Partnerships to Advance STEM Education: Building College Readiness for All Students

TABLE OF CONTENTS

RESPONSE TO PRIORITIES

Absolute Priority 3: Improving Science, Technology, Engineering, and Mathematics (STEM) Education

The National Math and Science Initiative's (NMSI) proposal, ***Partnerships to Advance STEM Education: Building College Readiness for All Students***, meets **Absolute Priority 3: Improving Science, Technology, Engineering, and Mathematics (STEM) Education** through the scale-up of its innovative and effective College Readiness Program (CRP). CRP increases the number and diversity of students **taking** and **earning qualifying scores (3 or above on a 5-point scale)** in Advanced Placement® (AP®) courses and exams in math, science, and English by transforming partner schools into centers of college readiness.[1] NMSI makes a dramatic difference in student achievement in only one year and transforms school culture over three years. Based on AP data from the College Board, the increase in qualifying scores after just one year of CRP implementation in six cohorts of NMSI partner schools (2009–2014) was 68% compared with the average national increase of 6.8% over the same time. CRP's lasting impact on students is discussed in more detail in the Significance section of this application (p. 9).

Specific math and science AP courses supported through CRP include: Calculus AB, Calculus BC, Statistics, Computer Science A, Chemistry, Biology, Environmental Science, Physics 1, Physics 2, Physics C: Mechanics, and Physics C: Electricity and Magnetism. NMSI includes English Language and Composition and English Literature and Composition as part of the program because mastery of language and reading skills is a critical component of college readiness and STEM preparedness.

If funded, the proposed project will increase the number of students earning qualifying scores on AP exams in math, science, and English. Research shows that success in AP positively influences college matriculation and graduation.[2] The overall goal of the program is to expand the number and diversity of students achieving at high levels, with a focus on STEM preparedness among traditionally underrepresented students.

Competitive Priority 1: Improving cost-effectiveness and productivity

<u>Substantially improving student outcomes without commensurately increasing per-student costs</u>

1. Advanced Placement® and AP® are registered trademarks of the College Board.
2. Geiser, Saul, and Veronica Santelices. 2004.

NMSI's College Readiness Program (CRP) has a strong track record of significantly improving student outcomes in partner schools. Research reflects a positive and statistically significant impact on student enrollment in Advanced Placement (AP) courses in math, science, and English and on students' success earning qualifying scores of 3 or higher on the associated AP exams.[3] In addition, the program has proven to have positive effects on SAT/ACT scores, college matriculation, college GPAs, and college persistence.[4] In addition to these short- and long-term improvements in student achievement outcomes, evidence suggests that there is an annual wage increase of roughly 3.7% per year for program participants.[5] This implies a per-student lifetime wage benefit of at least $16,650, which exceeds the current cost of the program.

Based on a project cost of $26,312,275 (i.e., the direct and indirect costs of the project, including NMSI's 50% match, but excluding the cost of the independent evaluation and related activities), the cost per student enrolled in AP courses is $940 (including 28,000 students enrolled in AP courses tracked as part of the evaluation study to assess impact); the cost per student drops to $439 when we include both AP students and pre-AP students reached by teachers in grades 3–12 trained as part of NMSI's Laying the Foundation (LTF) program (32,000 additional students). (See the Budget Narrative for details.)

Through CRP, NMSI works with existing schools and teachers to dramatically improve their capacity to support college readiness and better student outcomes. According to Jackson's 2014 study of CRP, "[V]ery little evidence has shown that one can improve students' long-run

3. Holtzman, Deborah J. 2010.

4. Jackson, C. K. 2010.

5. Jackson, C. K. 2014.

outcomes by adopting a program at their existing schools; the results of this study are encouraging about the potential efficacy of college-preparatory programs at improving the educational outcomes of disadvantaged students who attend inner city schools."[6] NMSI's cost-effective approach does not rely on opening new schools focused on college readiness that require significant infusions of private philanthropy and public financing to achieve improved student outcomes. The three-year per-student cost of CRP ($439) is dramatically lower than the estimated $3,500 per student in philanthropic support required to start a new college readiness–focused high school.[7]

These are one-time costs, and the results are sustained beyond the grant period because NMSI aims to transform schools' underlying cultures, such that all students — regardless of gender, ethnicity, or socioeconomic status — are seen as capable of achieving at high levels. The goal of the three-year program is to build knowledge and capacity in local education agencies (LEAs) so that LEA-based personnel can continue to lead the work after the grant period is complete (see p. 35-36 for details).

In addition, NMSI seeks to further reduce the per-student cost of CRP to make broad and rapid scale-up even more cost effective. We have identified several components of the approach for which we will pilot redesign during this grant period to increase productivity and enable a cost-effective approach. They include:

- **Selecting and training a locally based master corps of mentor teachers to serve each region.** Historically, NMSI has built its mentor teacher corps nationally, irrespective of

---

6. Jackson, C. K. 2014.
7. Bellwether Education Partners analysis.

location, and spent up to $873 per mentor annually on travel costs. We expect to save $644,122 in travel costs over the life of the grant by using local mentors.

- **Enabling school-based employees to manage key operational procedures.** Rather than managing these processes at NMSI's central office, we will provide teachers and other school- or district-based employees with stipends to cover the costs of their time required to complete these activities, thereby affording cost savings in terms of staff salaries, supplies, shipping costs, etc. We expect $271,542 in savings over the life of the grant.
- **Hosting teacher trainings locally.** All teacher trainings will be hosted locally. The estimated annual savings for AP teachers in the program will be $2,688 per year with $1,175 per year in savings for pre-AP teachers. Expected savings over the life of the grant total $3,254,672.

As part of the project evaluation (see p. 37), UCLA will assess the degree of fidelity of program implementation. NMSI will use these data, along with financial management system data to assess whether CRP is maintaining or reducing the cost per student while improving student outcomes.

Competitive Priority 2: Enabling the broad adoption of effective practices

The impact of the proposed scale-up project will far exceed the 60,000 students reached during the grant period. Specifically, this project will enable broad adoption in several ways:

- NMSI has developed a program that supports the strategic implementation of one of the only nationwide STEM college readiness programs: the College Board's Advanced Placement (AP) program. This helps enable broader adoption because the AP program is already offered in 14,000 schools nationwide.[8]

8. College Board.

- Our proposed project includes enrolling 1,000 educators who will teach 32,000 students, grades 3–12, in the feeder patterns for high schools enrolled in CRP. During the project period, we will open select Laying the Foundation training sessions to *all* educators employed within the LEAs with which we are partnered—not just the 1,000 teachers assigned to feeder-pattern schools. Conservatively, these sessions could be *open* to roughly 10,000 educators who annually instruct 300,000 students.[9]
- At the conclusion of the grant period, we will publish a guidebook to train teachers on how to adopt content-rich instructional techniques, set high classroom expectations, and shift students to advanced levels of thinking and learning. We will distribute the guidebook via NMSI's UTeach alumni network of more than 2,100 math and science teachers, at national conferences (e.g., Association for Supervision and Curriculum Development), and through other relevant dissemination methods. We will also host a convening of superintendents and assistant superintendents of instruction to help these leaders build a plan to increase college readiness and STEM proficiency using the guidebook as a roadmap.
- In addition, the Strategy to Scale section (see p. 20) summarizes our plan for disseminating results from the UCLA independent evaluation of the proposed i3 scale-up project.

9. This includes all teachers instructing in core subject areas for grades 3–12 in the partner LEAs. There are roughly 300,000 grade 3–12 students in these LEAs who will not directly participate in this project; we assume a 30:1 ratio of students to core teachers.

INTRODUCTION

The National Math and Science Initiative (NMSI), a 501(c)(3) nonprofit organization, is applying for this i3 scale-up grant project, ***Partnerships to Advance STEM Education: Building College Readiness for All Students***, to catalyze nationwide implementation of the proven College Readiness Program (CRP) by working with 10 local education agency (LEA) partners in eight states across the country to dramatically improve STEM outcomes. NMSI was formed to address one of this nation's greatest economic and intellectual threats—the declining number of students who are prepared to take rigorous college courses in math and science and are equipped for careers in those fields. NMSI's CRP is raising the academic bar in public schools by demonstrating that more students, especially high-need students, can master rigorous Advanced Placement (AP) coursework, with a particular emphasis on math and science.

CRP partners with schools to increase the number of students taking and earning qualifying scores on AP math, science, and English exams. There are three critical elements of CRP's success: teacher support, student support, and school support (see Quality of Project Design and Management Plan, p. 25, for more detail). During the project, we will reach approximately 60,000 students in 40 schools, including 28,000 high school students directly enrolled in AP courses and an additional 32,000 students in grades 3–12 who will participate in pre-AP activities through our Laying the Foundation (LTF) program.

Committed LEA partners for this i3 project include: Atlanta Public Schools (GA), Cleveland Metropolitan School District (OH), Detroit City School District (MI), Houston Independent School District (TX), Noble Network of Charter Schools (Chicago, IL), Oakland Unified School District (CA), St. Louis Public Schools (MO), and three small districts in North Dakota (Bismarck Public Schools, West Fargo Public Schools, and Mandan Public School

District). These LEAs reflect a variety of contexts (including size, percentage of students who are economically disadvantaged, and racial composition) and governing models (including traditional district school and charter school). Across LEA partners, an average of 75.9% of students qualify for free or reduced lunch. Among the LEAs are schools that serve dramatically higher concentrations of African American, Hispanic, and Native American students than the national average—minority groups who are traditionally underrepresented in STEM fields. The Brookings Institution identified seven of the eight geographies in which partner LEAs are located in the top half of rankings of "STEM-intensive" metropolitan areas in which a high percentage of available jobs require STEM knowledge.[10] See Table 2 on p. 17-18 for more detail about students served in LEA partners and nationwide demographics.

