Pathways to STEM Initiative (PSI): Evaluation Report for an Investing in Innovation (i3) Development Grant

Jonathan Nakamoto
Juan Carlos Bojorquez

May 5, 2017

WestEd.org

WestEd — a national nonpartisan, nonprofit research, development, and service agency — works with education and other communities to promote excellence, achieve equity, and improve learning for children, youth, and adults. WestEd has 15 offices nationwide, from Washington and Boston to Arizona and California, with its headquarters in San Francisco. For more information about WestEd, visit WestEd.org; call 415.565.3000 or, toll-free, (877) 4-WestEd; or write: WestEd / 730 Harrison Street / San Francisco, CA 94107-1242.

Table of Contents

Executive Summary

Clark County School District (CCSD) was awarded an Investing in Innovation (i3) Development grant in 2012 to develop and evaluate the Pathways to STEM Initiative (PSI) in middle and high schools. PSI involved a combination of project-based science, technology, engineering, and math (STEM) coursework; extra-curricular opportunities for students to explore STEM concepts and real-world applications alongside STEM professionals; and teacher professional development and support with emphasis on the needs of students with learning disabilities and English language learners in a project-based environment. The project integrated Project Lead the Way's (PLTW) Gateway To Technology (GTT) curriculum into science classes in the four participating middle schools (i.e., Findlay, Garside, Gibson, and Johnston Middle Schools).

WestEd conducted an external evaluation of the middle school component of PSI and reported the findings from Years 1 and 2 of the study to the U.S. Department of Education as part of the National Evaluation of i3. WestEd's evaluation included an implementation study that was based on the fidelity of implementation of critical program components as outlined in the project's logic model (see Appendix A). Additionally, the evaluation employed a quasi-experimental design with matched comparison schools to examine the impact of PSI on students and teachers. A summary of the program implementation findings during Year 2, impacts on science teachers during Year 2, and impacts on students across Years 1 and 2 are outlined below.

Program Implementation

Overall implementation during Year 2 was classified as emerging to moderate based on the scoring rubric that WestEd and CCSD program staff jointly developed. As shown in Exhibit ES-1, two or more of the PSI schools had emerging or moderate implementation on six of the eight components. Schools generally had high levels of implementation with summer extracurricular activities but struggled reaching similar high levels of implementation in school-year-based extracurricular activities. Schools also had moderate to high levels of implementation of the program training and curricula.

Component 1: Teacher Professional Development. Garside, Gibson, and Johnston had moderate levels of implementation for Component 1. Over two-thirds of the teachers in these three schools participated in all or nearly all of the PLTW Readiness Training Modules and Core Training. However, less than 34% of teachers participated in two optional PD sessions.

Component 2: Implementation of CCSD and PLTW Curricula. Garside and Johnston had high levels of implementation for Component 2. Over two-thirds of the teachers at Garside and Johnston had GTT objectives incorporated into their lesson plans. Further, over two-thirds of the teachers at these two schools had GTT activities observed during lessons in which the activities were scheduled.

Exhibit ES-1. Number of Schools with Emerging, Moderate, and High Implementation on the Eight Program Components

Component	Number of Schools		
	Emerging Implementation	Moderate Implementation	High Implementation
Component 1: Teacher Professional Development	1	3	0
Component 2: Implementation of CCSD and Project Lead the Way Curricula	0	2	2
Component 3: Ongoing Teacher Support	0	3	1
Component 4: Classroom Technology	0	0	4
Component 5: Weekly Sessions with STEM Professionals	3	1	0
Component 6: STEM Club	2	2	0
Component 7: STEM Summer Camp	0	1	3
Component 8: Science and Math Tutoring	2	1	1

Component 3: Ongoing Teacher Support. Findlay received a high level of teacher support from the PSI project coordinator, whereby the coordinator visited and provided support to teachers during more than 70% of school weeks. Garside, Gibson, and Johnston received moderate levels of teacher support. The project coordinator provided support visits during 55% to 68% of the school weeks.

Component 4: Classroom Technology. Findlay, Garside, Gibson, and Johnston had high implementation for Component 4. Nearly all of the laptops used by students and teachers in all four schools were functioning.

Component 5: Weekly Sessions with STEM Professionals. Garside had moderate implementation for Component 5. More than 85% of weekly sessions had at least one teacher participate. Garside also had an average of 13 students attend weekly sessions over the school year. Gibson and Johnston had at least one teacher participate in 26% to 85% of weekly sessions, and average student attendance was fewer than 11 students. At Findlay, less than 26% of weekly sessions had at least one teacher participate, and average student attendance was fewer than 11 students. STEM professionals were not involved in weekly sessions.

Component 6: STEM Club. Gibson and Johnston had moderate implementation for Component 6. More than 75% of STEM Club activities at both schools had at least one teacher participate. Gibson had an average of 14 students participate in STEM Club over the year, and Johnston had an average of six students participate in STEM Club over the year.

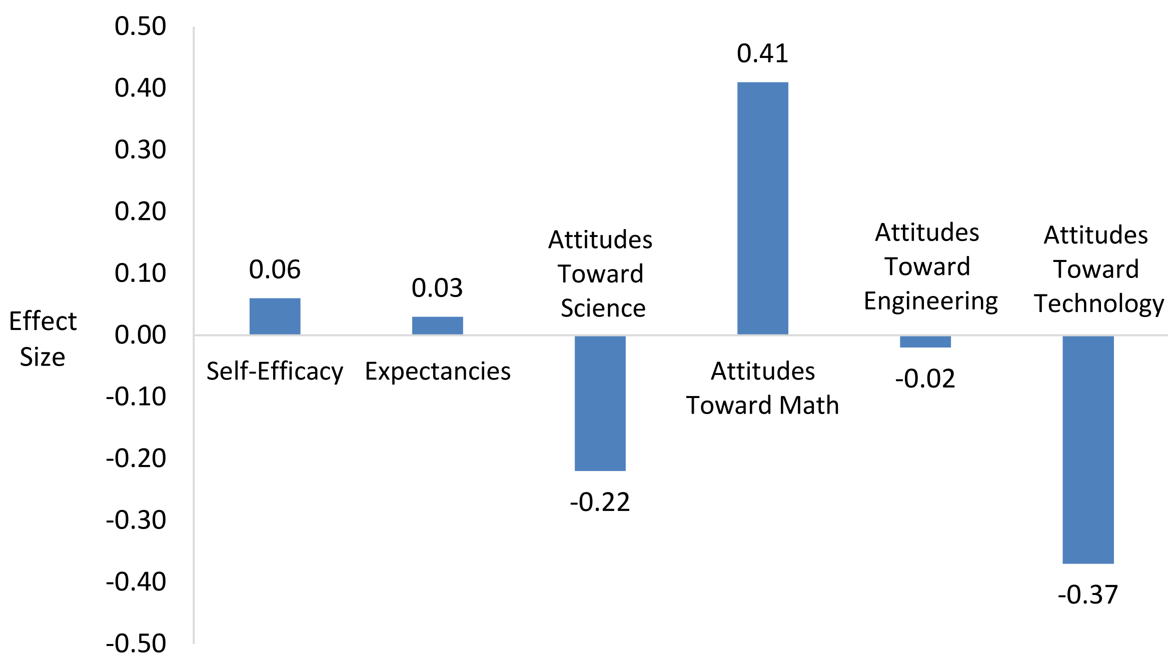
Component 7: STEM Summer Camp. Garside, Gibson, and Johnston had high implementation for Component 7 and Findlay had moderate implementation. Similar to the first year of PSI, only one summer camp was held for students from across the four schools. Student attendance ranged from 29 to 45 students per school.

Component 8: Science and Math Tutoring. Gibson had high implementation for Component 8, and Johnston had moderate implementation. Both math and science tutoring were implemented at Gibson; only science tutoring was implemented at Johnston. When tutoring was implemented, the majority of the tutoring sessions at these two schools had at least one teacher participate and an average of 6–10 students attended the sessions.

Program Impacts on Teacher Outcomes

The impact analyses did not show that PSI had a statistically significant effect on the science teachers' beliefs about science and attitudes toward STEM. However, teachers in PSI and non-PSI schools generally reported very positive attitudes toward science and high self-efficacy for teaching science. Of the six scales on the teacher survey, half favored the treatment group and the other half favored the comparison group (see Exhibit ES-2). The effect sizes indexing the differences between the treatment and comparison teachers ranged from -0.37 to 0.41. The small sample size for the teacher analyses meant that effect sizes in the small to moderate range, which would be statistically significant with larger samples, were not statistically significant.

