_2013    132.258 175.555 140.166 198.831     -0.042 -0.393 0.695 0.068   0.926
mse_allstudents_total_2013:<NA>  0.243   0.429   0.243   0.429      0.000 -1.570 0.116 0.000   0.999
mse_allstudents_pass_2013      64.849 115.895  84.121 155.067     -0.138 -1.495 0.135 0.083   0.770
mse_allstudents_pass_2013:<NA>   0.243   0.429   0.243   0.429      0.000 -1.570 0.116 0.000   0.999
mse_minoritydemo_2013            0.228   0.289   0.240   0.306     -0.039 -0.315 0.753 0.060   0.976
mse_minoritydemo_2013:<NA>       0.243   0.429   0.243   0.429      0.000 -1.570 0.116 0.000   0.999
mse_femaledemo_2013              0.570   0.116   0.554   0.186      0.098  0.955 0.340 0.092   0.658
mse_femaledemo_2013:<NA>         0.243   0.429   0.243   0.429      0.000 -1.570 0.116 0.000   0.999

$ks.max.ATE
                                tx.mn   tx.sd   ct.mn   ct.sd std.eff.sz   stat     p    ks ks.pval
mse_allstudents_total_2013    127.800 172.660 140.182 198.845     -0.066 -0.602 0.547 0.068   0.942
mse_allstudents_total_2013:<NA>  0.293   0.455   0.243   0.429      0.108 -0.591 0.555 0.049   0.727
mse_allstudents_pass_2013      62.981 112.310  84.113 155.056     -0.152 -1.606 0.108 0.090   0.728
mse_allstudents_pass_2013:<NA>   0.293   0.455   0.243   0.429      0.108 -0.591 0.555 0.049   0.727
mse_minoritydemo_2013            0.223   0.289   0.240   0.306     -0.057 -0.433 0.665 0.058   0.986