Rigorous research confirms strong evidence of effectiveness at the national level for CRP (see Significance, p. 9). In 2012, NMSI's four schools in Hawaii accounted for 89% of the entire state's increase in qualifying math, science, and English scores, and NMSI's two schools in Oklahoma accounted for 35% of that state's increase. The track record continued in 2013 and 2014 with similar results in Colorado, Maryland, Indiana, Mississippi, California, and Pennsylvania (see Appendix C for additional detail).

Expected outcomes of the proposal include: 1) student enrollment in AP courses, particularly among traditionally underrepresented populations, will increase from the baseline year by at least 80% for each LEA partner in the first year and 140% over three years; and 2) students' qualifying scores in program schools in AP math, science, and English exams will

10. Rothwell, J. 2013. The only LEA partner geography not represented in the Brookings Institution analysis is North Dakota.

increase by at least 70% for each LEA in the first year, and at least 125% over three years.

Building on the success of NMSI's 2011 validation i3 grant, the goal of this project is to catalyze the next phase of national scale-up of CRP by enabling NMSI to create and/or deepen eight regional hubs across the country, focused on where the need is the greatest. The LEA partners identified for this project were intentionally selected because of their concentration of high-need students, the historical opportunity gaps within these regions, and/or the "STEM-intensive" economic environment in which our partners educate students.

SIGNIFICANCE

The extent to which the proposed project addresses a challenge for which there is a national need for solutions that are better than the solutions currently available

Ensuring that all students have access to and excel in STEM fields is essential for our nation's economic growth and future prosperity.[11] In 2011, roughly one-third of U.S. bachelor's degrees were awarded in science and engineering fields, compared to 60% in Japan and 50% in China.[12] Our knowledge capital, which fuels innovation and economic growth, is at risk.

Despite the national publicity and attention focused on these needs, they stubbornly persist. It is estimated that in 2014, only 43% of U.S. high school graduates were ready for college-level math, and only 37% were ready for college-level science.[13] These problems are even more pronounced for the high-need and traditionally underserved students whom the proposed LEA

11. Langdon, D., G. McKittrick, D. Beede, B. Khan, and M. Doms. 2011.

12. National Science Board. 2014.

13. ACT, Inc. 2014.

partners serve in high concentrations; these students face hurdles because of policies and mindsets that limit their ability to access rigorous coursework (see p. 22 in the Strategy to Scale section for more details). Recent National Science Foundation (NSF) data found that more than one-fourth of ninth graders in NSF's lowest socioeconomic status category were not enrolled in any science courses (27%), compared with 11% of students in the highest income category. These differences in access lead to achievement gaps that persist through college and beyond. The gap between white students' six-year college graduation rates and their African American peers is 22 percentage points, and the gap between white students and their Hispanic peers is 10 percentage points.[14]

A growing body of evidence indicates that 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 AP exams, hence increasing their achievement and college readiness, but also has significant and longer-term positive postsecondary and economic impacts.** The program's consistent elements produce reliably successful and sustained outcomes across settings, states, subject areas, and students, including those students traditionally underrepresented in STEM. Across studies, research questions consistently relate to the extent to which implementation of CRP is associated with increased percentages of high school students taking AP exams and increased percentages of students scoring 3 or higher on these exams, particularly in math, science, and English.

14. Kena, G., et al. 2014. See Table 326.10: Graduation rate from first institution attended for first-time, full-time bachelor's degree-seeking students at 4-year postsecondary institutions.

Summarized in the table titled “**Assessing Strong Evidence of Effectiveness to Demonstrate Eligibility for i3 Scale-up Grant”** (Appendix D), the four studies upon which we focus represent an array of well-designed, well-implemented research studies that present solid evidence of the effectiveness of CRP, from impact on immediate outcomes related to AP, 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 collective group, we purport that CRP is supported by the strong evidence of effectiveness required for the proposed i3 scale-up grant.

Holtzman (2010) found that in its first year, **CRP had a positive and statistically significant 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 their 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 areas/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 and Song (2014, 2015), as part of their current i3 evaluation of CRP in two states, provide longer-term evidence of CRP success, showing positive impacts on students' AP performance based on <u>multiple years</u> 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; $p < 0.001$).

TABLE 1. Two-Year Impacts of CRP on the Percentage of Students Taking AP Exams and the Percentage of Students Scoring 3 or Higher on AP Exams in Cohort 1 Schools in Colorado and Indiana, By Subject

Outcome	**Average Deviation from Baseline Mean**		**Difference**	**Standard Error of Difference**	**P-Value**
	Treatment Group	**Comparison Group**			
Percentage Taking AP Exam					
Math/science/English	7.80	-2.29	10.09	1.44	0.000***
Math/science	5.23	-1.71	6.94	1.02	0.000***
Math	3.35	-0.82	4.17	0.80	0.000***
Science	3.81	-0.71	4.51	0.84	0.000***
English	4.47	-1.25	5.72	1.25	0.000***
Percentage Passing AP Exam					
Math/science/English	3.28	-0.48	3.76	0.68	0.000***
Math/science	2.93	-0.26	3.19	0.50	0.000***
Math	1.84	-0.30	2.14	0.42	0.000***
Science	1.90	0.01	1.90	0.37	0.000***
English	1.58	-0.30	1.88	0.60	0.002**

NOTES: Number of schools = 36 (18 treatment, 18 comparison). Average deviation from baseline mean for the treatment group is unadjusted average deviation across treatment schools; average deviation for the comparison group was computed by subtracting the estimated group difference from the unadjusted average deviation for the treatment group. P-values are based on two-tailed t-tests. * $p < .05$; **$p<.01$; ***$p<.001$.

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 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.

Taken together, the results of the Holtzman, Sherman and Song, Jackson, and Brown and Choi studies suggest that participation in **CRP is expected to have 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.

The potential replicability of the proposed project or strategies, including, as appropriate, the potential for implementation in a variety of settings.

As a national entity with experience scaling its program in 727 schools across 25 states over time, NMSI is well positioned to guide further scale-up and adoption of CRP in a range of settings that serve high-need students. The CRP model works in a variety of settings — urban and rural, disadvantaged and affluent, charter and traditional district schools, from coast to coast — and for a variety of students. CRP has scaled successfully within school districts serving low-income students — including Dallas Independent School District (DISD)[15] (described in more detail on p. 22), Pittsburgh Public Schools[16], and IDEA Public Schools[17], for example — and has a proven track record of improving results for traditionally underrepresented minorities and women. For example, the U.S. average one-year increase in qualifying scores in math, science, and English among African American and Hispanic students is 9.7%; at NMSI partner schools it is 81%. Over three years, the average national increase for minorities is 48%; among NMSI partner schools it is 179%. Similarly, among young women, the first-year increase in qualifying scores in math, science and English is 6.5% nationally and 68% for NMSI partner schools. Over

15. Free or reduced meal rate of 89.6% in 2013-2014.

16. Free- or reduced- meal rate of 71.0% in 2013-2014.

17. Free- or reduced- meal rate of 83.2% in 2012-2013.

three years, the average national increase for young women is 22% nationally; among NMSI partner schools it is 122%.[18]

This grant will allow NMSI to continue replication of the program to more regions across the country, including districts with very different state standards and requirements. Our proposed LEA partners range in size from 3,599 to 215,157 students. They have different student population compositions, with free-or-reduced-lunch percentages ranging from 20.8% to 100%[19] (the blended average across all districts is 75.9%, reflecting a high concentration of high-need students). Among the LEAs are schools that serve higher concentrations of African American, Hispanic, and Native American students than the national average—minority groups who are traditionally underrepresented in STEM fields. The table below provides more detail about students served in LEA partners.

TABLE 2. Details and Demographics on Proposed CRP Partner LEAs

LEA name	State	# of students	# of high schools	% FRL	% African American	% Hispanic	% Native American
National Data (demographics)	N/A			49.6	15.3	26.4	1.0
Cleveland Metropolitan School District	OH	38,725	23	100	66.9	14.4	*
Oakland Unified School District	CA	48,181	7	73.4	29.7	39.3	*

18. College Board (2013a).

19. The Cleveland Metropolitan School District reports a free-or-reduced-lunch rate of 100% because it sought a waiver to enroll all of its schools and students in the free-or-reduced-lunch program.

St. Louis Public Schools	MO	24,869	17	88.8	82.7	*	*
Noble Network of Charter Schools	IL	10,235	16	90.9	50.6	45.7	*
West Fargo Public Schools	ND	9,074	3	27.6	10.4	3.3	2.8
Bismarck Public Schools	ND	12,003	4	20.8	2.7	2.6	8.4
Mandan Public School District	ND	3,599	1	28.4	2.5	2.5	7.6
Houston Independent School District	TX	215,157	40	75.5	24.9	62.1	*
Atlanta Public Schools	GA	50,253	24	77.6	80.4	3.7	*
Detroit City School District	MI	47,959	24	75.5	83.2	12.7	*

* Indicates no data available or a percentage too small to be noted in a separate category. Sources: data provided by each partner LEA (2015 enrollment estimates, 2012 FRL data).