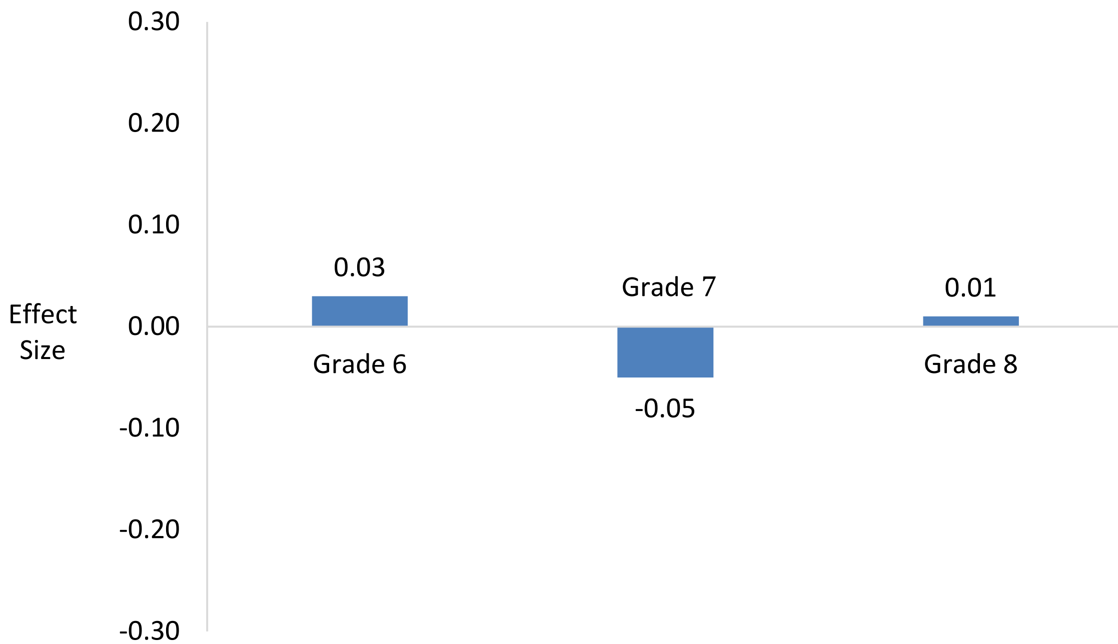
Exhibit ES-2. Effect Sizes from the Teacher Impact Analyses



Program Impacts on Student Outcomes

The impact analyses did not show that PSI had a statistically significant effect on the students' science achievement. There was no significant effect on 6th grade students who participated in PSI for one year. There was also no significant effect on 7th and 8th grade students who participated in PSI for two years. On the Partnership for the Assessment of Standards-Based Science (i.e., the PASS Assessment), the treatment students in grade 6 correctly completed 0.18 more items than the comparison students and the treatment students in grade 7 correctly completed 0.29 fewer items than the comparison students. The effect sizes indexing these differences were small (see Exhibit ES-3). On the grade 8 science CRT, the treatment students' scale scores were similar to the comparison students' scores (299.86 and 299.16 respectively) and had a small effect size (*ES* = 0.01). Finally, exploratory subgroup analyses by grade level for gender, English learner status, and ethnicity revealed no consistent positive impact of PSI participation for any subgroup.

Exhibit ES-3. Effect Sizes from the Student Impact Analyses



Introduction and Evaluation Overview

In 2012, Clark County School District (CCSD) was awarded a three-year Investing in Innovation (i3) Development grant to develop and evaluate the Pathways to STEM Initiative (PSI) in middle and high schools. PSI involved a combination of project-based science, technology, engineering, and math (STEM) coursework; extra-curricular opportunities for students to explore STEM concepts and real-world applications alongside STEM professionals; and teacher professional development and support with emphasis on the needs of students with learning disabilities and English language learners in a project-based environment.

During Years 1 and 2, the project integrated Project Lead the Way's (PLTW) Gateway To Technology (GTT) curriculum into all middle school science classes in the four participating schools. Middle school teachers received professional development and ongoing support in implementing GTT curricula while implementing the GTT curricula in their science classes. In Year 3, implementation at grade 6 was modified in that GTT was offered as an elective course, while in grades 7 and 8 GTT continued to be integrated into all science classes in participating schools. Additionally, CCSD's PSI project included some elective PLTW courses and internship, mentoring, and shadowing experiences at the high school level.

WestEd conducted a formative and summative evaluation focused on the middle school component of the program given the project's approach of comprehensive implementation and exposure at this level. Formative interim reports were submitted annually to CCSD during implementation. WestEd's external evaluation included an implementation study based on the fidelity of implementation of critical program components as outlined in the project's logic model (see Appendix A). The evaluation also employed a quasi-experimental design with matched comparison schools to examine the overall impact of PSI on students and science teachers. The current report includes the student impact and implementation findings from Years 1 and 2 that WestEd reported to the U.S. Department of Education as part of the National Evaluation of i3.

The impact study was based on student achievement as measured by Nevada's Criterion Referenced Test (CRT) in grade 8 and by the Partnership for the Assessment of Standards-Based Science (PASS) in grades 6 and 7. Impact analyses were conducted at grade 6 for students who participated in Year 2 of PSI and at grades 7 and 8 for students who participated during Years 1 and 2 of PSI. The impact study also included an examination the program's impact on teachers' STEM-related attitudes, preparedness to implement curricula, and efficacy teaching science. The teacher analyses were conducted on the sample of teachers who participated in PSI during Year 2.

Implementation study evaluation research questions addressed in this report included:

- How and to what extent were the critical components of PSI implemented?
- To what extent did students and teachers participate in project activities?
- What was the overall level of fidelity of implementation?

Impact study evaluation research questions addressed in this report included:

- Did one year of participation in teacher-related PSI activities have a positive impact on teachers' science attitudes?
- Did one year of participation in middle school PSI activities have a positive impact on 6th grade students' science achievement?
- Did two years of participation in middle school PSI activities have a positive impact on 7th and 8th grade students' science achievement?

We present an overview of the study methodology, including sample demographics, participation rates, analysis plan, and baseline comparisons of treatment and comparison schools. This is followed by a discussion of the findings on implementation and the impact analyses based on the student achievement measures and the teachers' science attitudes.

Method

This section of the report outlines the methodology used for the evaluation. We first describe the program implementation measures used to assess whether the PSI program was implemented as planned. This section also describes surveys administered to teachers, student achievement measures, and student and teacher demographic data used in the evaluation. The statistical analyses are reviewed in this section of the report with additional detail in the Appendix. Finally, the teachers and students in the treatment and comparison schools are compared prior to the start of the intervention.

Program Implementation Measures

Prior to implementation of PSI, WestEd collaborated with CCSD program staff and Project Lead the Way (PLTW) program developers to identify the following eight critical components of PSI: 1) Teacher Professional Development, 2) Implementation of CCSD and PLTW Curricula, 3) Ongoing Teacher Support, 4) Classroom Technology, 5) Weekly Sessions with STEM Professionals, 6) STEM Club, 7) STEM Summer Camp, and 8) Science and Math Tutoring. These components were used to construct a measure for fidelity of program implementation. Each component had between one and four subcomponents. These components and their subcomponents are articulated in the PSI logic model (see Appendix A).

In order to create a well-documented and quantitative method to measure the implementation of the program, WestEd and CCSD program staff jointly developed a scoring rubric for the components and subcomponents. WestEd and CCSD followed the i3 evaluation technical assistance providers' guidance for measuring fidelity of implementation when creating the scoring rubric (Abt, 2013). The scoring rubric was used in Years 1 and 2 of the evaluation. For each subcomponent, the scoring rubric produced scores of 0, 1, or 2 for teachers and schools. To calculate the score for each component, we summed the school-level scores for each subcomponent. The scores for each component ranged from 0 to 8, depending on the number of subcomponents, and allowed us to classify each treatment school as having emerging, moderate, or high implementation for the eight components. Thresholds for emerging, moderate, or high were identified separately for each component. Finally, we calculated a sample-level measure of implementation for each component that categorized the entire sample as having emerging/moderate or high implementation.

Component 1: Teacher Professional Development

The scoring rubric for the four subcomponents of Component 1: Teacher Professional Development is shown in Exhibit 1. The data for each subcomponent were collected for each participating teacher in the four treatment schools. For example, the number of PLTW training modules completed was counted for each teacher. Scores were given to teachers based on the number of modules they completed (e.g., teachers who completed 4–5 modules received a score of 2). Next, school-level scores were calculated based on the percentage of teachers in each school that received a score of 2. A school received a score of 2 if greater than 66% of teachers at that school received a teacher-level score of 2. The school-level scores for the four subcomponents were summed to calculate the school-level score for the full component. Schools with scores of 5 or more on the full component were classified as having high implementation. Finally, the percentage of schools with high levels of implementation was calculated to obtain the sample-level score.

Exhibit 1. Scoring Rubric for Component 1: Teacher Professional Development

Subcomponent	Data Source	Teacher-Level Score	School-Level Score
1) Participation in PLTW Readiness Training Modules	Online sign-in and tracking system	0 = 0–1 modules completed 1 = 2–3 modules completed 2 = 4–5 modules completed	0 = Less than 34% of teachers at level 2 1 = 34–66% of teachers at level 2 2 = Greater than 66% of teachers at level 2
2) Participation in PLTW Core Training	Attendance sheets collected by PSI coordinator	0 = An average of less than 3 days a week 1 = An average of 3–4 days a week 2 = An average of more than 4 days a week	0 = Less than 34% of teachers at level 2 1 = 34–66% of teachers at level 2 2 = Greater than 66% of teachers at level 2
3) Participation in PLTW Ongoing Training - Virtual Academy	Online sign-in and tracking system for each session	0 = Not registered for Virtual Academy 1 = Registered for Virtual Academy	0 = Less than 51% of teachers at level 1 1 = 51% or more teachers at level 1
4) Attendance at Optional Clark County School District Professional Development (PD) Sessions	Attendance sheets collected by PSI coordinator	0 = Did not attend 1 = Attended one PD session 2 = Attended two PD sessions	0 = Less than 34% of teachers at level 2 1 = 34–66% of teachers at level 2 2 = Greater than 66% of teachers at level 2
School-Level Score for the Full Component (Sum of Scores for the 4 Subcomponents)	**Emerging implementation = Scores of 0–2** **Moderate implementation = Scores of 3–5** **High implementation = Scores of 6–7**		
Sample-Level Score	**Emerging/Moderate implementation = Less than 50% of schools have high levels of implementation of teacher PD** **High implementation = 50% or more of schools have high levels of implementation of teacher PD**		

Component 2: Implementation of Clark County School District (CCSD) and Project Lead the Way (PLTW) Curricula

The scoring rubric for the two subcomponents of Component 2: Implementation of CCSD and PLTW curricula, Gateway to Technology (GTT), is outlined in Exhibit 2.