mse_minoritydemo_2013:<NA>       0.293   0.455   0.243   0.429      0.108 -0.591 0.555 0.049   0.727
mse_femaledemo_2013              0.568   0.115   0.554   0.186      0.086  0.827 0.408 0.092   0.699
mse_femaledemo_2013:<NA>         0.293   0.455   0.243   0.429      0.108 -0.591 0.555 0.049   0.727
```

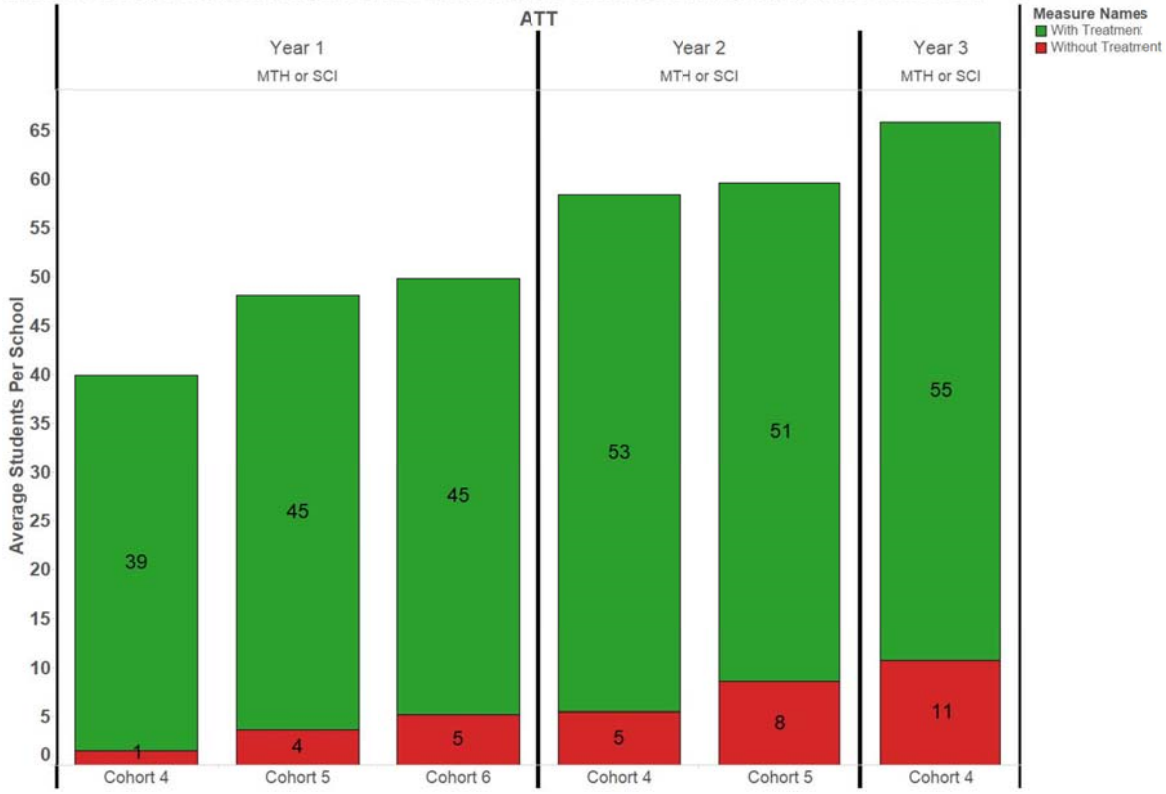
Appendix C:
ATT Analysis Subject Subgroups

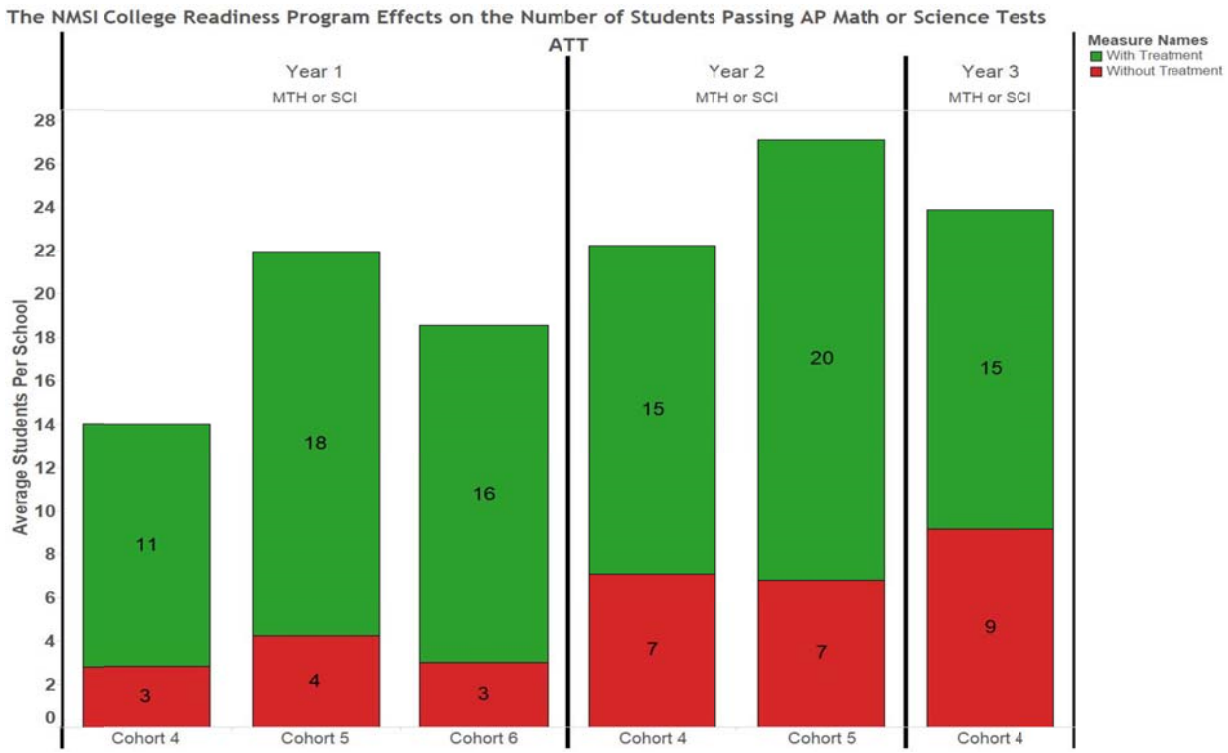