The need is very significant in all of these LEAs — for example, in Detroit, only 37 of the 5,620 African American and Hispanic 11th and 12th graders earned a qualifying score on a math or science AP exam in 2013; in St. Louis that number was only five of 2,782. The average percentage of African American and Hispanic 11th and 12th graders earning a qualifying score on a math or science AP exam across the LEAs (excluding North Dakota and Cleveland) was only 1.7%. In North Dakota, there were only 18 qualifying math and science AP scores for every 1,000 high school juniors and seniors — this is less than one-fifth the number for the U.S.[20]

The extent to which the proposed project involves the development or demonstration of promising new strategies that build on, or are alternatives to, existing strategies.

NMSI's proposed i3 project focuses on scaling up CRP, a program that increases participation and performance in rigorous AP math, science, and English courses. CRP addresses the need to improve STEM education, increase academic intensity, and improve student

20. College Board (2013a).

achievement in order to decrease the college readiness gap,[21] especially among traditionally underrepresented and high-need students.

The most important factor that differentiates the CRP model is its ability to demonstrate measurable student outcomes on a rigorous metric. The strong evidence of this effectiveness at the national level is summarized beginning on p. 10. In addition, we summarize three other factors that enable CRP to uniquely address challenges faced by high-need students.

***(1) Focus on high school reform**:* CRP dares to tackle the STEM crisis at the high school level, when evidence shows it can be very difficult to catch up those students who lag behind.[22] Participating CRP schools rethink their AP culture by adopting open enrollment and recruiting more students, including high-need and traditionally underrepresented students, into AP courses, thereby allowing many more students to succeed at that level.

***(2) Investment in building the pipeline of students prepared for rigorous coursework**:* CRP also focuses on using vertical teaming, curriculum alignment, and scaffolding of course content to build a pipeline of students who enter high school prepared for rigorous coursework.[23] NMSI's Laying the Foundation (LTF) Teacher Training Program for educators of students in grades 3–12 is an important component of NMSI's CRP partnership with schools. The program equips pre-AP teachers with the content knowledge and instructional know-how they need to set

---

21. For purposes of this application, the college readiness gap is measured by the number of high-need students who take and earn qualifying scores on AP exams, because the AP exam is one of the few nationally accepted proxies for college readiness

22. Dougherty, Chrys and Steve Fleming. 2012.

23. Jackson, C. K. 2010.

high classroom expectations and shift students to advanced levels of thinking and learning. The LTF program complements CRP by training teachers to facilitate students' progression through the academic pipeline toward advanced coursework starting as early as third grade. Developed by experienced teachers and content experts, the program provides:

- **Comprehensive, hands-on training led by a national corps of expert classroom teachers.** LTF trainers guide participants through content-rich instruction that moves beyond *what* to teach to *how* to deepen student understanding of key concepts.
- **Classroom-ready materials and resources aligned with state standards.** Training participants receive a full suite of print and online classroom-ready materials that encourage higher-order thinking.
- **Instructional best practices for increasing academic rigor and building college and career readiness.** LTF training emphasizes research-based instructional strategies including: inquiry-based learning, instructional scaffolding (techniques and guidance for delivering differentiated instruction), and vertical alignment (education about the knowledge and skills that students need to master at each grade level).

(3) Program structure designed to work within the framework of existing schools: The literature on CRP reflects that it stands alone among evidence-based programs that focus on adopting a program at existing schools.[24] NMSI's cost-effective approach does not rely on opening new schools focused on college readiness that require significant infusions of private philanthropy and public financing to achieve improved student outcomes. See Competitive Priority 1 (beginning on p. 2) for a summary of the cost effectiveness of this approach.

24. Jackson 2014.

STRATEGY TO SCALE

Four key elements of our scaling strategy, during this project and over time, are as follows: 1) enter eight regions that will become "regional hubs" for NMSI; 2) build a strong STEM human capital pipeline; 3) address school- and LEA-level barriers that inhibit our ability to scale up; and 4) enable broad adoption by sharing best practices and tools.

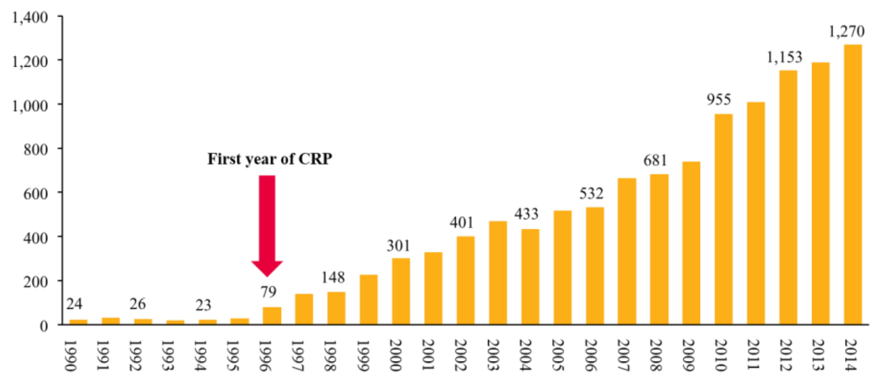
<u>There is unmet demand for CRP that will enable NMSI to reach the level of scale that is proposed in the application</u>

There is clear demand nationwide for CRP. When NMSI issued an RFP to assess interest from local communities in replicating CRP in 2008, it received applications from organizations in 28 states. In addition, the U.S. Department of Defense (DoD) partnered with NMSI because of the significant demand for the program on military bases across the country, where continuity of high-quality curriculum and instruction is a particular concern for military families. (See Appendix J-1 for more information about NMSI's partnership with DoD.)

In 2011, NMSI won a $15 million i3 validation grant enabling replication of CRP in two regional hubs (Colorado and Indiana) and assessment of the quality of the results. The goal of NMSI's regional hub strategy is to establish a CRP presence within a target geography serving schools within one or more LEA. By establishing an initial foothold in these metropolitan areas, we will grow NMSI's footprint, diversify its partnership base, and explore deeper engagements once we have local results that can be replicated and scaled. Regional hubs also demonstrate to other potential expansion regions what is possible through CRP.

NMSI's partnership with Dallas Independent School District (DISD) provides an example of what CRP's impact in a geographic hub looks like over time. DISD is a large, urban school district in which 89.6% of students currently qualify for free or reduced lunch. CRP (then called APTIP) was piloted in 10 DISD high schools in 1996. As detailed in Figure 1, African American and Hispanic student qualifying scores increased **43 times higher** over 18 years.

FIGURE 1. Number of AP Qualifying Scores Earned by African American and Hispanic students in Math, Science and English in 10 DISD Program Schools, 1990–2014



Source: College Board data

The demonstrated success of the DISD pilot led to demand for CRP's replication in 34 DISD schools and in 26 LEAs in Texas and success in 92 schools of all sizes, demographics, and locations. Success in one district rippled to affect students across the state, and encouraged policymakers to adopt policies that enable more students to access the rigorous STEM coursework necessary for success.

Given the success of this regionally-based approach, NMSI is now prepared to scale up eight regional hubs across the country. We have commitments from 10 LEAs that represent a range of schools and settings (see p. 17 for more details) in these regions to partner with NMSI on the implementation of this i3 project (see MOUs in Appendix G). Over time, we expect that we will work with a high percentage of high schools within these eight regions, just as we did within DISD.

<u>The extent to which NMSI will use grant funds to address a particular barrier that prevented NMSI, in the past, from reaching the level of scale proposed in the application</u>

As part of this project, we will directly address barriers to scale related to 1) upfront investment required to implement CRP; 2) availability of STEM-qualified mentor teachers to implement the program in selected regions; and 3) school-level conditions that enable successful implementation of CRP. By conducting activities outlined on p. 5 to enable broad adoption of CRP practices, we will also pave the way for addressing policy barriers that can limit adoption.

Upfront investment required to implement CRP: Although the three-year, per-student cost of the proposed project is only $439, the upfront investment required for a school or LEA within a new NMSI region is still significant given constrained district and school budgets. Expansion to new geographies enables additional proof points and increased awareness of the program, which in turn increases adoption and accelerates the scaling process described above.

Availability of STEM-qualified teacher mentors: Historically, NMSI has developed a nationwide corps of STEM-qualified teacher mentors to serve all of its schools because we did not have the scale required within a given region to hire and support local talent. Under this proposed project, we will focus on recruiting and training regional corps of master teachers who will serve as mentor teachers in each of the partner LEAs. Identifying and training this talented corps will enable rapid scaling within the given regions over time.

School-level conditions that enable successful implementation of CRP: Our capacity to implement our work effectively in the past has been limited by conditions within schools that can inhibit effective program implementation, including lack of instructional coherence, inexperienced teaching staff, low expectations for students, and lack of cultural responsiveness. As part of this proposed project, we have identified 10 LEA partners (in eight states) that are committed to enabling change in schools with which NMSI will partner as a condition of their participation (see MOUs in Appendix G). Having a district-level commitment to CRP as part of the i3 scale-up grant will help us overcome many school-level barriers.

Finally, we face significant policy barriers at the district level, including policies that limit enrollment of low-income students, and at the state level, including state accountability systems that penalize schools for low AP rates of qualifying scores, thereby incentivizing reduced access to low-income students. Although we will not directly change these policies as part of this work, our activities to disseminate learnings will indirectly highlight the negative effect of these policies, particularly for traditionally underrepresented students.

The mechanisms NMSI will use to broadly disseminate information on its project so as to support further development or replication

We chose our partner LEAs for their diversity and breadth of implementation settings, as detailed in the table on p.17. Scaling CRP across such a varied group of LEAs ensures that other interested districts or policymakers will be able to clearly see how the program can work in their specific context, with their specific constraints and unique assets.