Exhibit 2. Scoring Rubric for Component 2: Implementation of Clark County School District and Project Lead the Way Curricula

Subcomponent	Data Source	Teacher-Level Score	School-Level Score
1) Preparation of lesson plans aligned to the GTT and CCSD benchmark calendar	Lesson plans posted on the CCSD curriculum planning website	0 = GTT objectives included in less than 19% of lesson plans 1 = GTT objectives included in 19–25% of lesson plans 2 = GTT objectives included in greater than 25% of lesson plans	0 = Less than 34% of teachers at level 2 1 = 34–66% of teachers at level 2 2 = Greater than 66% of teachers at level 2
2) Classroom observations showed integration of GTT activities prescribed in GTT benchmark calendar and based on teacher lesson plans	Classroom observations conducted by PSI project coordinator	0 = GTT activities observed in less than 19% of lessons 1 = GTT activities observed in 19–25% of lessons 2 = GTT activities observed in greater than 25% of lessons	0 = Less than 34% of teachers at level 2 1 = 34–66% of teachers at level 2 2 = Greater than 66% of teachers at level 2
School-Level Score for the Full Component (Sum of Scores for the 2 Subcomponents)	**Emerging implementation = Scores of 0–1** **Moderate implementation = Scores of 2–3** **High implementation = Score of 4**		
Sample-Level Score	**Emerging/Moderate implementation = Less than 75% of schools have high levels of implementation of CCSD and PLTW curricula** **High implementation = 75% or more of schools have high levels of implementation of CCSD and PLTW curricula**		

Note: A greater number of classroom observations were scheduled during quarters when more GTT activities were planned based on CCSD's schedule for science lessons.

The PSI project coordinator reviewed and scored the lesson plans for subcomponent 1, and the PSI project coordinator and other staff from CCSD's Instructional Design and Professional Learning Division conducted classroom observations for subcomponent 2. The data for both subcomponents were collected for each participating teacher in treatment schools. For example, lesson plans posted on the CCSD curriculum planning website were reviewed to determine the extent to which lesson plans used by treatment teachers were aligned to the GTT/CCSD benchmark calendar. Scores were given to teachers based on the percentage of their lesson plans that included GTT objectives. Next, school-level scores were calculated based on the percentage of teachers in each school that received a score of 2. A school received a score of 2 if greater than 66% of teachers at that school received a teacher-level score of 2. The school-level scores for the two subcomponents were summed to calculate the school-level score for the full component. Schools with a full component score of 4 were classified as having high implementation. In addition, the percentage of schools with high levels of implementation was calculated to obtain the sample-level score.

Component 3: Ongoing Teacher Support

The scoring rubric for Component 3: Ongoing Teacher Support is shown in Exhibit 3. Ongoing Teacher Support had one subcomponent and was measured for each school as opposed to for each teacher and school, as was done for Components 1 and 2. Data for the amount of weekly on-site support from the PSI project coordinator were obtained through a review of the project coordinator's calendar. We calculated the school-level scores based on the percentage of weekly visits that occurred. The project coordinator scheduled school visits to take place each week during the school year and a school received a score of 2 if greater than 70% of the scheduled weekly visits occurred. A score of 2 was classified as high implementation and the percentage of schools with high levels of implementation was tallied to obtain the sample-level score.

Exhibit 3. Scoring Rubric for Component 3: Ongoing Teacher Support

Subcomponent	Data Source	School-Level Score
1) Weekly on-site support from PSI project coordinator	Project coordinator's electronic work calendar	0 = Less than 50% of weekly visits occurred 1 = 50–70% of weekly visits occurred 2 = Greater than 70% of weekly visits occurred
School-Level Score for the Full Component	**Emerging implementation = Score of 0** **Moderate implementation = Score of 1** **High implementation = Score of 2**	
Sample-Level Score	**Emerging/Moderate implementation = Less than 75% of schools have high levels of ongoing teacher support** **High implementation = 75% or more of schools have high levels of ongoing teacher support**	

Component 4: Classroom Technology

Exhibit 4 shows the scoring rubric for Component 4: Classroom Technology. The number of available laptops for students and teachers in each school constituted the single subcomponent for the Classroom Technology component. The data were obtained from quarterly computer status reports completed by the PSI project coordinator and site-based information technology staff. The school-level scores were based on the percentage of student and teacher computers that were functioning. A school received a score of 2 if greater than 75% of the computers were functioning. A score of 2 was classified as high implementation, and we calculated the percentage of schools with high levels of implementation to create the sample-level score.

Exhibit 4. Scoring Rubric for Component 4: Classroom Technology

Subcomponent	Data Source	School-Level Score
1) 108 student laptops are available for students and 9 laptops are available for teachers	Quarterly computer status report	0 = Less than 50% of computers were functioning 1 = 50–75% of computers were functioning 2 = Greater than 75% of computers were functioning
School-Level Score for the Full Component	**Emerging implementation = Score of 0** **Moderate implementation = Score of 1** **High implementation = Score of 2**	
Sample-Level Score	**Emerging/Moderate implementation = Less than 75% of schools have high levels of implementation of classroom technology** **High implementation = 75% or more of schools have high levels of implementation of classroom technology**	

Component 5: Weekly Sessions with STEM Professionals

Exhibit 5 displays the scoring rubric for the three subcomponents of Component 5: Weekly Sessions with STEM Professionals. The data for the three subcomponents were collected for each participating treatment school. For example, we measured the extent of teacher participation in each of the weekly sessions using extra-duty time sheets submitted by participating teachers. A school received a score of 2 if at least one teacher participated in greater than 85% of the weekly sessions. The school-level scores for the three subcomponents were summed to calculate the school-level score for the full component. Schools with scores of 4–6 across subcomponents were classified as having high implementation. In addition, to create the sample-level score we calculated the percentage of schools with high levels of implementation.

Exhibit 5. Scoring Rubric for Component 5: Weekly Sessions with STEM Professionals

Subcomponent	Data Source	School-Level Score
1) One teacher participates in the weekly sessions	Extra-duty time sheets submitted by participating teachers	0 = One teacher participated in less than 26% of weekly sessions 1 = One teacher participated in 26–85% of weekly sessions 2 = One teacher participated in greater than 85% of weekly sessions
2) STEM professionals participate in weekly sessions	Quarterly Support Calendar and name of STEM Professional submitted by participating teachers	0 = Less than 60% of weekly sessions were supported by a STEM Professional 1 = 60–75% of weekly sessions were supported by a STEM Professional 2 = Greater than 75% of weekly sessions were supported by a STEM Professional
3) Up to 30 students participate in weekly sessions	Attendance sheets submitted by participating teachers	0 = An average of less than 11 students per weekly session 1 = An average of 11–20 students per weekly session 2 = An average of more than 20 students per weekly session
School-Level Score for the Full Component (Sum of Scores for the 3 Subcomponents)	**Emerging implementation = Scores of 0–2** **Moderate implementation = Score of 3** **High implementation = Scores of 4–6**	
Sample-Level Score	**Emerging/Moderate implementation = Less than 50% of schools have high levels of implementation for the weekly sessions** **High implementation = 50% or more of schools have high levels of implementation for the weekly sessions**	

Component 6: STEM Club

The scoring rubric for the two subcomponents of Component 6: STEM Club is shown in Exhibit 6. The data for the two subcomponents were collected for each participating treatment school. For example, we measured extent of teacher participation in STEM Club activities using extra-duty time sheets submitted by participating teachers. A school received a score of 2 if at least one teacher participated in greater than 75% of STEM Club activities. We summed school-level scores for the two subcomponents to create the school-level score for the full component. Schools with full component scores of 4–6 were classified as having high implementation. In addition, to create the sample-level score we calculated the percentage of schools with high levels of implementation.

Exhibit 6. Scoring Rubric for Component 6: STEM Club

Subcomponent	Data Source	School-Level Score
1) One teacher participates in STEM Club activities	Extra-duty time sheets submitted by participating teachers	0 = One teacher participated in less than 60% of STEM Club activities 1 = One teacher participated in 60–75% of STEM Club activities 2 = One teacher participated in greater than 75% of STEM Club activities
2) Up to 40 students participate in STEM Club activities	Attendance sheets submitted by participating teachers	0 = An average of less than 11 students per STEM Club activity 1 = An average of 11–20 students per STEM Club activity 2 = An average of more than 20 students per STEM Club activity
School-Level Score for the Full Component (Sum of Scores for the 2 Subcomponents)	**Emerging implementation = Scores of 0–1** **Moderate implementation = Scores of 2–3** **High implementation = Score of 4**	
Sample-Level Score	**Emerging/Moderate implementation = Less than 50% of schools have high levels of implementation for the STEM Club** **High implementation = 50% or more of schools have high levels of implementation for the STEM Club**	

Note: The school-level scores for subcomponent 2 in Year 1 were: 0 = An average of less than 21 students per STEM Club activity; 1 = An average of 21–30 students per STEM Club activity; 2 = An average of more than 30 students per STEM Club activity.

Component 7: STEM Summer Camp

The scoring rubric for the three subcomponents of Component 7: STEM Summer Camp is shown in Exhibit 7. Data for the three subcomponents were collected for each participating treatment school. The original plan for the implementation of the PSI program was to have one summer camp at each school. However, given the low enrollment during Year 1, this component was modified to hold one summer camp for all four schools. We measured whether the nine planned STEM Summer Camp activities were implemented as scheduled. All four schools received a score of 2 if greater than 80% of the planned activities were implemented. The other two subcomponents were based on the number of teachers and students that participated in the summer camp and were calculated individually for each school. We summed the school-level scores for the three subcomponents to create the school-level score for the full component. Schools with full component scores of 4–6 were classified as having high implementation. In addition, to create the sample-level score we calculated the percentage of schools with high levels of implementation.