The NMSI College Readiness Program Effects on the Number of Students Passing AP English Tests

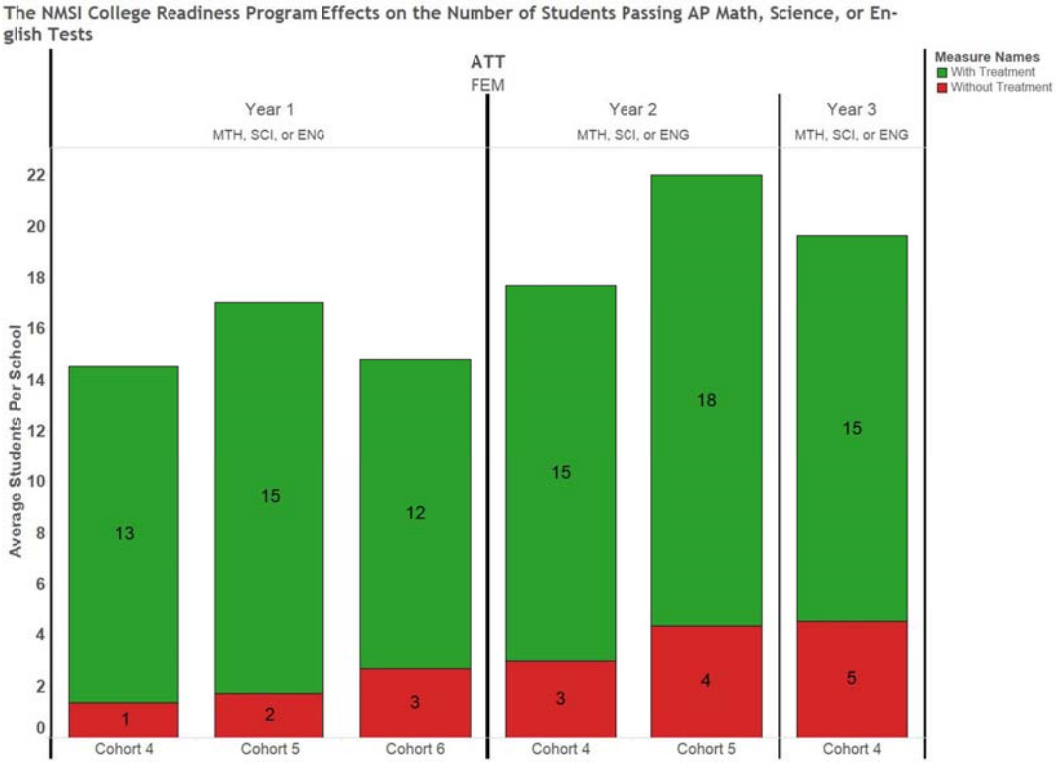


The NMSI College Readiness Program Effects on the Number of Students Taking AP Math or Science Tests

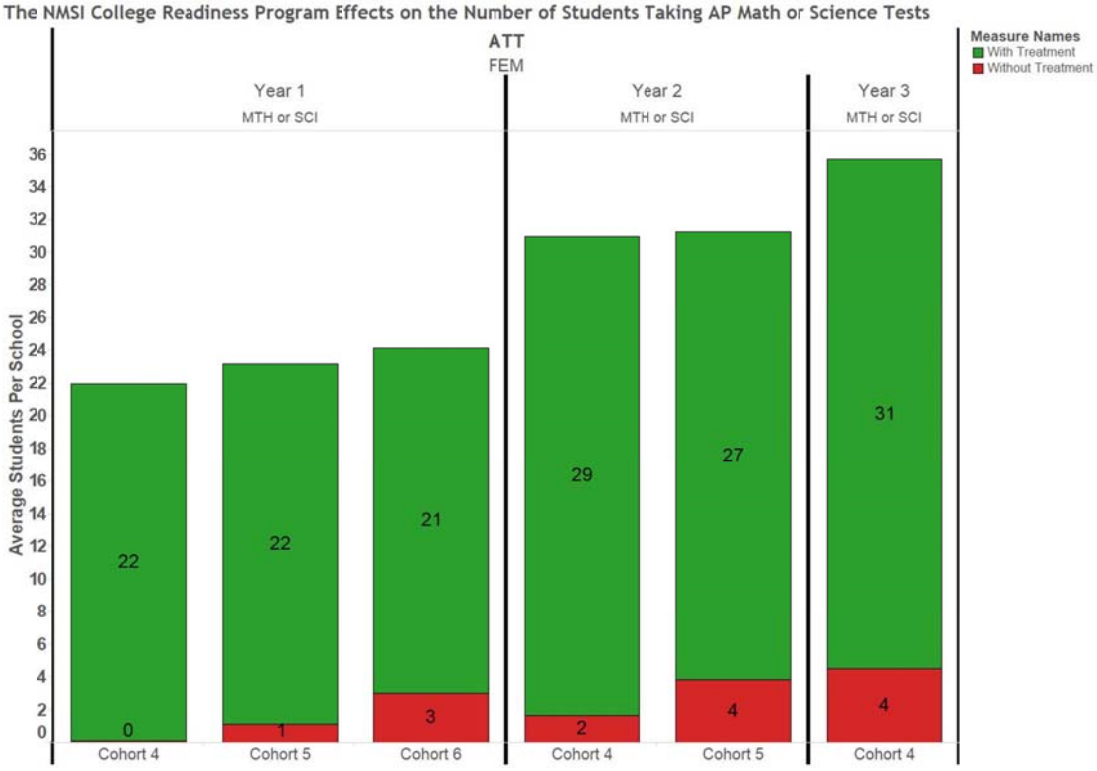



The NMSI College Readiness Program Effects on the Number of Students Passing AP Math or Science Tests
ATT
Measure Names
With Treatment
Without Treatment
Year 1