- During the project period, we will conduct LTF trainings that will be open to *all* educators employed within the LEAs with which we are partnered (as well as LEAs in surrounding regions). See p. 5 for more information on this strategy to enable broad adoption.
- At the conclusion of the grant period, we will publish a guidebook to train teachers on how to adopt content-rich instructional techniques to set high classroom expectations and shift students to advanced levels of thinking and learning. See p. 5 for more information.
- Project results from the UCLA independent evaluation will be disseminated both through the NMSI website and the Center for Research on Evaluation in Standards and Student Testing (CRESST) website (www.cresst.org), at regular conferences and workshops, and in peer-

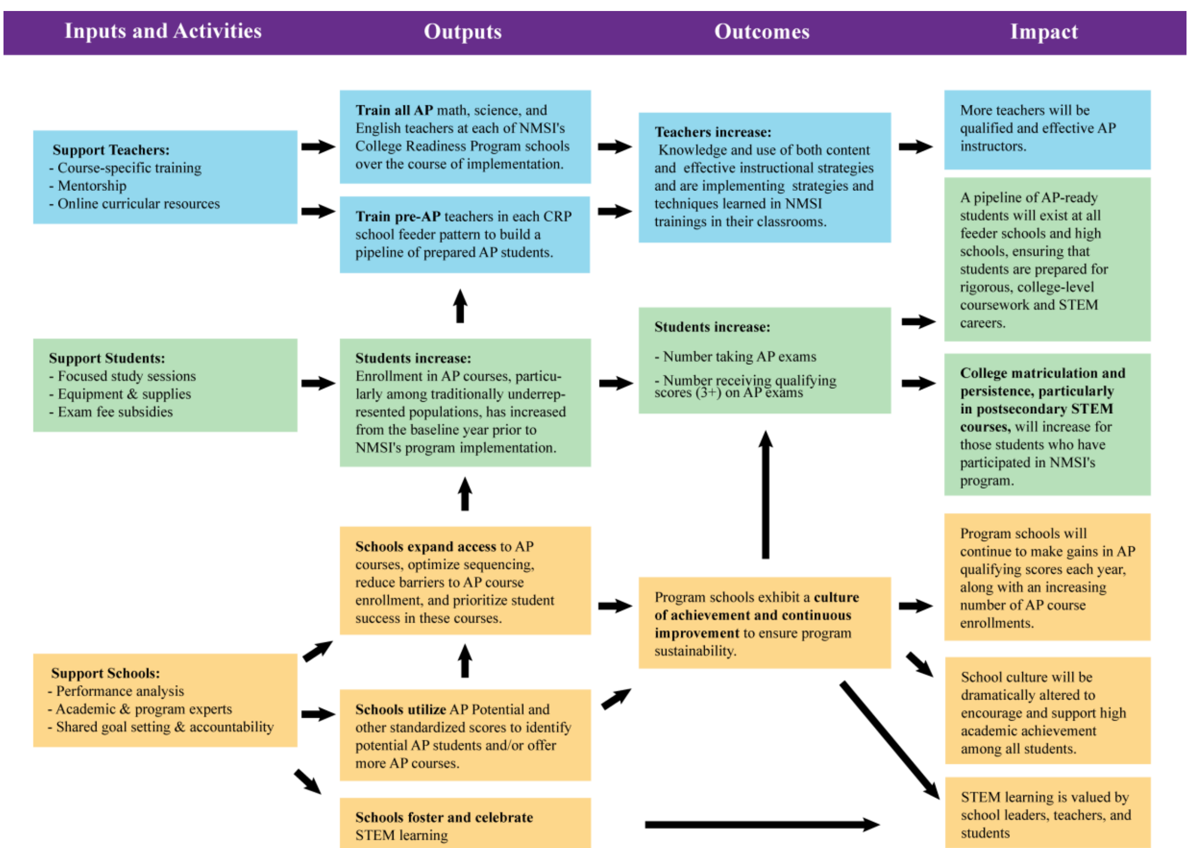
reviewed publications. UCLA plans to formally disseminate research results as publications in peer-reviewed substantively oriented journals and methodologically oriented journals. UCLA will also collaborate with NMSI on providing feedback to participating schools.

QUALITY OF PROJECT DESIGN AND MANAGEMENT PLAN

<u>The extent to which the goals, objectives, and outcomes to be achieved by the proposed project are clearly specified and measurable.</u>

The objective of CRP is to dramatically increase the number of students taking and earning qualifying scores on AP math, science, and English exams. The CRP logic model (see below; also attached as Appendix J-2) identifies the key factors of CRP that are necessary ingredients for success across students, teachers, and schools.

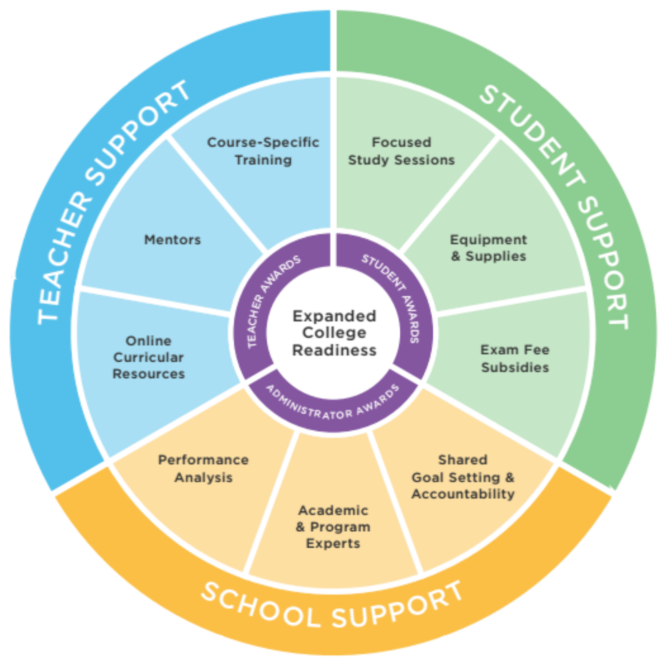
FIGURE 2. NMSI's College Readiness Program (CRP) Logic Model



Description of key logic model inputs / activities*:* Within the logic model for CRP, NMSI has identified the key factors that are indispensable to scaling the program with fidelity. These Elements of Success (see Figure 3 below) are the foundation for successfully scaling CRP, and, as such, NMSI requires strict adherence to each. See Appendix J-3 (CRP Implementation Processes) for detailed information about how each element of success is implemented.

i. Teacher Support

FIGURE 3. Elements of Success



Course-Specific Training: Teacher-to-teacher instruction and support including:

- Four-day AP Summer Institutes, followed by two-day workshops each fall to reinforce content knowledge.
- Two-day training in the fall and mock exam grading in the spring.
- NMSI's Laying the Foundation training program for teachers of grades 3–12 to build the pipeline of students ready for rigorous AP courses (summarized in more detail on p. 19).

Mentors: One-on-one support—including curricular help, pacing guidance, and instructional feedback—provided by expert teachers with a proven record of success in AP performance.

Online Curricular Resources: CRP teachers have access to two key online resources:

- **The NMSI Teacher Portal:** Each subject has a separate section for teachers with details regarding successful implementation of AP programs, targeted toward the most important information for increased achievement on AP exams.

- **The Quest Homework System:** An extensive knowledge bank of more than 60,000 questions and answers covering math, biology, chemistry, computer science, and physics that provides a user-friendly, formative assessment system that allows teachers to customize their instruction and homework assignments toward a student's individual learning needs.

ii. *Student Support*

Study Sessions Focused on Student Support: At least three study sessions per AP subject, typically held on Saturdays, led by state and national AP experts who teach NMSI-created lessons as students' classroom teachers observe.

Equipment and Supplies: Access to the latest laboratory equipment, graphing calculators, and other classroom materials needed to support rigorous coursework.

Exam Fee Subsidies: Program funds typically covering 50% of the cost of each AP exam.

iii. *School Support*

Performance Analysis: Annual review of program components and compliance to ensure maximum program effectiveness.

Academic and Program Experts: Detailed curricular, programmatic, and logistical support provided by experienced academic content directors and program managers.

Shared Goal Setting and Accountability: Mutually agreed upon expectations for program participation and support, as well as performance goals for teachers, students, and schools.

iv. *Awards*

Offering financial awards sends a message to students and teachers alike that success in rigorous courses and extra studying and teaching time are valued. Prior research shows that students who participated in CRP in high school went on to attend college in greater numbers and had improved college GPAs, and that African American and Hispanic students were more likely to stay in college beyond their freshman year.[25] Concerns that awards-based interventions may lead to “teaching to the test” and cheating were not realized, while the benefits of CRP awards that induce students to reach for higher standards lasted in post-secondary education.

- **Teacher Awards:** $100 for each qualifying score earned by a student who appears on a teacher’s AP class roster, plus a $1,000 bonus based on a qualifying score threshold set for specific AP courses.
- **Student Awards:** $100 for each qualifying AP exam score earned.
- **Administrator Awards:** Awards granted to designated school administrators or lead teachers, subject to the specific grant agreement for each school.

Measurable goals and outcomes

As shown in the logic model above — through CRP’s student support, teacher support, and school support — a number of short- and long-term outcomes will be achieved that will help transform partner LEAs into centers of college readiness.

Magnitude of project impact on students, teachers, and leaders:

- We will reach approximately **60,000** students over the course of the project, including **28,000** students directly enrolled in AP courses and an additional **32,000** students in grades 3–12 who will participate in pre-AP activities.
- We will train approximately **1,000** teachers from CRP schools and feeder elementary and middle schools via our LTF program during the grant period. Select trainings will also be available to the more than 10,000 educators within the partner LEAs who teach grades 3–12.