Exhibit 7. Scoring Rubric for Component 7: STEM Summer Camp

Subcomponent	Data Source	School-Level Score
1) Nine planned STEM Summer Camp Activities are implemented	Summer Camp Activity Log	0 = Less than 80% of planned STEM activities were implemented 1 = 80% of planned STEM activities were implemented 2 = Greater than 80% of planned STEM activities were implemented
2) Two teachers from each school participate in the STEM Summer Camp	Extra-duty time sheets submitted by participating teachers	0 = 1 teacher per school participated in the summer camp 1 = 2 teachers per school participated in the summer camp 2 = 3 or more teachers per school participated in the summer camp
3) Number of students attending STEM Summer Camp	Attendance sheets submitted by participating teachers	0 = Fewer than 30 students per school attended the summer camp 1 = 30–40 students per school attended the summer camp 2 = More than 40 students per school attended the summer camp
School-Level Score for the Full Component (Sum of Scores for the 3 Subcomponents)	**Emerging implementation = Scores of 0–2** **Moderate implementation = Score of 3** **High implementation = Scores of 4–6**	
Sample-Level Score	**Emerging/Moderate implementation = Less than 50% of schools have high levels of implementation for the STEM Summer Camp** **High implementation = 50% or more of schools have high levels of implementation for the STEM Summer Camp**	

Component 8: Science and Math Tutoring

Exhibit 8 displays the scoring rubric for the four subcomponents of Component 8: Science and Math Tutoring. The data for the four subcomponents were collected for each participating treatment school. For example, we measured whether at least one teacher participated in each of the science tutoring sessions using extra-duty time sheets submitted by participating teachers. A school received a score of 2 if one teacher participated in greater than 85% of the tutoring sessions. The school-level scores for the four subcomponents were summed to calculate the school-level score for the full component. Schools with full component scores of 6–8 were classified as having high implementation. In addition, to create the sample-level score we calculated the percentage of schools with high levels of implementation.

Exhibit 8. Scoring Rubric for Component 8: Science and Math Tutoring

Subcomponent	Data Source	School-Level Score
1) One teacher participates in the Science Tutoring	Extra-duty time sheets submitted by participating teachers	0 = One teacher participated in less than 26% of tutoring sessions 1 = One teacher participated in 26–85% of tutoring sessions 2 = One teacher participated in greater than 85% of tutoring sessions
2) One teacher participates in Math Tutoring	Extra-duty time sheets submitted by participating teachers	0 = One teacher participated in less than 26% of tutoring sessions 1 = One teacher participated in 26–85% of tutoring sessions 2 = One teacher participated in greater than 85% of tutoring sessions
3) Number of students participating in Science Tutoring	Attendance sheets submitted by participating teachers	0 = An average of fewer than 6 students per tutoring session 1 = An average of 6–10 students per tutoring session 2 = An average of more than 10 students per tutoring session
4) Number of students participating in Math Tutoring	Attendance sheets submitted by participating teachers	0 = An average of fewer than 6 students per tutoring session 1 = An average of 6–10 students per tutoring session 2 = An average of more than 10 students per tutoring session
School-Level Score for the Full Component (Sum of Scores for the 4 Subcomponents)	**Emerging implementation = Scores of 0–2** **Moderate implementation = Scores of 3–5** **High implementation = Scores of 6–8**	
Sample-Level Score	**Emerging/Moderate implementation = Less than 50% of schools have high levels of implementation for the Science and Math Tutoring** **High implementation = 50% or more of schools have high levels of implementation for the Science and Math Tutoring**	

Comparison School Identification

At the start of the evaluation, WestEd used a multivariate matching algorithm to select the comparison schools for the study (see Appendix B for additional details on the matching procedure). The inclusion of comparison schools in the study allowed us to contrast the treatment schools' outcomes with outcomes at similar schools that did not participate in PSI.

The comparison schools provided an estimate for what would have happened in the treatment schools without PSI. The matching procedure utilized the schools' science, math, and reading CRT scores, student transiency rate, percentage of students qualifying for free or reduced-price lunch (FRL), and percentage of White, Black/African American, and Hispanic/Latino students from the 2011–12 school year. We created a pool of middle schools in CCSD that were not participating in PSI and selected two non-PSI schools for each of the four PSI schools. Due to resource limitations and our goal of limiting the data collection burden on the schools, only teachers in these 12 schools were included in the survey administration. Additionally, students in grades 6 and 7 in these 12 schools completed an independent science assessment (i.e., the PASS Assessment) developed by WestEd's Science, Technology, Engineering, & Mathematics (STEM) Program because there was no science CRT for those two grades. For the analysis of the CRT data, we selected two additional comparison schools for each treatment school so that the analyses included a total of four PSI schools and 16 non-PSI schools. Following the matching, we compared the CRT scores and the demographic variables for the treatment and comparison schools and found that the matching process produced a group of comparison schools that was equivalent to the treatment schools prior to the start of the intervention.

Surveys

Fall and Spring Teacher Survey

During Year 2, the teachers completed paper and pencil surveys in September 2014 (i.e., the fall survey) and in May–June 2015 at the end of the school year (i.e., the spring survey). The teachers received their surveys along with the packets of student surveys and completed their surveys at the same time as their students.

Response Rates

Overall, 44 of the 95 science teachers in the four treatment and eight comparison schools consented to participate in the study and completed a survey in the fall and spring, which produced a response rate of 46.3%. The response rate was higher for the treatment (55.9%) schools than the comparison (41.0%) schools. The response rates varied considerably across schools and ranged from 11.1% to 100.0%.

Science Teaching Outcome Expectancy Scale Instrument (STEBI)

We included a 12-item efficacy beliefs scale from the Science Teaching Outcome Expectancy scale instrument (STEBI; Riggs & Enochs, 1989). One item from the original efficacy beliefs scale was not included in the current survey. The current survey also included a 12-item expectancies scale from the STEBI that assessed teachers' beliefs about the efficacy of science instruction. The efficacy beliefs items (e.g., "I am continually finding better ways to teach science.") and the expectancy items (e.g., "If students are underachieving in science, it is most likely due to ineffective science teaching.") were rated on a 1 (*strongly disagree*) to 5 (*strongly agree*) scale. The scales showed good reliability in the fall and spring with Cronbach's alphas ranging from $\alpha = .78$ to $\alpha = .83$.

STEM Semantics Survey

The teacher survey contained four scales assessing general attitudes and perceptions of science, math, engineering, and technology from the STEM Semantics Survey (STEM-T; Tyler-Wood, Knezek, & Christensen, 2010). The scales had teachers rate where they felt science, math, engineering, and technology were on seven-point scales with five adjective pairs (e.g., 1 [*boring*] to 7 [*interesting*]) as anchors on the scales. The reliability of the scales in the fall and spring was good to excellent and the Cronbach's alphas ranged from $\alpha = .77$ to $\alpha = .93$ for all of the scales except the spring science scale. The Cronbach's alpha for the spring science scale was low ($\alpha = .59$) in part because nearly all of the teachers reported the highest level of attitudes to science for several of the items and there was little variability in their responses.

Student Surveys

As part of the evaluation, the science teachers administered a paper and pencil survey to the students who did not opt out of the study during their science classes in September 2014 prior to the start of the intervention (i.e., the fall survey) and in May 2015 at the end of the school year (i.e., the spring survey). The survey included items from the Test of Science-Related Attitudes (TOSRA; Fraser, 1981), the STEM Semantics Survey for students (STEM-S; Tyler-Wood et al., 2010), and the School Attitude Assessment Survey-Revised (SAAS-R; McCoach & Siegle, 2003). The fall survey response rate was 61.0% and the spring response rate was 72.7%. WestEd used data from these fall and spring surveys for an internal report. However, CCSD's Institutional Review Board determined that the findings from the student survey could not be made available publicly because the evaluation team used a passive consent process.

Student Achievement Measures

Partnership for the Assessment of Standards-Based Science (PASS) Assessment

WestEd's STEM program developed science tests for the 6th and 7th grade using multiple choice items from the PASS because the state did not administer a science assessment in these two grades. WestEd's STEM program reviewed CCSD's Science Benchmark Calendar to align PASS items with NDE's Science Standards.

Since the inception of the PASS Assessment in 1996, students in 463 districts in 27 states and Puerto Rico have completed the exam. PASS Assessments include valid and reliable items that are aligned to national science standards (i.e., National Science Education Standards and Benchmarks for Science Literacy). A team of grade-level teachers, science specialists/supervisors, scientists, and measurement specialists, including staff from WestEd's STEM program, developed the items on the PASS. Additionally, an advisory board consisting of individuals from three states oversaw the development of the PASS. This assessment was developed with the support of a National Science Foundation (NSF) grant.

The PASS Assessment was administered by the teachers during April and May of 2015 and used to assess science achievement after participation in PSI. Given resource limitations and our goal of limiting the data collection burden on the comparison schools, the PASS Assessment was administered only in the eight comparison schools that completed the survey in addition to the four treatment schools. The tests for both grades contained 30 multiple choice items. We used the Kuder-Richardson formula 20 (KR-20; McDonald, 1999) to assess the reliability of the PASS Assessment because the items were binary (correct = 1; incorrect = 0). The tests showed good reliability in grade 6 (KR-20 = .83) and grade 7 (KR-20 = .78). We used the raw number of correct responses as the outcome variable in our impact analyses.

Nevada Criterion Referenced Tests (CRTs)

WestEd used the Nevada science CRT, which was administered April 20–May 15, 2015 to students in grade 8, as an outcome variable in our impact analyses. For the PASS analysis with the students in grade 6, we used the science, math, and reading CRTs administered in the spring of 2014 as covariates to adjust for preexisting differences between the treatment and comparison groups. In 2014, 153 of the elementary schools in CCSD piloted the Smarter Balanced Assessment Consortium (SBAC) tests, which resulted in a higher rate of missing data for the CRTs than in previous years. We used the science, math, and reading CRTs administered in spring 2013 as covariates in the PASS analysis with the students in grade 7. In addition, we used the math and reading CRTs from spring 2013 as covariates in the grade 8 science CRT analysis. The use of the CRTs from 2013 as covariates for the grade 7 and 8 analyses allowed us to assess the impact of the intervention after two years of participation. The math, reading, and science CRTs exhibited high reliability (Cronbach's alphas ranged from $\alpha = .89$ to $\alpha = .92$) in prior testing administrations (Nevada Department of Education, 2012). We used the tests' scale scores that ranged from 100 to 500 in our analyses.