MTH or SCI
Year 2
MTH or SCI
Year 3
MTH or SCI
Average Students Per School
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
11
3
18
4
16
3
15
7
20
7
15
9
Cohort 4
Cohort 5
Cohort 6
Cohort 4
Cohort 5
Cohort 4

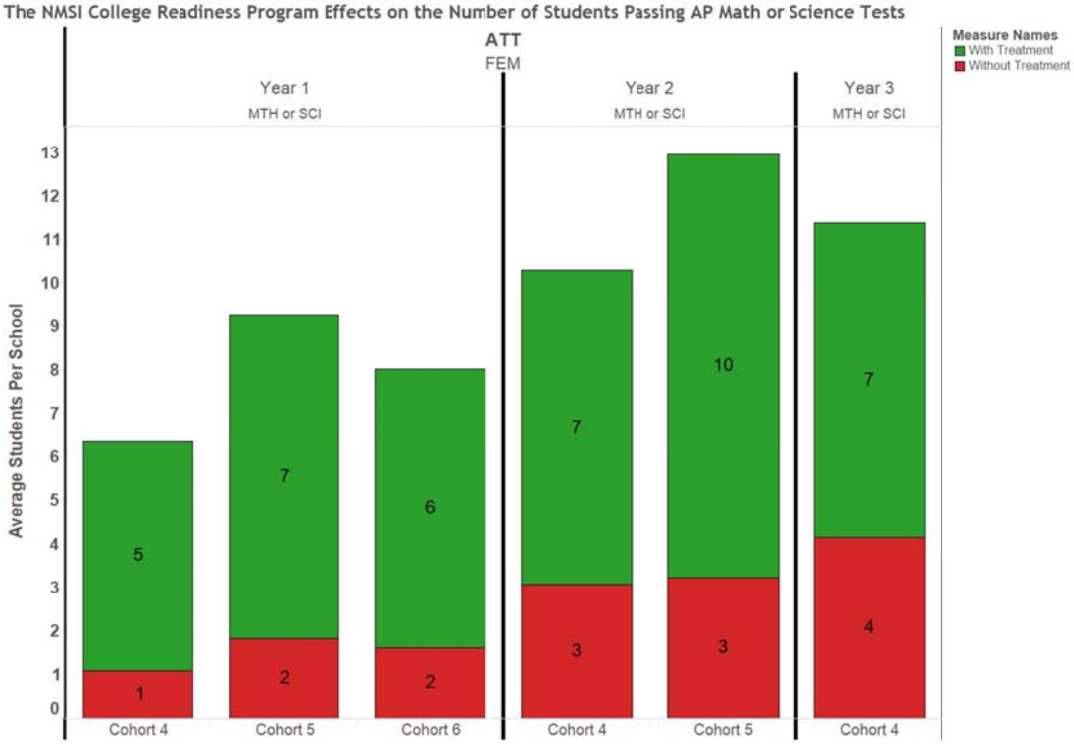
Appendix D:
Female Sample Only



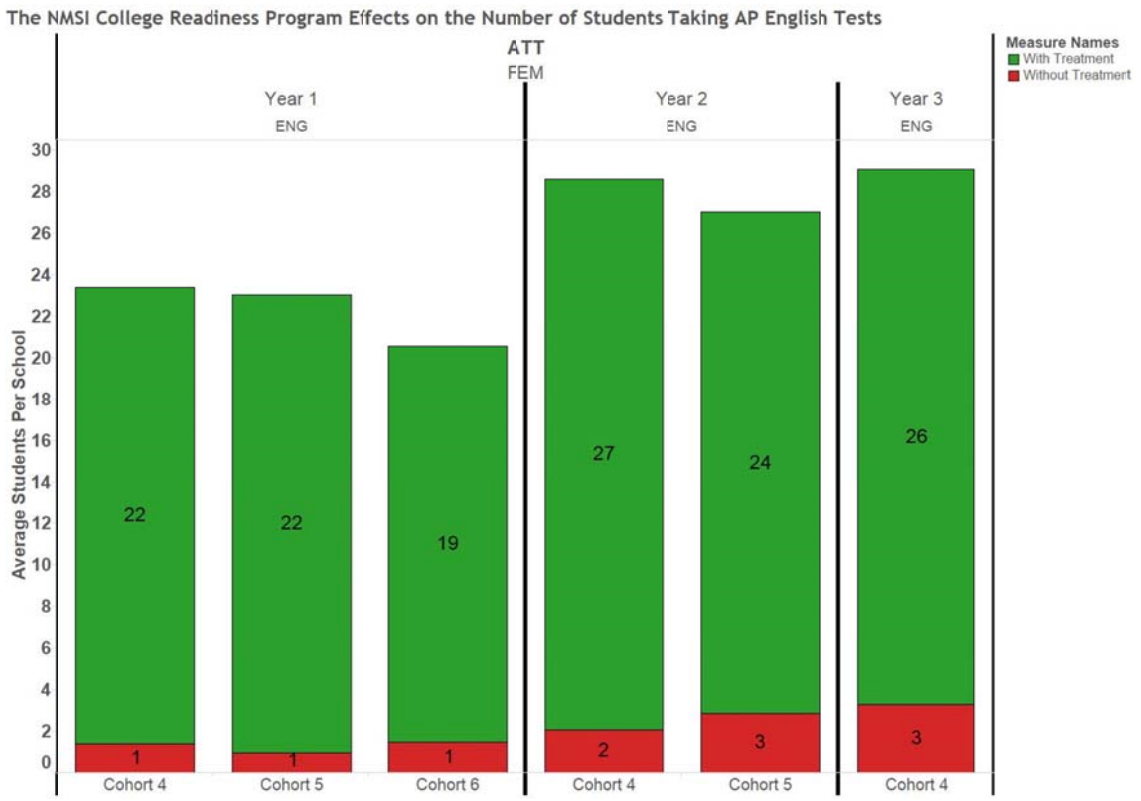
The NMSI College Readiness Program Effects on the Number of Students Taking AP Math or Science Tests



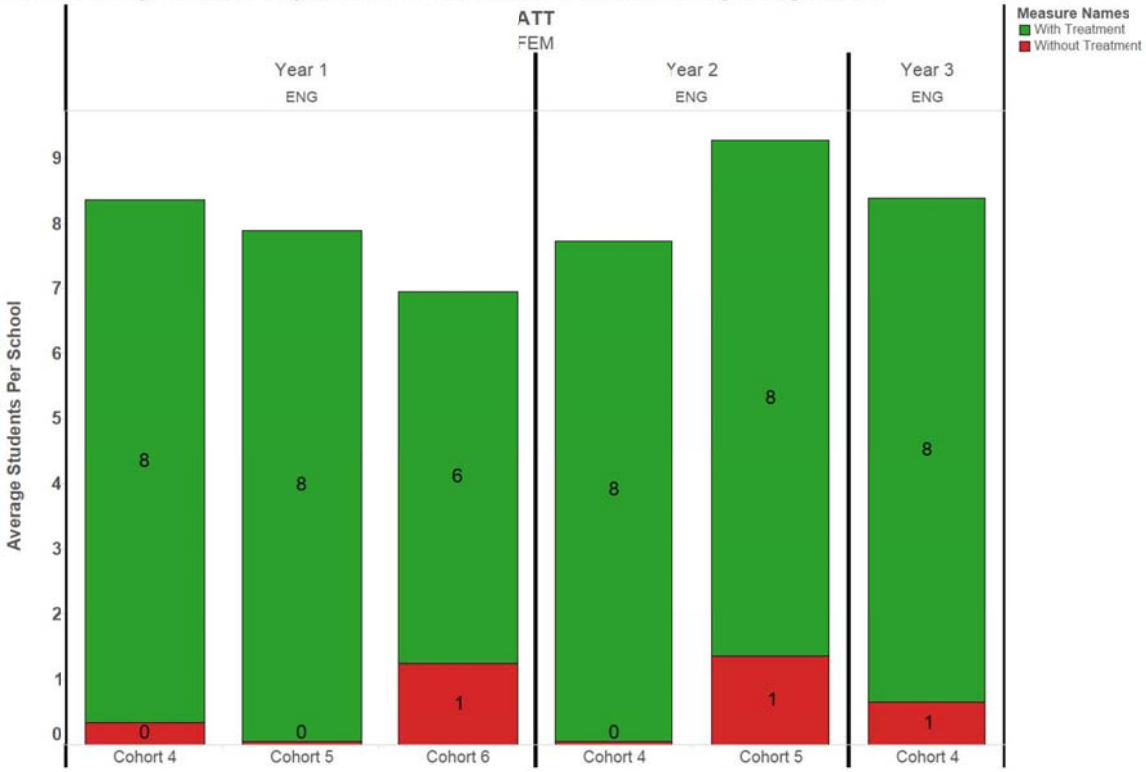