25. Jackson 2014.

Outcomes over the course of the grant: Even over the relatively short time span of three years, CRP will have a deep impact on its **10** LEA partners and 40 of their schools, with a goal eventually affecting all **159** of their high schools (see p. 17 for number of high schools by LEA).

- Students' qualifying scores in program schools in AP math, science, and English will increase by at least **70%** for each LEA after the first year of CRP, and at least **125%** over the three-year grant period. *Measurement:* AP qualifying scores, collected annually.
- Teachers will report increased knowledge and use of both content and effective instructional strategies and will implement strategies and techniques learned in NMSI trainings in their classrooms. *Measurement*: Formal and informal surveys, site visits, and NMSI mentor feedback, collected annually.
- Student enrollment in AP courses, particularly among traditionally underrepresented populations, will increase from the baseline year prior to NMSI's program implementation by at least **80**% for each LEA partner in the first year and 140% over three years. *Measurement*: Course enrollment data shared by schools, collected annually.
- Schools will make necessary changes to facilitate expanded access to AP courses and prioritize student success in these courses. This will include adding AP courses and altering AP sequencing based on NMSI's recommendations. *Measurement*: Schools' adding AP courses and altered AP sequencing based on NMSI's recommendations, and administrator and teacher implementation of programmatic feedback provided by NMSI's program team, all collected annually.

Long-term outcomes: We anticipate that this project will have several important long-term impacts that will continue well after the completion of the i3 grant period. These long-term outcomes include:

- Other, non-program schools in the LEAs will see what's possible and the LEAs may provide funding through their own means or other partners for their non-program schools.
- School culture in partner LEAs will be dramatically altered to encourage and support high academic achievement among all students.
- Program schools will continue to make gains in AP qualifying scores each year, along with an increasing number of AP course enrollments.
- A pipeline of AP-ready students will exist at all LEA partner feeder schools and high schools, ensuring that students are prepared for rigorous, college-level coursework.
- STEM learning will be fostered and celebrated across partner LEAs.
- Policy makers in LEA partner districts and states will become more aware of the importance of expanding access to AP for all high school students.
- College matriculation and persistence, particularly in postsecondary STEM courses, will increase for those students who have participated in NMSI's program.

The adequacy of the management plan to achieve the objectives of the proposed project on time and within budget

NMSI has developed a robust management plan to ensure it meets its project objectives on time and within budget, consistent with previous success in implementing large-scale grants, including the successful 2011 i3 validation grant. The table below summarizes key responsibilities, timelines, and milestones for accomplishing key project tasks. A detailed CRP implementation timeline is provided in Appendix J-4.

TABLE 3. Key Activities and Milestones

Activities and milestones	Responsible Party	Year and Quarter				
		Y1	Y2	Y3	Y4	Y5
Prepare for implementation: Hire positions needed, meet with evaluator, confirm data-sharing agreements with College Board, and execute contracts with each LEA partner	CAO CRS UCLA	Q1-Q2				
Execute contracts with each LEA partner and identify schools	CAO CRP team	Q1-Q2	Q1-Q2			
Human capital: Identify mentor teachers to serve as hub of CRP at each school	CRP team	Q1-Q2	Q1-Q2			
Agree upon annual participation and performance goals for teachers, students, and schools	CAO CRP team	Q3	Q3	Q3	Q3	Q3
Develop comprehensive evaluation plan and management plan for submission to the Department of Education	CAO CRS GM	Q1				
Communication: In each community/LEA in which NMSI will expand, launch tailored teacher, staff, parent, and community awareness program, including grant announcement event	CRP team Communications	Q2-Q3	Q2-Q3	Ongoing		
Enroll teachers in NMSI systems	CRP team	Q2-Q3	Q2-Q3			
Order AP course equipment (e.g., science lab materials, graphing calculators)	CRP team	Q2	Q2	Q2	Q2	
AP teachers attend summer institute	CRP team	Q3	Q3	Q3	Q3	
Teachers for grades 3–12 attend Laying the Foundation teacher training	CRP team	Q3	Q3	Q3	Q3	
Student study sessions begin	CRP team	Q4	Q4	Q4	Q4	
Teachers attend AP two-day workshop	CRP team	Q4	Q4	Q4	Q4	
Students complete mock exams	CRP team		Q1	Q1	Q1	Q1
Students complete AP exams	LEAs		Q2	Q2	Q2	Q2
Collect annual feedback from students, teachers, administrators, and staff to inform continuous improvement	CRP team IT		Q2-Q3	Q2-Q3	Q2-Q3	Q2-Q3
AP scores received; verification of schools and participation confirmed; teacher and student payments made	CRP team Finance		Q2	Q2	Q2	Q2
Make semi-annual updates to program to reflect feedback from key stakeholders, partners, and participants	CAO CRP team GM		Q1, Q3	Q1, Q3	Q1, Q3	Q1, Q3

CAO=Chief Academic Officer; CRS=Chief Research Scientist; GM=Grant Manager

NMSI's project leaders have experience managing large, complex, and rapidly growing projects. NMSI's leadership team for this grant includes: (1) the CEO, who has successfully managed NMSI's i3 validation grant as well as many large federal grants at other organizations; (2) the Chief Academic Officer, who designed the CRP program and **will lead implementation of the i3 grant**; (3) The Chief Research Scientist, **who is responsible for working with UCLA to support the external evaluation**; and (4) the Chief Financial Officer, who has overseen the budgets for several federal grants including the 2011 i3 validation grant. The "CRP Team" includes a Content Director for each subject area (math, science, and English) four Program Managers (supported by two Program Assistants), who act as the account executives by traveling to the regional hubs frequently. In addition, the Grant Manager will manage all reporting requirements.

Finally, UCLA's CRESST (National Center for Research on Evaluation, Standards & Student Testing) will lead the independent project evaluation. The principal investigator for this project is Dr. Eva Baker. Brief biographies for the senior project leadership team can be found below, with complete résumés for each key staff member available in Appendix F.

Matthew Randazzo, Chief Executive Officer: As CEO, Matthew leads the strategic direction of the organization. Prior to joining NMSI, Matthew served as founding CEO of Choose to Succeed, where he mobilized $45 million to support exceptional, tuition-free public education options for families in San Antonio, Texas. Previously, he served as Chief Growth Officer for IDEA Public Schools, where he oversaw the execution of aggressive growth plans for a network of K–12 public charter schools serving more than 20,000 students in 36 schools.

Gregg Fleisher, Chief Academic Officer: As CAO, Gregg provides strategic direction and oversight of NMSI's core programs and **will lead the proposed i3 scale-up project.** In 1995, Gregg started the AP Incentive Program (now known as NMSI's College Readiness Program) in Dallas ISD while working as a math teacher. In 2000, he was asked to lead the nonprofit Advanced Placement Strategies (APS), which managed the College Readiness Program for districts in the state of Texas.

Richard Brown, Chief Research Scientist: As CRS, Rich is responsible for a program of research, systems, and measures to evaluate the effectiveness of NMSI's core programs in improving teacher performance and student achievement and **will manage the external evaluation of the proposed i3 scale-up project.** He previously served as an associate research professor at University of Southern California's Rossier School of Education. Rich is contracted to NMSI through West Coast Analytics, an organization providing contract support for this grant.

Tammy Knapp, Chief Financial Officer: As CFO, Tammy is responsible for all financial matters related to NMSI operations, including budget development and oversight as well as financial reporting and compliance related to numerous public and private grants. Tammy previously worked for Bank of America, where she held a variety of positions including Chief Financial Officer of a subsidiary bank.

Eva Baker, Principal Investigator, UCLA CRESST: Dr. Eva L. Baker is a Distinguished Professor of Education at UCLA, and has served as Director of CRESST since 1975.

<u>The clarity and coherence of NMSI's multi-year financial and operating model and accompanying plan to operate the project at a national level during the project period</u>

NMSI has overseen $300 million in public-private funds since 2007. The organization had an annual operating budget of $30.8 million in 2014. NMSI has smoothly operated its i3 validation grant since 2011 and is therefore familiar with reporting and accountability standards at the federal level. The budget narrative reflects an overall project budget of **$29,997,913, including a federal request of $19,998,619 and philanthropic matching budget of $9,999,294.** Please see the budget narrative for more detail.

In preparation for the proposed project, NMSI carefully reviewed its staffing structure, and this project includes a budget request for additional staff where needed to successfully implement the project including the addition of four program managers and two program assistants. As described above, we have identified staff with the highest qualifications, experience, and expertise to ensure fidelity of implementation. The budget narrative includes detailed information about time allocated by staff member to each project. Résumés are attached in Appendix F.

NMSI is accustomed to operating at a national level, enabled by a strong board of directors, chaired by Tom Luce. Mr. Luce served as United States Assistant Secretary of Education for Planning, Evaluation, and Policy from 2005 to 2006. His biography, along with all other members of the board of directors, is attached in Appendix F.

The LEAs that will be implementing this project in partnership with NMSI — Atlanta Public Schools (GA), Cleveland Metropolitan School District (OH), Detroit City School District (MI), Houston Independent School District (TX), Noble Network of Charter Schools (Chicago, IL), Oakland Unified School District (CA), St. Louis Public Schools (MO), and three small districts in North Dakota (Bismarck Public Schools, West Fargo Public Schools, and Mandan Public School District) — all have MOUs included in Appendix G that clearly define roles and responsibilities for each partner LEA in project implementation.

