#

#

#

RESEARCH WHITE PAPER #16-01

The Effectiveness of The National Math + Science Initiative's College Readiness Program

Summary of Independent Research

DATE: May 2, 2016

#

INTRODUCTION

A growing body of evidence indicates that National Math + Science Initiative's College Readiness Program (CRP), previously known as the Advanced Placement Training and Incentive Program (APTIP) or the Advanced Placement Incentive Program (APIP), not only increases the probability that students will take and earn qualifying scores on Advanced Placement exams, thereby increasing their achievement and college readiness, but also has significant positive postsecondary and economic impacts. **The program produces successful and sustained outcomes across settings, states, subject areas, and students, including students traditionally underrepresented in STEM subject areas**. Across studies, CRP is associated with increased percentages of high school students taking AP exams and increased percentages of students earning qualifying scores on these exams, particularly in math, science, and English. This report will briefly review a set of four research efforts that demonstrate these effects.

Holtzman (2010)

In this study, Holtzman (2010) found that CRP had a positive and statistically significant first-year impact on student enrollment in AP courses in math, science, and English and on students' success on related AP exams, as measured by exam scores of 3 or higher. Using a comparative interrupted time series (CITS) design, Holtzman matched 64 program schools with 128 other schools within the program site states that were equivalent, without any statistical adjustments, on pre-treatment values for each of the three pre-implementation years and were also equivalent in enrollment, percentage urban, and percentage rural. Selecting two comparison schools per program school, the nearest above and nearest below neighbors on a composite value, enhanced power for the analysis and the balance between the comparison and program schools on the pre-implementation outcomes.

Fixed-effects regressions showed that in all five of the subject area combinations, implementation of CRP was associated with large and statistically significant increases in the percentages of students taking AP exams. Notably, program implementation was associated with a 12-point increase in the percentage of students taking at least one math, science, or English AP exam — growth of more than a full standard deviation. In addition, CRP implementation was associated with strongly significant increases in the percentages of students earning qualifying scores, with effect sizes up to 0.5. Although the effects on exam taking clearly indicated that more students attempted AP exams in program schools than in non-program schools, it is also true that more students earned qualifying scores. This suggests the possibility that while CRP expands access to AP opportunities, it also supports an expanded pool of students who succeed.

Sherman & Song (2014, 2015)

Sherman and Song (2014, 2015), as part of their i3 evaluation of CRP in two states, provide longer-term evidence of CRP success, showing positive impacts on students' AP performance based on multiple years of program implementation across two cohorts of schools in Colorado and Indiana. Again using a CITS design, changes in average AP outcomes

over time of high schools implementing CRP (N=18) were compared with the changes in matched comparison schools that were not implementing the program (N=18). The authors utilized a two-level hierarchical linear model (HLM) nesting four student cohorts within each school and controlling for school background characteristics.

First-year outcomes in this study again show that CRP schools significantly outperformed the comparison group schools both in the percentage of students taking an AP exam in math, science, and/or English and in the percentage of students earning qualifying exam scores in these subjects. In the second year, using the same cohort of schools, the study found that treatment schools significantly outperformed comparison schools in the percentage of students taking AP exams and the percentage earning qualifying scores across all subject areas and all analyses (see Table 1). For example, the percentage of students who took an AP exam in math, science, or English increased by 7.80 percentage points for the treatment schools, but decreased by 2.29 percentage points for the comparison schools over the same time period (significant difference of 10.09 percentage points; $p < 0.001$). Similarly, the percentage of students earning qualifying scores on AP exams in math, science, or English increased by 3.28 percentage points, but decreased by .48 percentage points for the comparison schools over the same time period (significant difference of 3.76 percentage points; $p < 0.001$).

Jackson (2007, 2010, 2014)

Jackson's first two studies (2007, 2010), both quasi-experimental in nature, used a differences-in differences (DID) regression approach with matched comparison schools that wanted to implement the program. Both examined the impact of the early Texas APTIP program (now known as the College Readiness Program), extending the research beyond K-12 outcomes into the longer-term rationale for the program: success in the postsecondary years. **The earlier study, deemed consistent with WWC evidence standards with reservations in 2008, found positive effects on AP course enrollment, SAT/ACT scores, and college matriculation for students in participating schools (Jackson, 2007). The latter also identified longer-term outcomes of the program, reporting positive effects on college matriculation, college GPAs, and college persistence (Jackson, 2010).**

Jackson's 2014 work extends these outcomes by investigating not only the long-run educational effects of CRP, but also enduring labor-market outcomes, such as wages. It shows not only that CRP works, but also that it contributes to the desired end. Again using a quasi-experimental DID strategy, Jackson compares the change in outcomes between observationally similar students from the same high school before and after CRP adoption to the change in outcomes across cohorts from other high schools that did not adopt CRP over the same time period. Jackson's findings are derived from a sizable sample of students within schools that adopted the program (58 schools representing 137,704 students) and schools that did not adopt the program (1,413 schools representing 156,858 students). Through the study's design, access to extensive longitudinal data across multiple sectors, and use of a series of empirical tests, Jackson both builds a compelling case for the impact of CRP and successfully addresses a range of potential threats to validity.

#

Short-run AP outcomes examined over four years were significant, showing the program's positive effect on AP exams taken and qualifying scores earned (both p=.01). Postsecondary outcomes were also significant, with the program's effect being positively related to retention in college (e.g., "ever being a freshman", "ever being a sophomore"; both p=.01) and freshman year grade point average (p=.05). Jackson also reported a positive CRP effect on earnings, with an overall 2.7% increase that was largest and statistically significant (p=.05) for the second post-adoption cohort (3.8% increase in earnings).