The NMSI College Readiness Program Effects on the Number of Students Passing AP Math or Science Tests



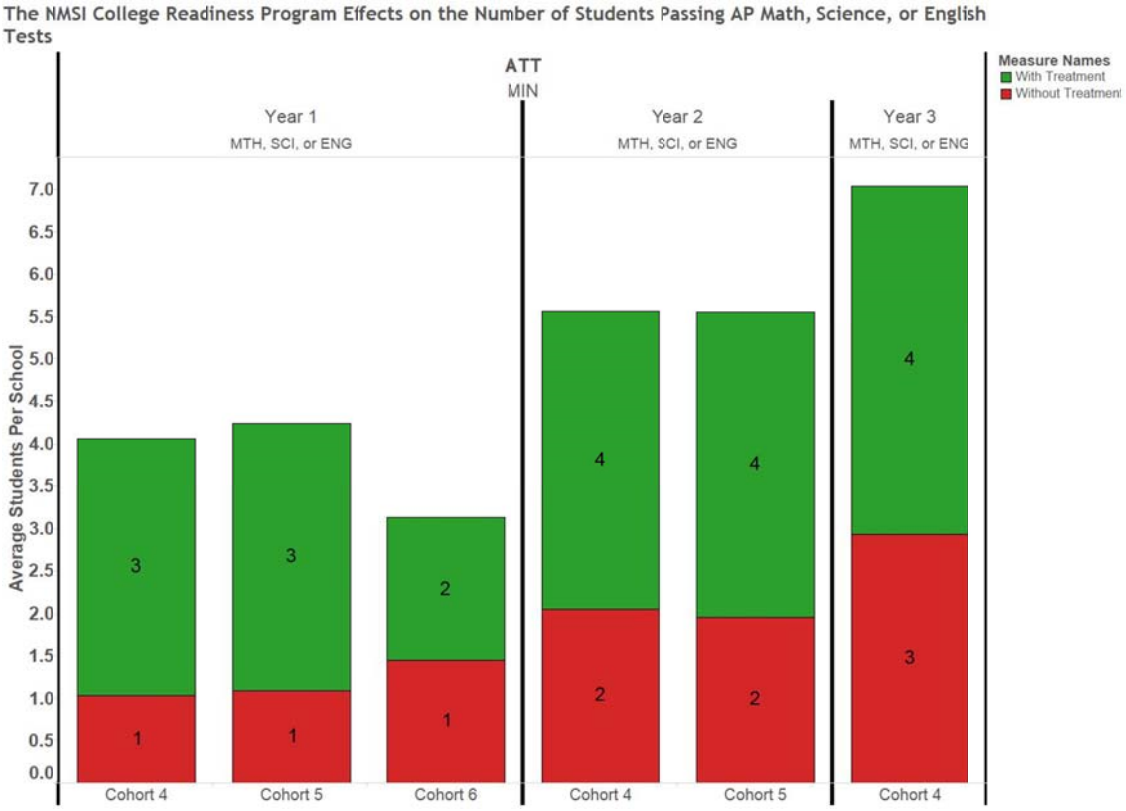
The NMSI College Readiness Program Effects on the Number of Students Taking AP English Tests



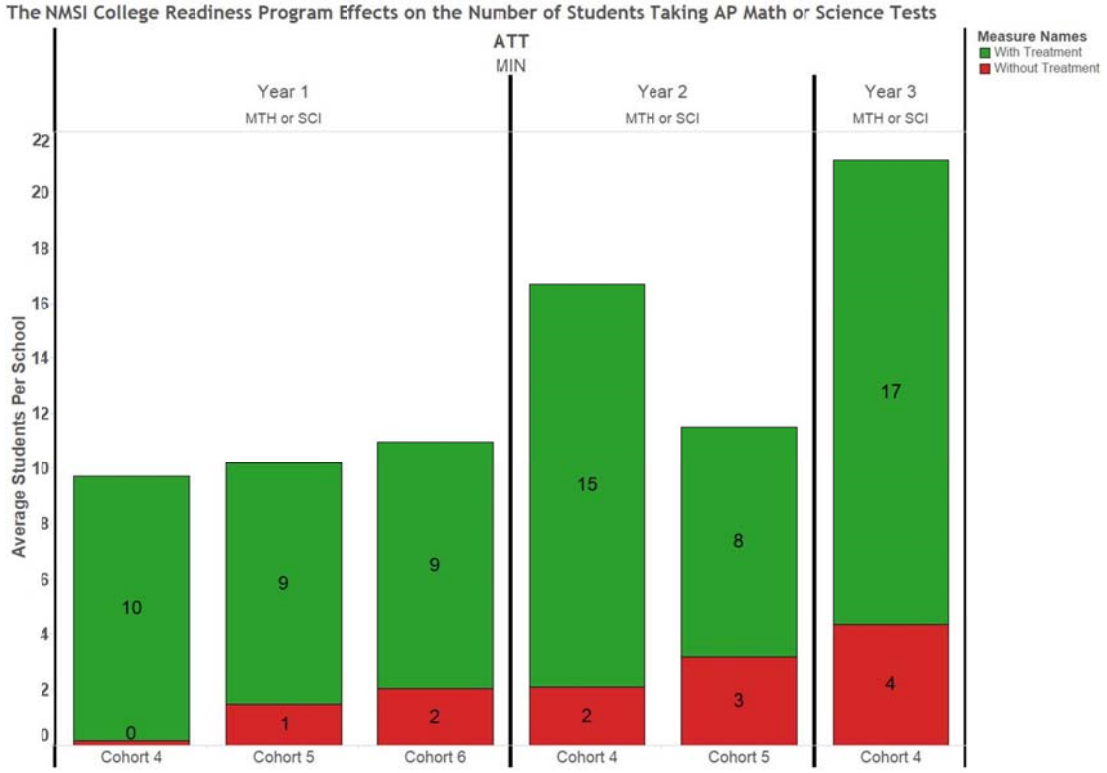
The NMSI College Readiness Program Effects on the Number of Students Passing AP English Tests



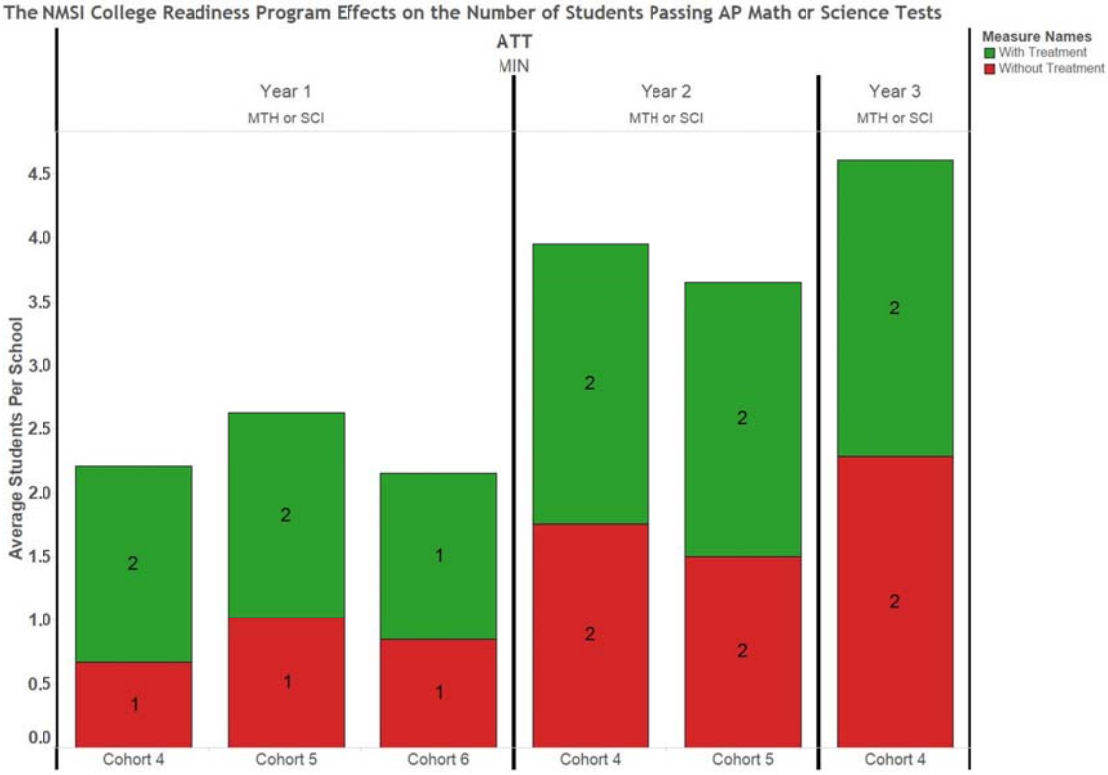
Appendix E:
Minority Sample Only



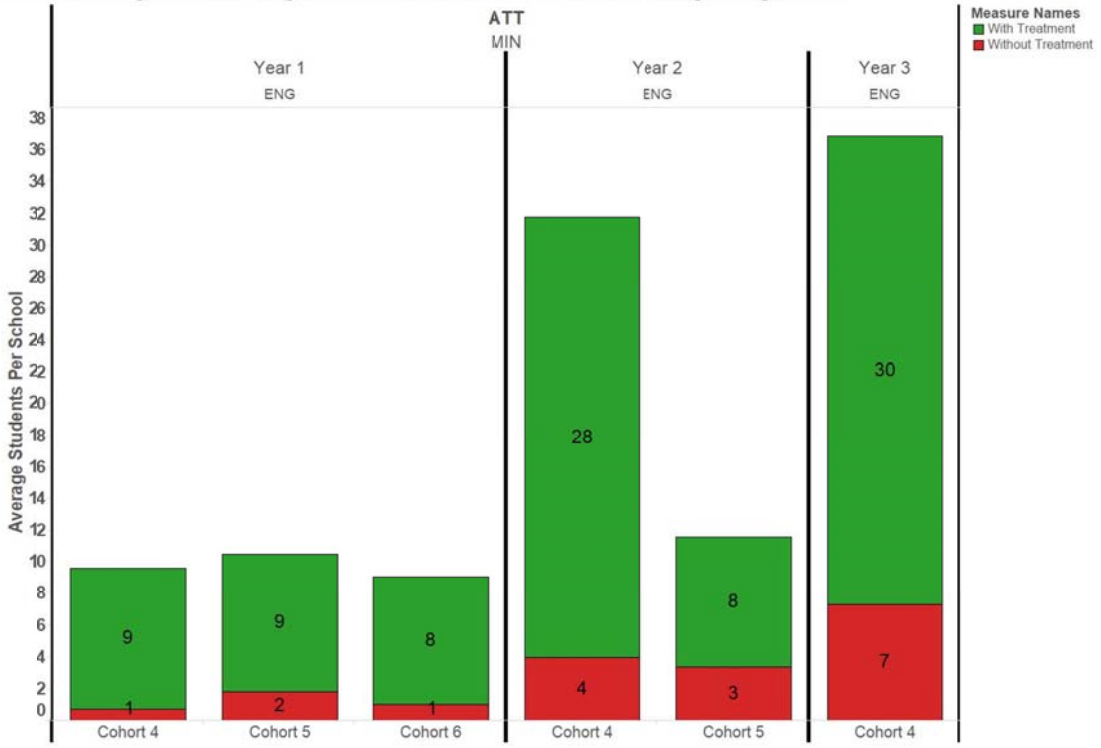
The NMSI College Readiness Program Effects on the Number of Students Taking AP Math or Science Tests



The NMSI College Readiness Program Effects on the Number of Students Passing AP Math or Science Tests



The NMSI College Readiness Program Effects on the Number of Students Taking AP English Tests



The NMSI College Readiness Program Effects on the Number of Students Passing AP English Tests