Student and Teacher Demographic Data

CCSD provided data on the students' race/ethnicity, English language learner (ELL) status, individualized education program (IEP) status, and gender from the 2013–14 school year. The district could not provide student-level FRL data. The students' race/ethnicity was categorized as American Indian/Alaskan Native, Asian, Black/African American, Hispanic/Latino, Multi-Racial/Multi-Ethnic, Native Hawaiian/Pacific Islander, or White. To include race/ethnicity as a statistical control in our analyses, we combined the two smallest categories (i.e., American Indian/Alaskan Native and Native Hawaiian/Pacific Islander) and created five dummy-coded variables that contrasted each racial/ethnic group with Hispanic/Latino (i.e., the largest racial/ethnic group). In addition, we created dummy-coded variables for ELL status (i.e., ELL = 1; non-ELL = 0), IEP status (i.e., IEP = 1; non-IEP = 0), and gender (i.e., male = 1; female = 0).

Exhibit 9. Demographic Characteristics of the Students in Grades 6–8 in the Treatment and Comparison Schools

	Treatment Schools (n = 4)		PASS Comparison Schools (n = 8)		CRT Comparison Schools (n = 16)	
	n	%	n	%	n	%
Race/Ethnicity						
American Indian/Alaskan Native	21	0.4%	44	0.5%	69	0.4%
Asian	204	3.9%	445	4.7%	750	4.0%
Black/African American	1,256	23.8%	1,463	15.4%	2,697	14.5%
Hispanic/Latino	2,881	54.6%	5,782	61.0%	11,929	64.0%
Multi-Race/Multi-Ethnic	245	4.6%	412	4.3%	748	4.0%
Native Hawaiian/Pacific Islander	68	1.3%	136	1.4%	200	1.1%
White	601	11.4%	1,204	12.7%	2,240	12.0%
Gender						
Female	2,578	48.9%	4,486	47.3%	9,045	48.5%
Male	2,698	51.1%	5,000	52.7%	9,588	51.5%
ELL Students	**953**	**18.1%**	**2,122**	**22.4%**	**4,433**	**23.8%**
Students with an IEP	**641**	**12.1%**	**1,133**	**11.9%**	**2,099**	**11.3%**

Note: ELL = English language learner. IEP = individualized education program.

The demographic characteristics of the students in grades 6–8 in the treatment and comparison schools on the fall count day are shown in Exhibit 9. The demographic characteristics of the treatment and comparison groups were fairly well-matched. The treatment schools had slightly more Black/African American students and slightly fewer Hispanic/Latino students than the comparison group. The comparison schools had slightly more ELL students and had comparable proportions of students with an IEP. Additionally, the demographic characteristics of the students in each grade level were consistent with the percentages shown in Exhibit 10 for the full sample of students in grades 6–8.

CCSD provided the participating teachers' employment start date with the district. We used these data to calculate the teachers' years of experience working in CCSD, which was a proxy for teaching experience. The treatment teachers (M = 8.90, SD = 6.06, n = 19) had less experience in CCSD than the comparison teachers (M = 11.80, SD = 8.33, n = 25) and the effect size indexing this difference was -0.39. This variable was used as a covariate in our impact analyses with the teacher survey data to statistically control for preexisting differences between the groups.

Identification of Eligible Students, Crossovers, and Attrition

All students in grades 6–8 in the four treatment schools participated in PSI and were eligible for the study. Similarly, all students in grades 6–8 in the comparison schools were eligible for the study. We defined the sample as all students in those three grades who were enrolled on the fall count days in Year 1 and 2. In Year 1, this included 5,299 students in the treatment schools and 18,569 students in the 16 comparison schools. In Year 2, this included 5,276 students in the treatment schools and 18,633 students in the 16 comparison schools. Students who entered the participating schools after the fall count day were excluded from the analyses because the students who entered the treatment schools did not receive the full intervention.

The attrition rates for the PASS Assessment and science CRT are shown in Exhibit 10. The attrition rates were calculated by dividing the number of students who did not have outcome, pre-test, and demographic data by the total number of students enrolled on the fall count day. The attrition rates for the PASS Assessment in grades 6 and 7 were between 29% and 42% for the treatment and comparison groups. For the PASS Assessment, the attrition resulted from students not completing the PASS Assessment and/or the CRTs and students leaving participating schools. The attrition for grade 6, which required the students to stay in the study for only one year, resulted from large groups of students at certain schools not taking the math and reading CRTs used for the pre-tests because their school piloted the SBAC tests in 2014. Although the What Works Clearinghouse (WWC) guidelines for attrition apply only to randomized controlled trials, the high differential attrition rate for the grade 6 analyses suggest that the findings could have high levels of bias. The attrition rates for grades 7 and 8 exceeded 23% but the rates for these two grades are in the acceptable range based on WWC guidelines (U.S. Department of Education, 2014).

Exhibit 10. Attrition Rates for the PASS Assessment and CRT Analyses

	Treatment Schools (n = 4)			Comparison Schools (n = 8; n = 16)		
	Attrition Rate	Students Lost to Attrition	Students Enrolled on Count Day	Attrition Rate	Students Lost to Attrition	Students Enrolled on Count Day
PASS Assessment						
Grade 6	40.9%	709	1,734	29.6%	916	3,095
Grade 7	42.4%	747	1,763	35.2%	1107	3,145
Science CRT						
Grade 8	23.9%	433	1,810	24.7%	1,536	6,231

Note: There were eight comparison schools in the student survey and PASS Assessment analyses and 16 comparison schools in the CRT analyses.

For the achievement analyses, WestEd calculated the number of students who changed from the treatment group to the comparison group or from the comparison group to the treatment group (i.e., the number of crossovers) between the fall count days and the times when the outcome measures were administered. Four the 1,025 6th graders originally in the treatment group and six of the 2,179 6th graders originally in the comparison group attended a school in the opposite condition when the PASS Assessment was administered. In addition, 37 of the 1,016 7th graders originally in the treatment group and 36 of the 2,038 7th graders from the comparison group were in the opposite condition at the end of their second year of the study when the PASS Assessment was administered. A total of 107 of the 1,377 8th graders originally in the treatment condition were in a comparison or a non-participating school when the science CRT was administered at the end of their second year of participation.[1] Finally, 79 of the 4,695 8th graders originally in the comparison group attended a treatment school when the science CRT was administered.

Analyses

Baseline Balance Testing

WestEd conducted baseline balance testing on teacher and student measures to ensure that treatment and comparison groups were equivalent prior to the start of the intervention. For the baseline balance testing, we compared the mean scores of the treatment and comparison groups and calculated the effect size (i.e., the mean difference between the groups divided by the pooled standard deviation) indexing these differences. Additional details on baseline balance testing are included in Appendix C. When the treatment and comparison groups have differences that are below effect sizes of +/- 0.25 they are considered to be equivalent (U.S. Department of Education, 2014). Stronger conclusions about the impact of interventions can be drawn from quasi-experimental designs that find the baseline measures of the outcome variables are equivalent for the treatment and comparison groups (Shadish, Cook, & Campbell, 2002). For the teachers, we assessed the equivalence of each of the six survey scales from the fall for the sample of teachers included in the analyses. For the students, we compared 6th graders' science, math, and reading CRT scores from 2014 and 7th graders' science, math, and reading CRT scores from 2013. Finally, we compared the 8th graders' math and reading CRT scores from 2013.

Impact Analyses

WestEd conducted the impact analyses to examine whether PSI had an effect on the participating teachers and students. The analyses that examined the impact of PSI on the teachers utilized the six survey scales from the spring assessing the teachers' efficacy and attitudes toward STEM as the outcome variables. The outcome variables for the analyses that examined the program's impact on the students utilized the PASS Assessment scores and the science CRT scores from 2015.

We used a type of regression analysis named hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002) for the impact analyses with the student assessment outcomes in order to appropriately account for the grouping of students within schools. The HLM analyses included the students' demographic characteristics and pre-test CRT scores from the spring of 2013 or 2014 as control variables to adjust for differences between the treatment and comparison groups. Additional details on the HLM models are included in Appendix D. After conducting the HLM analyses, we calculated the effect sizes to benchmark the treatment effects.

For the impact analyses that utilized the teacher survey data, we used a different type of regression model that accounted for the nesting of teachers and students within schools. Although HLM is the

---

[1] Unlike the PASS Assessment, which was administered in only the 12 participating schools, the science CRT was administered across the district, and WestEd received data for participating students even if they were not enrolled in one of the 12 participating schools when the test was administered.

most appropriate type of analysis to use for nested educational data, the models will not always run properly (i.e., the statistical models will not converge on a unique solution) when there are a small number of teachers and/or schools included in the analyses. Exploratory HLM analyses with the teacher survey data revealed this issue with the current data and the alternative type of regression model was used instead. The impact analyses for each of the survey scales utilized the fall measure of the same scale and the demographic characteristics as covariates to control for preexisting differences between the groups.