In order to best steward its resources, NMSI engages in a vetting process to determine which schools within partner LEAs will benefit most from CRP. Using a consistent selection process helps ensure that teachers and administrators demonstrate the commitment and key mindsets required for the success of CRP. NMSI looks to work with schools that demonstrate the following characteristics:

- ***Growth mindset:*** educators must be committed to the principle that all students can succeed
- ***Implementation capacity:*** schools must be committed to focusing on the implementation of CRP and have the resources to do so, such as teachers available to expand AP course offerings, the scheduling capability to meet expanded course needs, etc.
- ***Leadership:*** school leaders must be willing to address the barriers preventing schools, teachers, and students from maximizing success in AP performance and to make changes within the schools to leverage the full benefits of the CRP, which often includes amending school policies (related to grading, scheduling, and course admittance).
- ***Teacher commitment:*** Teachers must commit to attend training sessions, implement key program elements into their instruction, and be willing to utilize feedback and new instructional concepts in their classrooms.

A copy of NMSI's Initial Visit Assessment template, used to help evaluate potential schools, can be found in Appendix J-5.

The goal of the three-year program is to build knowledge and capacity in LEAs so they can continue to do this work on their own after the grant period. Because a number of factors affect the specific amount and type of sustainment support needed (e.g., total student population, AP course offerings, and teacher turnover) NMSI will work directly with school leaders to determine priority areas and create explicit budgets. NMSI will provide schools with a fundraising toolkit to help them pursue additional support from their own local communities. Sustainment funds may be used to provide teacher training and support, exam fee subsidies, student study sessions, and mock exams (to be determined by priority by individual schools).

NMSI has significant financial capacity to continue to scale CRP nationally. To date, more than $300 million has been invested in NMSI by the Bill & Melinda Gates Foundation, the Michael and Susan Dell Foundation, IBM, Texas Instruments, ExxonMobil, Northrop Grumman, BAE Systems, Lockheed Martin, the Boeing Company, the Department of Defense Education Activity, the Office of Naval Research, and others. **NMSI does not seek a waiver of the i3 match requirement; our budget reflects a commitment to a 50% match of $9,999,294.** We will leverage the i3 funding with our philanthropic partners. NMSI has secured letters of support for the project from a broad range of philanthropic supporters (Appendix G).

The adequacy of procedures for ensuring feedback and continuous improvement in the operation of the proposed project.

Performance management and continuous improvement are a cornerstone of NMSI's CRP. NMSI plans to implement a continuous improvement process that reflects the need to engage LEA and school-based partners and maintain flexibility to course-correct quickly. To enable continuous improvement, we will:

Use data-driven decision making to refine approach: NMSI's online data management system provides timely, Web-based quality control that allows NMSI to gather and analyze national-, regional-, and school-based data. This includes formative, benchmark, and annual summative data from participating schools. For example, when students begin their AP courses in the fall, NMSI analyzes increases in AP enrollment, broken out by subject, by discipline (math, science, English), by gender, by ethnicity, and by socio-economic status to ensure that all schools are on track for ambitious increases in AP scores across all subgroups. When AP results are released, NMSI analyzes AP scores to calculate percentage increases and to evaluate unexpected outcomes. These data allow NMSI to generate an effective and informative feedback loop that facilitates CRP's constant improvement and targets lagging schools. NMSI program managers act as "account executives" by traveling to the states each month to manage day-to-day implementation. Over time, program managers build the capacity of school-level teams to complete this goal-setting and review process on their own.

Seek feedback at frequent intervals: NMSI surveys AP teachers and mentor teachers at least twice annually to assess a wide range of measures including depth of content knowledge and satisfaction implementing instructional strategies learned during NMSI training sessions.

Implement an ongoing grant compliance structure: NMSI's Grant Compliance Committee meets monthly to assess financial and programmatic compliance. The committee is headed by the CFO. The committee will approve and improve the grant implementation plan, assess metrics, develop action plans for improvement, and communicate implementation progress.

QUALITY OF PROJECT EVALUATION

This section describes UCLA/CRESST's evaluation plan to estimate the impact of NMSI's College Readiness Program (CRP) on selected student outcomes, evaluate the fidelity of program implementation, and examine factors that may be associated with successful implementation. This study consists of four parts. First, we will conduct a randomized cluster trial (RCT) with 20 treatment schools and 20 delayed treatment schools as control schools, **providing an opportunity to generate strong evidence of effectiveness without reservations**

per What Works Clearinghouse requirements.[26] Second, a comparative interrupted time series (CITS) analysis will be performed in which selected student outcomes will be compared before and after treatment implementation for treatment schools, delayed treatment schools, and comparison schools. Beginning with school year 2017–18, initial treatment schools and delayed treatment schools will all be treatment schools, and comparison schools will be selected using propensity score matching techniques. Third, fidelity of implementation of CRP and factors that may relate to successful implementation and selected students will be measured and evaluated for treatment schools. Fourth, an exploratory analysis will investigate the impact of CRP using student-level AP exam data from the states where CRP has been implemented since NMSI's inception. The comparison will be made between CRP schools and comparison schools using propensity score matching techniques and potential outcomes estimation (Rubin 2005).

Framework

The logic model for CRP shows that the CRP intervention intends to improve student outcomes by improving classroom teaching. Thus, the CRESST evaluation design addresses the working hypothesis that CRP implementation will produce intermediate outcomes in the form of improved conditions for learning and teaching and improved instruction. Our working hypothesis

[26] We expect that each school will enroll roughly 250 students per grade, with ~233 total students annually (across grades 9 through 12) participating in at least one STEM-related (math, science, or English) AP course (equivalent to 700 students participating per school over a three-year period as referenced in the Project Design section of application). This assumption is based on past enrollment data from CRP schools.

further asserts that these intermediate outcomes will ultimately lead to improved student outcomes.

The evaluation questions are threefold as conceptualized in Raudenbush and Sadoff (2008):

1. What are the effects of the CRP treatment (*Z*)[27] on student outcomes (*Y*)?
2. What are the effects of the CRP treatment (*Z*) on intermediate outcomes (*Q*)?
3. What is the association between intermediate outcomes (*Q*) and student outcomes (*Y*)?

The first question involves a two-level analysis (students are nested within schools, with schools being randomization level) with binary outcomes (e.g., whether a student takes a STEM-related AP course, or whether a student chooses a STEM-related major in college), which can be readily achieved by fitting a standard hierarchical generalized linear model (HGLM; Raudenbush and Bryk 2002). The key interest of this model is focused on the comparison between CRP schools and comparison schools in terms of student outcomes. However, it is noted that for a binary outcome of whether a student scores 3 or higher on STEM-related AP exams, a three-level HGLM would be more appropriate if there were a large number of STEM-related AP teachers per school. In our hierarchical model building process, we will also include student background characteristics such as gender, race, and ethnicity for examining the differential effects of the treatment.

The second question, whether CRP leads to improved conditions for teaching and learning and classroom instruction, will be examined using a three-level model as Raudenbush and Sadoff (2008) suggest: a standard two-level model with *Z* and *Q* both defined on teachers nested within schools. However, because *Q* is measured with errors, the measurement model for *Q* at Level 1,

27. Z is a treatment indicator, which denotes Trt_k in Equations 1 and 2 in the Appendix.

which takes into account measurement errors in the model, will be additionally specified. In this model, *Q* is used as the outcome (e.g., increased teacher knowledge; increased use of effective instructional strategies; increased content knowledge), which is presumably a continuous scale, and *Z* is the explanatory variable.

The third question, whether the intermediate outcome *Q* is associated with student outcome *Y*, will also be examined using a two-level model, where students are nested within teachers or schools. We assume that the number of STEM-related AP teachers per school is very small (e.g., one or two in mathematics), so teacher level would be considered a Level 2 unit and school flag variables will be included as fixed effects. *Q* is defined at the teacher level (Level 2) as an exploratory variable. In addition, we first will apply the Raudenbush and Sadoff (2008) approach for correcting inferences regarding association between *Q* and *Y* when *Q* is assumed to be measured with error. Because this approach has a couple of limitations in correcting measured error issues in *Q*, we will employ an advanced latent variable modeling approach (LV) in a two-level hierarchical generalized modeling setting.

Research Questions

The evaluation will focus on the research questions outlined below. The first six questions are related to the first aim, estimating the CRP impacts on student outcomes. Research Question 7 deals with efforts to estimate mediation effects of program implementation. Specifically, Research Questions 4, 5, and 6 address moderating effects of student-level variables. Research Questions 8–11 are focused on outcomes from student and teacher surveys and the relationships of such outcomes to the fidelity of implementation.

1. What is the impact of the program on the likelihood that students take STEM-related AP courses?

2. What is the impact of the program on the likelihood that students will achieve a qualifying score of 3 or higher on STEM-related AP exams?
3. What is the impact of introducing the program on postsecondary outcomes of high school students, including matriculation and persistence?
4. What is the impact of the program on school-level rates of obtaining a score of 3 or higher (i.e., a score eligible for college credit) on STEM-related AP exams by gender/race/ethnicity?
5. What is the impact of the program on the likelihood of declaring a STEM-related major among students by gender/race/ethnicity?
6. What is the impact, by gender/race/ethnicity, of the program on stated declaration of a STEM-related major among students who graduated from treatment and comparison schools at the end of the first semester of enrollment in a postsecondary institution?
7. Are variations in program implementation systematically associated with differences in program outcomes?
8. What is the impact of the program on student reports of teacher effectiveness?
9. What is the impact of the program on teachers' self-reported effectiveness?
10. To what extent is NMSI's program implemented with fidelity at the treatment sites?
11. What are the facilitators and barriers to implementation?