Brown & Choi (2015)

Brown and Choi's approach (2015) employs a potential outcomes modeling approach (Rubin 2005) to estimate the causal effect of CRP program participation on first-, second-, and third-year improvements over base year in AP exam taking and AP qualifying score earning in math and science AP subjects. In addition to showing the impact of the program on the desired outcomes, it also shows the manner in which the impact happens. Using a propensity weighting approach (Rubin 2005), Brown and Choi accessed data from 287 treatment schools and 10,097 non-treatment schools.

Brown and Choi's results indicate substantial and significant increases in both AP exam taking and qualifying score earning for all students. In addition, significant first-year effects for AP exam taking and qualifying score earning were found for female students and minority students when analyzed separately. The average effect size (Cohen's d) for first-year increases over both average treatment on treated and average treatment effects for all students, all subgroups of students, both outcomes, and all disciplines was 0.64, showing a substantial positive causal impact. These first-year effects persisted into the second year (average effect size of 0.64) but diminished slightly in the third year (average effect size of 0.59). The effects are stronger when looking only at the average treatment on the treated effects, where the average effect size for first-year effects was 0.69. This increased to 0.73 for average second-year effects and returned to 0.68 for average third-year effects.

METHODOLOGY

Each study applies quasi-experimental design methods to key elements of the CRP model, and we believe each meets the evidence requirement of moderate internal validity and strong external validity. In addition, each study implements key elements of the CRP intervention (APTIP, APIP) under typical conditions to large and diverse samples. The analytic methods vary among comparative interrupted time series, difference-in-differences (DID) analyses, and potential outcomes modeling with causal estimation. Taken together, these studies provide strong evidence of the effectiveness of NMSI's College Readiness Program. The methodological specifications of the four distinct research efforts are detailed in the table below.

#

Table 1. Summary of Research on NMSI College Readiness Program

Evidence of Program Effectiveness		***Citation***			
		Holtzman (2010)	**Sherman & Song (2015)**	**Jackson (2014)**	**Brown & Choi (2015)**
Outcomes / valid measures	AP test taking	✓	✓	✓	✓
	AP test	✓	✓	✓	✓
	Postsecondary outcomes			✓	
	Earnings			✓	
Type of Study		QED	QED	QED	QED
Design		CITS	CITS	DID	POM
Sample	Large?	✓	✓	✓	✓
	Multi-site?	✓	✓	✓	✓
	Similar to i3 scale-up?	✓	✓	✓	✓
Groups	Treatment	64 schools	18 schools	58 schools / within school cohorts	287 schools
	Comparison	128 schools	18 schools	1,413 schools / within school cohorts	10,097 schools
Matching method	Propensity Scores	Nearest neighbors	Nearest neighbors	DID	Propensity Score weighting
Equivalence/Balance?		✓	✓	✓	✓
Potential confounds addressed?		✓	✓	✓	✓

#

	Hierarchical Linear Model	Hierarchical Linear Model	Ordinary Least Squares	Propensity Score Weighted Logistic Regression
Significant findings /no unfavorable outcomes?	✓	✓	✓	✓
Effect sizes reported?	✓	✓	✓	✓
Proposed determination: **Meets WWC Standards with Reservations**	✓	✓	✓	✓

RESULTS & CONCLUSIONS

These four research efforts represent an array of well-designed, well-implemented research studies that present solid evidence of the effectiveness of NMSI's College Readiness Program. Collectively, they demonstrate the positive effect of the program from impact on immediate outcomes related to Advanced Placement, to postsecondary results, to longer-term lifelong impacts. Individually, we propose that each study meets the What Works Clearinghouse (WWC) standards with reservations. As a group, they provide a sound evidence base for CRP effectiveness.

Taken together, the results of the Holtzman, Sherman & Song, Jackson, and Brown & Choi studies provide evidence that participation in NMSI's College Readiness Program generates significant effects for students that will positively impact their achievement, college readiness, persistent enrollment, and potentially their lifetime earnings. Given the suggested economic benefits of participation, the relatively low per-student cost, and the reliability of outcomes across settings and over time, we expect practitioners and policymakers to benefit from further understanding how CRP is brought to scale nationally and its continued educational and economic impacts.

#

REFERENCES

Brown, R, & Choi, K. (2015). *Measuring the Causal Effect of National Math + Science Initiative's College Readiness Program (CSE Tech Rep. No. 847)*. Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing. Retrieved from http://www.cse.ucla.edu/products/reports/R847.pdf

Holtzman, Deborah J. (2010). *The Advanced Placement Teacher Training Incentive Program (APTIP): Estimating the Impact of an Incentive and Training Program on Students.* Unpublished manuscript, American Institutes for Research, Washington, D.C.

Jackson, C.K. (2007). *A Little Now for a Lot Later: A Look at a Texas Advanced Placement Incentive Program*. Retrieved from http://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1070&context=workingpapers

Jackson, C.K. (2010). *A stitch in time: The effects of a novel incentive-based high school*. Unpublished paper submitted for publication. Chicago, IL: Northwestern University, 2010. Retrieved from http://www.nber.org/papers/w15722

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

Sherman, D., & Song, M. (2014). *Year 1 Findings for APTIP Impacts on Students' AP Performance in Cohorts 1 and 2 Schools*. American Institutes for Research. June 20, 2014.

Sherman, D., & Song, M. (2015). *Year 2 Findings for APTIP Impacts on Students' AP Performance in Cohorts 1 and 2 Schools*. Memorandum for Gregg Fleischer, NMSI, Washington, D.C.: American Institutes for Research.