For the students in grades 7 and 8 and the teachers in Year 2, the analyses using the assessment and survey data could be designed to assess the impact of one or two years of program participation. To assess the impact of one year of program participation, the pre-test measures from the end of Year 1 (i.e., the CRTs from 2014) or the beginning of Year 2 (i.e., the Year 2 fall survey) are used as covariates. To assess the impact of two years of program participation, the pre-test measures from prior to Year 1 (i.e., the CRTs from 2013) or the beginning of Year 1 (i.e., the Year 1 fall survey) are used as covariates. WestEd's study plan for the National Evaluation of i3 called for the analyses with the assessment data for students in grades 7 and 8 to assess the impact of two years of program participation. For the analyses with the survey data, we opted to assess the impact of one year of program participation because the attrition rates were substantially higher for the two-year impact analyses. Conducting the one-year impact analyses instead of the two-year impact analyses allowed us to include many more teachers in the survey analyses.

Subgroup Analyses

We conducted exploratory subgroup analyses with the scores from the PASS Assessment and CRTs to investigate whether the impact of PSI differed for: 1) females and males; 2) ELL and non-ELL students; 3) Hispanic/Latino students; and 4) Black/African American students. The smaller sample sizes for the teacher survey precluded us from conducting subgroup analyses with teacher outcome measures. The student subgroup analyses utilized the same HLM models as the impact analyses but included one additional variable that tested whether the program effect differed for the subgroups of interest (i.e., a variable indexing the interaction between treatment status and the demographic characteristic).

Treatment of Missing Data

WestEd removed all teachers and students from the analyses with missing outcome, pre-test, or demographic data. These teachers and students are counted in the attrition numbers reported in Exhibit 11. Our missing data strategy is consistent with the WWC (2014) recommendations and allowed the impact and baseline equivalence analyses to be based on the same samples of teachers and students.

Baseline Comparisons between the Treatment and Comparison Groups

Teacher Measures

The results of the baseline comparisons with the teacher survey data showed that the treatment and comparison teachers were equivalent at baseline on the Efficacy scale from the STEBI and the attitudes toward science, math, and engineering scales from the STEM-T (see Exhibit 11). For these four scales, the differences between the means for the treatment and comparison teachers were small and not statistically significant. The standardized differences between the means (i.e., the effect sizes) ranged from -0.01 to 0.20 and were below the equivalency cut-off of +/- 0.25 standard deviations (U.S. Department of Education, 2014). The difference between the means for the treatment and comparison teachers were larger for the Expectancies scale from the STEBI and the attitudes toward technology scale from the STEM-T and indicated that the treatment teachers reported less positive beliefs and attitudes than the comparison teachers. Although these differences were not statistically significant after a multiple comparison adjustment (Benjamini & Hochberg, 1995),[2] the standardized differences were beyond the cut-off of +/- 0.25 standard deviations. As a result, the impact analyses with these two variables should be viewed cautiously.

Exhibit 11. Baseline Balance Testing Results for the Teacher Survey Measures

Fall Survey Scales	Treatment Schools (n = 4)			Comparison Schools (n = 8)			Treatment-Comparison Difference	Effect Size	p value
	Mean	SD	n	Mean	SD	n			
STEBI Efficacy[a]	4.40	0.39	19	4.32	0.38	25	0.08	0.20	.38
STEBI Expectancies[a]	2.96	0.55	19	3.21	0.38	24	-0.24	-0.53	.07
Attitudes Toward Science[b]	6.84	0.43	18	6.84	0.35	25	0.00	0.00	.99
Attitudes Toward Math[b]	5.20	1.54	18	5.21	1.11	25	-0.01	-0.01	.98
Attitudes Toward Engineering[b]	6.04	0.67	18	5.98	0.78	25	0.06	0.08	.80
Attitudes Toward Technology[b]	6.05	1.00	18	6.42	0.69	25	-0.38	-0.45	.02

Note: [a]The Science Teaching Efficacy Belief Instrument Revised (STEBI) items were rated on a 1 (*strongly disagree*) to 5 (*strongly agree*) scale. [b]The STEM Semantics Survey for Teachers (STEM-T) items were rated on a 1 to 7 scale with varying anchors (e.g., *boring* to *interesting*).

Student Measures

The results of the baseline comparisons for the samples of students included in the PASS Assessment and the science CRT analyses indicated that the treatment and comparison students' achievement on the CRTs were equivalent prior to the students' participation in PSI (see Exhibit 12). The CRT scores shown in Exhibit 13 for students in grade 6 reflect CRTs administered in 2014 when the students were in grade 5. Additionally, the CRT scores for grades 7 and 8 were from 2013

[2] We used the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995), which reduced the critical *p* value used to judge statistical significance to below .05. A statistical adjustment was needed to reduce the probability that a significant finding would be a result of chance because data from six survey scales were analyzed.

prior to the start of PSI when the students were in grades 5 and 6, respectively. The effect sizes indexing the differences between the treatment and comparison students in grades 6–8 ranged from -0.08 to 0.06 and are well within the +/- 0.25 cut-off for satisfying baseline equivalence (U.S. Department of Education, 2014). Additionally, none of the differences were statistically significant.

Exhibit 12. Baseline Balance Testing Results for the PASS Assessment and Science CRT Samples

	Treatment Schools (n = 4)			Comparison Schools (n = 8; n = 16)					
Grade Level/Test	Mean	SD	n	Mean	SD	n	Treatment-Comparison Difference	Effect Size	p value
Grade 6									
Science CRT	293.65	63.12	1,025	296.82	59.19	2,179	-3.16	-0.05	.42
Math CRT	311.18	82.73	1,025	306.26	79.01	2,179	4.92	0.06	.56
Reading CRT	305.42	77.27	1,025	306.54	74.62	2,179	-1.12	-0.01	.87
Grade 7									
Science CRT	287.70	59.34	1,016	292.15	58.87	2,038	-4.45	-0.08	.44
Math CRT	324.63	85.97	1,016	323.53	84.56	2,038	1.10	0.01	.86
Reading CRT	304.81	71.44	1,016	307.98	71.52	2,038	-3.17	-0.04	.45
Grade 8									
Math CRT	270.31	73.78	1,377	276.54	74.42	4,695	-6.22	-0.08	.46
Reading CRT	299.38	78.49	1,377	299.38	80.95	4,695	0.00	0.00	.99

Note: CRT = Criterion Referenced Test. The analyses for grades 6 and 7 included students in 8 comparison schools. The analyses for grade 8 included students in 16 comparison schools. The CRTs listed for the students in grades 6 were administered in 2014 when the students were in grade 5. The CRTs listed for students in grades 7 and 8 were administered in 2013 when the students were in grades 5 and 6, respectively.

Findings

The evaluation findings from the second year of PSI implementation are outlined in this section of the report. We first summarize the program implementation findings for the eight components that we assessed for Year 2. The program implementation findings from Year 1 are included in Appendix E. Finally, this section contains the results from the impact analyses that examined the effect of PSI on the teachers and students.

Program Implementation

Component 1: Teacher Professional Development

Three of four schools were rated at moderate implementation. As such, the sample-level implementation for Component 1: Teacher Professional Development was classified as emerging/moderate. The level of implementation for each treatment school based on the scoring rubric for Component 1: Teacher Professional Development is shown in Exhibit 13. For the full component, Garside, Gibson, and Johnston exhibited moderate implementation, and Findlay exhibited emerging implementation. For Component 1-1, six of the nine teachers at Findlay did not complete any of the training modules. Similarly, seven of the nine teachers at Findlay did not participate in the Core Training for Component 1-2. It should be noted, however, that three of the nine teachers at Findlay could not attend the trainings for Components 1-1 and 1-2 because their contracts started after the trainings took place, and one transferred into the school and was unaware of training requirements. All teachers in the four schools were registered for the Virtual Academy, which resulted in the high implementation scores for Component 1-3. All four schools had low implementation scores for Component 1-4 in part because two of the optional PD sessions were canceled due to low participation.

Exhibit 13. Implementation Levels for Component 1: Teacher Professional Development

School	Full Component	Subcomponent 1-1: Participation in PLTW Readiness Training Modules	Subcomponent 1-2: Participation in PLTW Core Training	Subcomponent 1-3: Participation in PLTW Ongoing Training - Virtual Academy	Subcomponent 1-4: Attendance at Optional CCSD PD Sessions
Findlay	Emerging implementation	Less than 34% of teachers completed 4–5 Readiness Modules	Less than 34% of teachers participated in Core Training for more than 4 days a week	51% or more teachers registered for Virtual Academy	Less than 34% of teachers attended two optional CCSD PD sessions
Garside	Moderate implementation	Greater than 66% of teachers completed 4–5 Readiness Modules	Greater than 66% of teachers participated in Core Training for more than 4 days a week	51% or more teachers registered for Virtual Academy	Less than 34% of teachers attended two optional CCSD PD sessions
Gibson	Moderate implementation	Greater than 66% of teachers completed 4–5 Readiness Modules	Greater than 66% of teachers participated in Core Training for more than 4 days a week	51% or more teachers registered for Virtual Academy	Less than 34% of teachers attended two optional CCSD PD sessions
Johnston	Moderate implementation	Greater than 66% of teachers completed 4–5 Readiness Modules	Greater than 66% of teachers participated in Core Training for more than 4 days a week	51% or more teachers registered for Virtual Academy	Less than 34% of teachers attended two optional CCSD PD sessions

Component 2: Implementation of Clark County School District (CCSD) and Project Lead the Way (PLTW) Curricula

The sample-level implementation for Component 2: Implementation of CCSD and PLTW Curricula was classified as emerging/moderate because only two of the four schools had high levels of implementation. The level of implementation for each treatment school based on the scoring rubric for Component 2 is displayed in Exhibit 14. For the full component, Garside and Johnston exhibited high implementation; Findlay and Gibson exhibited moderate implementation. For Component 2-1, there was considerable variability within each of the schools. For example, even though Findlay had the lowest levels of implementation for this subcomponent, three of the nine teachers at the school had GTT objectives included in greater than 25% of their lessons. Across the four schools, the average number of teachers per quarter with GTT objectives included in greater than 25% of lesson plans ranged from four to six. For Component 2-2, for each of the four schools more than 66% of teachers had GTT activities observed in greater the 25% of lessons. The percentage of lessons observed that included GTT activities ranged from 33% to 90% at Findlay, 60% to 80% at Garside, 0% to 90% at Gibson, and 50% to 100% at Johnston. It should be noted, however, that a greater number of classroom observations were scheduled during lessons when GTT activities were planned, based on CCSD's schedule for science lessons, in order for observations to inform CCSD's monitoring of implementation.