Study Design and Statistical Comparisons

To address our series of research questions, we propose to conduct an RCT with delayed treatment design, with schools randomly assigned either to treatment or comparison conditions and teachers within schools following the school assignment. Schools that agree to participate in the study will be randomly assigned to experiment conditions prior to program implementation, augmented with a CITS component. As presented in Figure 1 in Appendix J-6, in the first full implementation year (2016–17), there are three groups of schools: initial treatment group; delayed treatment group; and matched comparison group. The delayed treatment group comprises schools in which implementation of CRP involves a one-year delay.

For the RCT in year 2016–17, comparisons will be made between the treatment schools and delayed treatment schools as control schools. For 2017–18, both initial treatment schools and delayed treatment schools will constitute the treatment group and comparison schools will be selected using propensity score matching techniques.

Table 1 in the Project Evaluation Appendix (Appendix J-6) presents the study design and sampling plan for the four years of the program implementation period. A total of 40 schools from selected geographies and districts will be sampled for the study. At Year 1, 20 schools will be randomly assigned to the treatment group and the remaining 20 schools to the control condition. In Year 2, the delayed treatment schools will be assigned to the treatment group, so the total number of treatment and comparison schools are 40 and 40, respectively. We plan to focus on students in grades 11 and 12, with an average of 250 students per school, assuming that the likelihood of students taking AP course(s) in grade 10 is very low. However, if any 10th grade students take AP course(s), they will be included in our analytic models.

Our analytic models feature two important aspects: (1) The data has a nested structure (i.e., students are nested within schools); and (2) outcomes of interest are binary. We assume that it is rare that a student takes more than one math AP course in each grade and that there are fewer than three AP math teachers in a school. However, we are aware that there are multiple science AP teachers as there are more science AP subjects. Given these assumptions and that we plan to perform separate analyses for math and science, a two-level hierarchical generalized linear model is an appropriate analytic model for both binary outcomes. However, for outcome of likelihood of science AP course scores being 3 or higher, a three-level HGLM would be a possible analytic option. If there are too few science AP teachers in the sampled schools (e.g., two or three), we will analyze the data using two-level HGLM with teacher level being a Level 2 unit including school flag indicators as fixed effects. See Appendix J-6 for detailed statistical models.

During each year of the study period (Years 1, 2, 3, and 4), the CRP impact will be evaluated using CITS design (CITS; Shadish, Cook, and Campbell 2002). In this design, we will examine the change in the program schools' performance using student-level outcome when the program was implemented, benchmarked against the change for a similar set of comparison schools. A CITS involves two comparisons: comparing performance in the program schools before and after the program was implemented, and comparing this change in performance to the change in similar comparison schools. The first comparison (performance in the program schools before and after the program) identifies the program effect by assuming that a change in student outcomes at the time the program was implemented is likely due to the program. For example, if a student's likelihood of taking AP course(s) increases more than we would expect based on pre-program trends after the program is implemented, this would suggest that the program had a positive effect. The second comparison strengthens this analysis by comparing the change in student outcomes in the program schools to the change in student outcomes in other similar schools (comparison schools). As such, we control for changes that happened at the same time as the program by benchmarking the change in the program schools against the change in comparison schools. This method accounts for the effects of any event that affected both the program and comparison schools, assuming that all events influencing the program schools also affected comparison schools in the same way, and that there were no events that affected comparison schools but did not affect the program schools. Comparison schools are selected to be as similar to the program schools as possible in order to maximize the likelihood that these assumptions are met.

Table 2 in Appendix J-6 depicts a framework of statistical comparisons in each year. As noted above, two main statistical comparisons will be performed—CITS analysis and RCT analysis. In Year 1, CITS analysis compares students in grades 11 and 12 in 2014–15 and 2015–16 with the first-year implementation students in 2016–17. Students in 2014–15 and 2015–16 will be served as benchmarking students throughout the study years in CITS analysis. As to RCT analysis, we will perform a comparison between treatment schools and delayed treatment schools. This analysis aims to estimate a causal estimate of CRP impact on student outcome(s). In Years 2, 3, and 4, matched sample comparison analysis will be performed each year separately. CITS analysis in Year 2 will add students in 2017–18 to the Year 1 CITS study sample. Likewise, in Year 3 CITS analysis, students in 2018–19 will be added to the CITS Year 2 sample. Both CITS and RCT analytic models are presented in detail in Appendix J-6.

To comprehensively evaluate the impact of CRP, we will analyze all CRP schools' data with data from non-CRP schools operating in the same states. We will receive comprehensive student-level data from the College Board and will also use Common Core of Data from the U.S. Department of Education (2013) to select comparison schools (see Appendix J-7 for data sharing agreement with the College Board).

Statistical Power Analysis

To estimate power of the planned analysis, we draw on methods described in recent literature on power analysis for binary outcomes in group-randomized trials (Donner and Klar

1996; Spybrook, Raudenbush, Congdon, and Martinez 2009). The power analysis for student participation in STEM-related AP courses determines the minimum detectable effect size (MDES) in participation percentage units. We assume a two-tailed test, with 0.80 power, and a Type I error level of 0.05. Below are the key parameters used in our power analysis:

- **Number of schools.** We assume 20 treatment schools, 20 delayed treatment schools, and 40 not treated comparison schools in impact estimation.
- **Number of students per school.** We assume 250 students in each grade 11 and grade 12, resulting in a total of 500 students per high school.
- **Base participation rate.** For comparison schools, we assume that an average of 4% to 10% of the students would participate in STEM-related AP courses. To approximate the participation rates for STEM-related AP courses from publicly available data, we obtain first the information about the average rate of participation in all AP courses, and second, the information about the ratio of participation in STEM-related courses to participation in all AP courses. According to the data on AP exam participation by state, the average participation rates in the target states ranged from 10.4% to 19.7% (National Science Board 2012, Table 8-12). Based on the data on exam participation counts by course, 36% of AP course participation is in one of the STEM-related courses (College Board 2013b). Lastly, we multiply the two pieces of information to obtain the base participation rate in STEM-related courses, resulting in the range of 4% (10.4 × 0.36) to 7% (19.7 × 0.36). Note that because students who take STEM-related AP courses do not necessarily take AP exams, our power calculation based on the rate of taking STEM-related AP exams gives us more conservative estimates.

- **Intraclass correlation at the school level (ICC_s).** The ICC_s is the proportion of variance in the outcome that lies between schools relative to total variance. It is assumed to range from 0.10 to 0.15 based on previous literature about student achievement.[28]

Table 4 in Appendix J-6 presents the minimum detectable effects in percentage under the scenario, described above. Under the assumption that the true participation rate of the comparison school students is approximately 4%, and the ICC ranges from 0.10 to 0.15, a difference in participation rates of 7% or 11% between treatment and comparison students would be required to ensure at least 80% power of detecting the treatment effect. Thus, the study as proposed is sufficiently powered to identify the expected effect size.

Student performance in STEM-related AP courses is measured by a binary outcome that indicates whether a student obtains a score of 3 or higher on the AP exam. Thus, the power analysis for student performance in STEM-related AP courses uses the same methodology as above. It presents the results in two alternative ways. First, it determines the minimum detectable effect in percentage units. Second, it determines school sample sizes to detect 10% difference between treatment and comparison schools. Below we describe assumptions on key parameters used in the power analysis; the other assumptions remain the same as the above power analysis on student participation.

---

28. For student achievement outcomes, an ICC of 0.10 across schools within districts is in the range based on analysis of large-scale data sets (see, e.g., Bloom, Richburg-Hayes, and Black 2007; Jacob, Zhu, and Bloom 2010; Schochet 2005).

- **Number of students per school.** For purposes of the power analysis only, we *conservatively* assume that 35 to 100 students per high school take STEM-related AP courses across grades 11 and 12 annually.[29]
- **Base rate of qualifying scores.** For comparison schools, we assume that an average of 2% of students will obtain a score of 3 or higher in STEM-related AP exams.[30]

Table 4 in Appendix J-6 presents the minimum detectable effects in percentage under the scenario described above. Under the assumption that the qualifying score rate of the comparison school students is approximately 2% of the total grade 11 and 12 population at a school, and the ICC ranges from 0.10 to 0.15, a difference in qualifying scores rates of 6% to 8% between treatment and comparison students would be required to ensure at least 80% power of detecting the treatment effect. Note, Sherman and Song (2014, 2015) found significant effects for the program with fewer schools than proposed in this study. Thus, the study is sufficiently powered to detect the anticipated size of effect of the program.

Evaluating Fidelity of Program Implementation

The three goals of the implementation evaluation are to 1) provide formative feedback on CRP implementation; 2) measure implementation fidelity; and 3) describe the service contrast between the treatment and comparison schools. Research Questions 8–11 guide the

---

[29] All other aspects of the project design and evaluation assume that ~233 students per year participate in AP exams (700 students across the 3-year program).

30. Based on findings from implementation of NMSI CRP in i3 validation grant (Sherman and Song 2015), an average of 2–3% of students obtained a score of 3 or higher in math or science AP courses.

implementation evaluation. The evaluation will examine the extent to which key components of CRP are implemented as intended at the school level.

The fidelity of implementation study will be part of the overall study design described in Section 3. Administrators, teachers, and students in the treatment schools will be surveyed, and a subset of those participants will be identified for follow-up interviews or classroom observations. Likewise, a similar set of participants in the comparison condition will be surveyed and interviewed to gather information on how the AP program and courses are supported and delivered.