Exhibit 14. Implementation Levels for Component 2: Implementation of CCSD and PLTW Curricula

School	Full Component	Subcomponent 2-1: Lesson Plans Aligned to GTT/CCSD Benchmark Calendar	Subcomponent 2-2: Classroom Observations of Integration of GTT Activities
Findlay	Moderate implementation	Less than 34% of teachers had GTT objectives included in greater than 25% of lesson plans	Greater than 66% of teachers had GTT activities observed in greater than 25% of lessons
Garside	High implementation	Greater than 66% of teachers had GTT objectives included in greater than 25% of lesson plans	Greater than 66% of teachers had GTT activities observed in greater than 25% of lessons
Gibson	Moderate implementation	34–66% of teachers had GTT objectives included in greater than 25% of lesson plans	Greater than 66% of teachers had GTT activities observed in greater than 25% of lessons
Johnston	High implementation	Greater than 66% of teachers had GTT objectives included in greater than 25% of lesson plans	Greater than 66% of teachers had GTT activities observed in greater than 25% of lessons

Component 3: Ongoing Teacher Support

The sample-level implementation for Component 3: Ongoing Teacher Support was classified as emerging/moderate. This component reflects the level of support provided to schools by the PSI project coordinator. Only one school, Findlay, was rated high implementation. As shown in Exhibit 15, three schools had moderate levels of implementation whereby the percentage of weekly visits to the schools ranged from 58% to 63%. The PSI coordinator conducted school visits (visits that included providing teachers with support) during 58% to 63% of the 36 weeks of school for the schools with a moderate implementation rating, and during 72% of the school weeks at Findlay.

Exhibit 15. Implementation Levels for Component 3: Ongoing Teacher Support

School	Full Component	Subcomponent 3-1: Weekly On-Site Support from PSI Project Coordinator
Findlay	High implementation	Greater than 70% of weekly visits occurred
Garside	Moderate implementation	50–70% of weekly visits occurred
Gibson	Moderate implementation	50–70% of weekly visits occurred
Johnston	Moderate implementation	50–70% of weekly visits occurred

Component 4: Classroom Technology

For Component 4: Classroom Technology, the sample-level implementation was classified as high. As shown in Exhibit 16, all four schools had high levels of implementation. The percentage of teacher and student computers that were functioning ranged from 96% to 100% across schools.

Exhibit 16. Implementation Levels for Component 4: Classroom Technology

School	Full Component	Subcomponent 4-1: Laptops for Students and Teachers
Findlay	High implementation	Greater than 75% of the computers were functioning
Garside	High implementation	Greater than 75% of the computers were functioning
Gibson	High implementation	Greater than 75% of the computers were functioning
Johnston	High implementation	Greater than 75% of the computers were functioning

Component 5: Weekly Sessions with STEM Professionals

The sample-level implementation for Component 5: Weekly Sessions with STEM Professionals was classified as emerging/moderate. As shown in Exhibit 17, for the full component Garside had moderate implementation and Findlay, Gibson, and Johnston had emerging implementation. With respect to subcomponent 5-1, at least one teacher participated in 47% of weekly sessions at Johnston, 65% of weekly sessions at Gibson, and 88% of weekly sessions at Garside. At Findlay, at least one teacher participated in only 24% of weekly sessions. Due to challenges with securing STEM professionals' involvement, compounded by the transition to a new PSI project coordinator early in the school year, STEM professionals were not involved with weekly sessions at any schools during Year 2, leading to low ratings for all schools on 5-2. Finally, the average number of students over the school year attending each of the weekly sessions was 3.9 at Johnston, 4.4 at Gibson, 6.1 at Findlay, and 16.1 at Garside. All four schools had moderate average attendance in the first quarter (9 to 19 students), with decreasing attendance over the school year, with the exception of Garside which maintained average quarterly attendance above 13 students.

Exhibit 17. Implementation Levels for Component 5: Weekly Sessions with STEM Professionals

School	Full Component	Subcomponent 5-1: Teacher Participation in Weekly Sessions	Subcomponent 5-2: STEM Professional Participation in Weekly Sessions	Subcomponent 5-3: Student Participation in Weekly Sessions
Findlay	Emerging implementation	One teacher participated in less than 26% of weekly sessions	Less than 60% of weekly sessions were supported by a STEM Professional	An average of fewer than 11 students per weekly session
Garside	Moderate implementation	One teacher participated in greater than 85% of weekly sessions	Less than 60% of weekly sessions were supported by a STEM Professional	An average of 11–20 students per weekly session
Gibson	Emerging implementation	One teacher participated in 26–85% of weekly sessions	Less than 60% of weekly sessions were supported by a STEM Professional	An average of fewer than 11 students per weekly session
Johnston	Emerging implementation	One teacher participated in 26–85% of weekly sessions	Less than 60% of weekly sessions were supported by a STEM Professional	An average of fewer than 11 students per weekly session

Component 6: STEM Club

The sample-level implementation for Component 6: STEM Club was classified as emerging/moderate. For the full component, Gibson and Johnston had moderate implementation; Findlay and Garside had emerging implementation (see Exhibit 18). For subcomponent 6-1, at least one teacher participated in 79% and 76% of STEM Club activities at Gibson and Johnston, respectively. At Findlay and Garside, at least one teacher participated in 64% and 47% of STEM Club activities, respectively. There was substantial variability in STEM Club participation by students over the course of the year across the four schools. Gibson had an average of 14 students over the entire year, beginning strong with an average of 24 students in quarter one sessions and decreasing to seven students in quarter four sessions. Findlay had an average across the year of eight students, beginning the year with zero students in quarter one and ending the year with an average of 17 in quarter four. Johnston had an overall average of six students, with an average of 16 students in quarter one and four students in quarter four. Garside remained relatively stable over the year (an overall average of three students), with five students in quarter one and three students in quarter four.

Exhibit 18. Implementation Levels for Component 6: STEM Club

School	Full Component	Subcomponent 6-1: Teacher Participation in STEM Club	Subcomponent 6-2: Student Participation in STEM Club
Findlay	Emerging implementation	One teacher participated in 60%–75% of STEM Club activities	An average of fewer than 11 students per STEM Club activity
Garside	Emerging implementation	One teacher participated in less than 60% of STEM Club activities	An average of fewer than 11 students per STEM Club activity
Gibson	Moderate implementation	One teacher participated in greater than 75% of STEM Club activities	An average of 11–20 students per STEM Club activity
Johnston	Moderate implementation	One teacher participated in greater than 75% of STEM Club activities	An average of fewer than 11 students per STEM Club activity

Component 7: STEM Summer Camp

The sample-level implementation for Component 7: STEM Summer Camp was high. Three schools had high levels of implementation on the full component (see Exhibit 19). The nine activities that were planned for the summer camp were all implemented. These activities included: LEGO robotics, electronic circuit board design/soldering, VEX robotics, 3D engineering and printing, bottle biology, flight and space, engineering design, computer programming, and field trip. In addition, three to six teachers from three schools (Garside, Gibson, and Johnston) participated in the summer camp. Student participation in the summer camp was also relatively high, with more than 30 students participating from three of the four schools. At the fourth school, an average of 29 students participated in the summer camp. Across the four schools, the average number of students attending the summer camp ranged from 29 to 45 students.[3]

Exhibit 19. Implementation Levels for Component 7: STEM Summer Camp

School	Full Component	Subcomponent 7-1: Implementation of Planned Activities	Subcomponent 7-2: Teacher Participation in Summer Camp	Subcomponent 7-3: Student Participation in Summer Camp
Findlay	Moderate implementation	Greater than 80% of planned activities were implemented	2 or more teachers per school participated in the summer camp	Less than 30 students per school attended the summer camp
Garside	High implementation	Greater than 80% of planned activities were implemented	3 or more teachers per school participated in the summer camp	More than 40 students per school attended the summer camp
Gibson	High implementation	Greater than 80% of planned activities were implemented	3 or more teachers per school participated in the summer camp	More than 40 students per school attended the summer camp
Johnston	High implementation	Greater than 80% of planned activities were implemented	3 or more teachers per school participated in the summer camp	More than 40 students per school attended the summer camp

3 Findlay, 29 students; Garside, 38 students; Gibson, 45 students; Johnston, 41 students.

Component 8: Science and Math Tutoring

The sample-level implementation for Component 8: Science and Math Tutoring was emerging/moderate. As presented in Exhibit 20, for the full component, only Gibson was rated at high implementation and Johnston was rated at moderate implementation. Garside and Findlay were rated at emerging implementation; however, no tutoring activities took place at these schools during Year 2. At Johnston, only math tutoring occurred. At Gibson, both math and science tutoring occurred. Where tutoring occurred, teacher participation was high, with at least one teacher participating in more than 85% of tutoring sessions. Student participation in tutoring, in both math and science, was moderate (i.e., between 6 to 10 students) and fairly consistent over the year. Gibson and Johnston averaged eight and seven students in science tutoring sessions over the year, respectively; Gibson averaged six students over the year for math tutoring.