To evaluate fidelity of implementation, we plan to adopt the existing CRP logic model and associated fidelity of implementation measures. The CRP logic model posits that the key components of the intervention are program management, teacher support, student support, and awards. Fidelity indicators have already been developed and field-tested (e.g., Sherman, Darwin, and Stachel 2015). Fidelity will be measured separately for each key component of the intervention and threshold values defined (in collaboration with NMSI) to determine whether the intervention was implemented with fidelity. If not already existing, an implementation fidelity matrix will be developed that links the key components of the intervention to their indicators, the data source, the indicator scoring system, and the implementation threshold values.

The primary data sources for the evaluation data will be surveys administered to school administrators, teachers, and students; individual interviews; focus groups; and classroom observations. We will use the instruments and protocols already developed by NMSI and augment the instruments where necessary. The existing instrument pool includes surveys for administrators, teachers, and students; interview protocols; and classroom observation protocols. Surveys will be administered to participants at all treatment schools, with follow-up interviews and observations conducted in a subset of schools. The schools selected for in-depth interviews and classroom observations will be based on recommendations from NMSI and degree of implementation fidelity (e.g., high and low implementation schools).

Formative feedback on CRP will be provided to NMSI through the fidelity measures and indicator scores on the key components. This information will be used to identify schools for more in-depth examination. For example, administrators and teachers at schools with low fidelity of implementation scores could be interviewed to allow us to better understand any barriers and challenges they are facing in implementing the CRP at their school.

During Phase 1 (pre-implementation), we will adopt, refine, and develop measures and interview protocols and develop a data collection plan for comparison schools. Also during this phase, we will fully develop the fidelity implementation matrix, including refining the key program components, logic model, and indicator set. Indicator scoring and fidelity measures for each component will be designed as needed. During Phase 2 (implementation), data will be collected from the treatment and comparison conditions. Fidelity of implementation will be determined for each school, formative feedback reported to NMSI and participating schools, and measures and indicators refined. Recommendations will be developed with respect to the logic model, indicators, logistic issues, and challenges and barriers to implementation.

Outcomes and Key Variables

The outcomes of interest for this study are measures of students' STEM-related AP course experience that include (a) taking an AP course in math and science; (b) STEM-related AP exam scores of 3 or higher; and (c) STEM-related major in college. In addition, measures from student and teacher survey instruments will be used as intermediate outcomes, consistent with the logic model. Those include student reports of teacher effectiveness and teachers' self-reported effectiveness that are considered as both outcome variables and mediating variables in analytic models.

In addition to outcomes, we plan to include selected student-level variables as covariates in our analytic models. For example, student background characteristics of gender, race, and ethnicity will be included to examine moderating effects of those variables on student outcomes. School background characteristics such as size, demographic composition, and school average of eighth-grade state assessment score in mathematics will be considered to examine contextual effects.

Evaluation Expertise and Resources

As part of UCLA's Graduate School of Education & Information Studies, CRESST conducts rigorous research studies, evaluates educational programs, and promotes the sound use of data for improving education. Recent CRESST endeavors include projects funded by IES (including the National Centers for Assessment and Instructional Technology), NSF, DARPA, ONR, NIH, and the Bill and Melinda Gates Foundation. CRESST has previously served as an independent i3 evaluator.

The evaluation will be led by Dr. Eva Baker, whose biography is included in the Quality of the Management Plan section. Detailed resumes are provided in Appendix F.

The $2,844,500 that is budgeted for the evaluation will be sufficient to complete the evaluation as detailed above.

BIBLIOGRAPHY

ACT, Inc. (2014). *The Condition of College & Career Readiness 2014 - Attainment of College and Career Readiness*. Retrieved from http://www.act.org/research/policymakers /cccr14/readiness.html

Ansari, A., Jedidi, K., & Dube, L. (2002). Heterogeneous factor analysis models: A Bayesian approach. *Psychometrika, 67*, 49-77.

Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis, 29*(1), 30–59.

Borenstein, M., Hedges, L., & Rothstein, H. (2007). Meta-analysis: fixed effect vs. random effects. *Meta-Analysis.com*. Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. Routledge.

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

Chairman's Staff of the Joint Economic Committee. (2012). Calculations using data from the Bureau of Labor Statistics. Employment Projections: 2010-20. Table 1.7 Occupational Employment and Job Openings Data, Projected 2010-2020 and Worker Characteristics, 2010. February 2012. Retrieved from http://bls.gov/emp/.

Charter School Growth Fund. (2011). *Charter School Growth Fund 2011 Annual Report*. Retrieved from http://www.socialimpactexchange.org/sites/www.socialimpactexchange .org/files/2011%20CSGF%20Annual%20Report.pdf

College Board. (2013a). *AP score distributions—All subjects 1993-2013.* Retrieved from http://research.collegeboard.org/programs/ap/data/archived/2013

College Board. (2013b). *Pennsylvania_summary_13* [Excel data file]. Retrieved from http://research.collegeboard.org/programs/ap/data/archived/2013

Cronbach, L. J. (with assistance of Deken, J. E., & Webb, N.) (1976). *Research on classrooms and schools: Formulation of questions, design and analysis.* Occasional Paper of the Stanford Evaluation Consortium, Stanford University.

Dehejia, J. H., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association, 98,* 1053-1062.

Donner, A., & Klar, N. (1996). Statistical considerations in the design and analysis of community intervention trials. *Journal of Clinical Epidemiology, 49,* 435–9.

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*.

.Geiser, Saul & Santelices, Veronica (2004). The Role of Advanced Placement and Honors Courses in College Admissions. *Research & Occasional Paper Series No. CSHE 4.04*, University of California, Berkeley. Retrieved from http://cshe.berkeley.edu/ publications/docs/ROP .Geiser.4.04. pdf.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational researcher*, *5*(10), 3-8.

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.

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

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

Jackson, C. K. (2012). *Do college-prep programs improve long-term outcomes?* NBER Working Paper No. 17589. Cambridge, MA: National Bureau of Economic Research.

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

Jacob, R., Zhu, P., & Bloom, H. (2010). New empirical evidence for the design of group randomized trials in education. *Journal of Research on Educational Effectiveness, 3*(2), 157-198.

Kalaian, S. A. (2003). Meta-analysis methods for synthesizing treatment effects in multisite studies: Hierarchical linear modeling (HLM) perspective. *Practical Assessment, Research & Evaluation, 8*(15).

Kena, G., Aud, S., Johnson, F., Wang, X., Zhang, J., Rathbun, A., Wilkinson-Flicker, S., and Kristapovich, P. (2014). *The Condition of Education 2014*. U.S. Department of Education, National Center for Education Statistics. Washington, DC. Retrieved from http://nces.ed.gov/pubsearch.

Krull, J. L., & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, 23, 418-444.

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research*, 36, 249-277.

Langdon, D., McKittrick, G., Beede, D., Khan, B., and Doms, M. (2011). *STEM: Good Jobs Now and for the Future*. Economics and Statistics Administration, US Department of Commerce; 2012. Retrieved from http://www.esa.doc.gov/sites/default/files/stemfinaly july14_1.pdf

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Lawrence Erlbaum Associates.

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.

McKinsey & Company (2009). *The Economic Impact of the Achievement Gap in American Schools.* McKinsey Social Sector Office Retrieved from http://mckinseyonsociety .com/downloads/reports/Education/achievement_gap_report.pdf

Muthén, B. O., & Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 143-165). Boca Raton, FL: Chapman & Hall/CCM.

National Science Board. (2012). *Science and Engineering Indicators 2012*. Arlington VA: National Science Foundation. Retrieved from http://www.nsf.gov/statistics/seind12/

National Science Board (2014). *Science and Engineering Indicators 2014*. Arlington VA: National Science Foundation. Retrieved from http://www.nsf.gov/statistics/seind14/

Nordmann, A. J., Kasenda, B., & Briel, M. (2012). Meta-analyses: what they can and cannot do. *Swiss Medical Weekly, 142*:w13518**.**

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling*, 18, 161-182.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods, 15*, 209–233.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika, 69*, 167–190.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., & Sadoff, S. (2008). Statistical inference when classroom quality is measured with error. *Journal of Research on Educational Effectiveness, 1,* 138-154.

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.

Rothwell, J (2013). The hidden STEM economy (data download file). *Brookings Institution*. Available at: http://www.brookings.edu/research/reports/2013/06/10-stem-economy-rothwell.

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.

Schochet, P. (2005). *Statistical power for random assignment evaluations of education programs.* Princeton, NJ: Mathematica Policy Research.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton-Mifflin.

Sherman, D., Darwin, M., & Stachel, S. (2015, January). *First-year impacts of the National Math and Science Initiative's College Readiness Program* (NMSI). Paper presented at the International Congress for School Effectiveness and Improvement. Cincinnati, OH.

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.

Spybrook, J., Raudenbush, S. W., Congdon, R., & Martinez, A. (2009). *Optimal design for longitudinal and multilevel research: Documentation for the Optimal Design software.* Ann Arbor: University of Michigan.

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

U.S. Department of Education. (2013). National Center for Education Statistics, *Common Core of Data (CCD), "Public Elementary/Secondary School Universe Survey,"* SY 2012–13 Provisional Version 1a. Available at http://nces.ed.gov/ccd/ccddata.asp

Wood, William, B. Revising the AP Biology Curriculum. *Science* 25. September 2009: 1627-1628.