Exhibit 20. Implementation Levels for Component 8: Science and Math Tutoring

School	Full Component	Subcomponent 8-1: Teacher Participation in Science Tutoring	Subcomponent 8-2: Teacher Participation in Math Tutoring	Subcomponent 8-3: Student Participation in Science Tutoring	Subcomponent 8-4: Student Participation in Math Tutoring
Findlay	Emerging implementation	One teacher participated in less than 26% of tutoring sessions	One teacher participated in less than 26% of tutoring sessions	An average of fewer than 6 students per tutoring session	An average of fewer than 6 students per tutoring session
Garside	Emerging implementation	One teacher participated in less than 26% of tutoring sessions	One teacher participated in less than 26% of tutoring sessions	An average of fewer than 6 students per tutoring session	An average of fewer than 6 students per tutoring session
Gibson	High implementation	One teacher participated in greater than 85% of tutoring sessions	One teacher participated in greater than 85% of tutoring sessions	An average of 6–10 students per tutoring session	An average of 6–10 students per tutoring session
Johnston	Moderate implementation	One teacher participated in greater than 85% of tutoring sessions	One teacher participated in 26–85% of tutoring sessions	An average of 6–10 students per tutoring session	An average of fewer than 6 students per tutoring session

Program Impacts on Teacher Outcomes

The results of the analyses that examined the impact of the second year of PSI on the outcomes assessed by the teacher survey are presented in Exhibit 21. There was no clear pattern favoring the treatment or comparison group. The means for three of the scales were higher for the treatment group and the means for the other three scales were higher for the comparison group. Additionally, none of the differences were statistically significant after using the Benjamini-Hochberg correction for multiple comparisons.[4] The effect sizes for the two STEBI scales were close to zero, indicating that the teachers in the treatment and comparison groups reported very similar self-efficacy for teaching science and beliefs about the efficacy of science instruction. The difference between the groups on the teachers' attitudes toward science was small and both groups, on average, reported very positive attitudes toward science. The effect sizes for the attitudes toward math, engineering, and technology ranged from -0.37 to 0.41 and are considered at least medium sized impacts in the context of other educational interventions (Lipsey et al., 2012). However, with the attitudes toward math scale favoring the treatment group and the attitudes toward technology scale favoring the comparison group, conclusions cannot be drawn about trends favoring either group.

Exhibit 21. Program Impacts for the Teacher Survey Measures

	Treatment Schools (n = 4)			Comparison Schools (n = 8)			Treatment-Comparison		
Spring Survey Scales	Mean	SD	n	Mean	SD	n	Difference	Effect Size	p value
STEBI Self-Efficacy[a]	4.38	0.37	19	4.35	0.40	25	0.02	0.06	.79
STEBI Expectancies[a]	3.16	0.53	19	3.14	0.62	24	0.02	0.03	.90
Attitudes Toward Science[b]	6.87	0.36	18	6.93	0.23	25	-0.06	-0.22	.29
Attitudes Toward Math[b]	5.46	1.63	18	4.86	1.33	25	0.60	0.41	.03
Attitudes Toward Engineering[b]	6.00	1.12	18	6.02	1.07	25	-0.03	-0.02	.94
Attitudes Toward Technology[b]	6.05	1.28	18	6.41	0.65	25	-0.36	-0.37	.11

Note: [a]The Science Teaching Efficacy Belief Instrument Revised (STEBI) items were rated on a 1 (*strongly disagree*) to 5 (*strongly agree*) scale. [b]The STEM Semantics Survey for Teachers (STEM-T) items were rated on a 1 to 7 scale with varying anchors (e.g., *boring* to *interesting*). None of the differences were statistically significant after using the Benjamini-Hochberg correction for multiple comparisons.

Program Impacts on Student Outcomes

Exhibit 22 displays the results of the impact analyses with the PASS Assessment for students in grades 6 and 7 and the science CRT for students in grade 8. Analyses on grade 6 PASS were conducted on the grade 6 sample of students with one year of participation in PSI. Analyses on grade 7 PASS and grade 8 science CRT were conducted on the grade 7 and 8 sample of students who had participated in PSI for two years.

The results with the students in grade 6 indicated that one year of participation in PSI did not impact performance on the PASS Assessment. The impact estimate for grade 6 students who participated in the program during Year 2 showed that the treatment students, on average, correctly completed 0.18 items more than the comparison students. However, the difference was not statistically significant and the effect size (ES = 0.03) is considered very small (Lipsey et al., 2012). In addition, the results for grades 7 and 8 indicated that two years of participation in PSI did not have an impact on science achievement. The impact estimate for grade 7 indicated that the comparison

[4] We used the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995), which reduced the critical p value used to judge statistical significance to below .05. A statistical adjustment was needed to reduce the probability that a significant finding would be a result of chance because data from six survey scales were analyzed.

students, on average, correctly completed 0.29 items more on the PASS Assessment than the treatment students after two years of program participation. In grade 8, the science CRT scale score for the treatment students was 0.71 points higher than the scale score for the comparison students after two years of program participation. Consistent with the results from grade 6, the results from grades 7 and 8 were not statistically significant and the effect sizes were very small.

Exhibit 22. Program Impacts on the PASS Assessment for Grades 6 and 7 and the Science CRT for Grade 8

	Treatment Schools (n = 4)			Comparison Schools (n = 8; n = 16)			Treatment-Comparison		
Grade Level/Test	Mean	SD	n	Mean	SD	n	Difference	Effect Size	p value
Grade 6 PASS	16.09	5.61	1,025	15.92	6.02	2,179	0.18	0.03	.84
Grade 7 PASS	13.40	5.14	1,016	13.69	5.26	2,038	-0.29	-0.05	.37
Grade 8 CRT	299.86	70.30	1,377	299.16	75.77	4,695	0.71	0.01	.89

Note: The PASS scores are the number of items answered correctly on the 30-item assessment. CRT = Criterion Referenced Test. The impact estimate for grade 6 is based on one year of program participation and the impact estimates for grades 7 and 8 are based on two years of program participation.

The distributions of the treatment and comparison students' scores on the PASS Assessment are shown in Exhibits 23 and 24 for grades 6 and 7, respectively. Additionally, the distributions of the treatment and comparison students' scores on the grade 8 science CRT are shown in Exhibit 25. Although there are many more comparison students in each exhibit, the distributions of scores are very similar for the treatment and comparison students and agree with the impact estimates in Exhibit 22 showing no differences between the groups.

Exhibit 23. Scores on the PASS Assessment for Treatment and Comparison Students in Grade 6

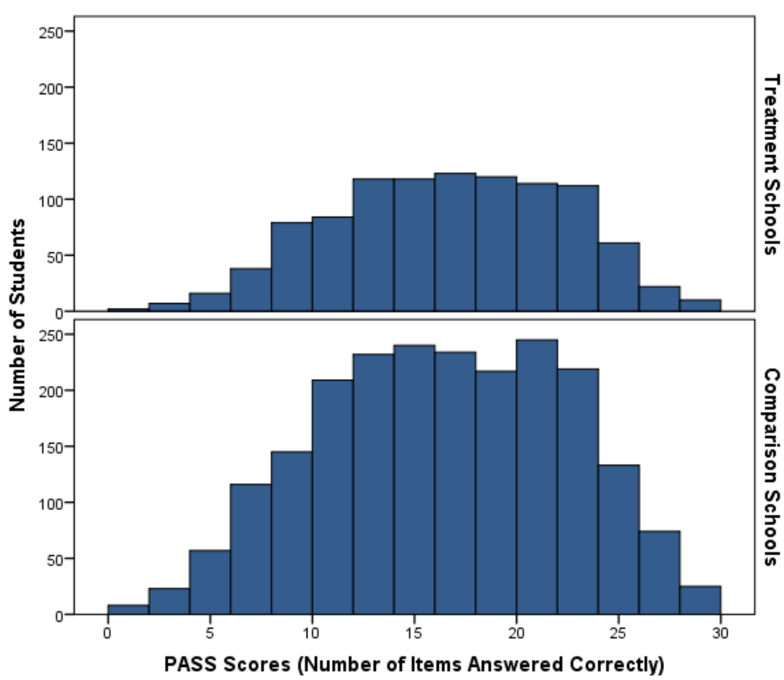


Exhibit 24. Scores on the PASS Assessment for Treatment and Comparison Students in Grade 7

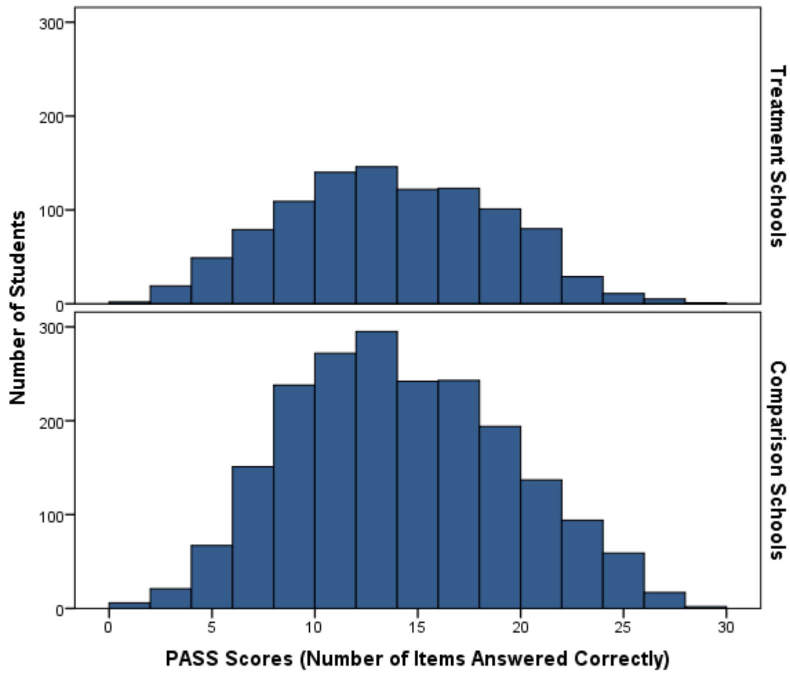
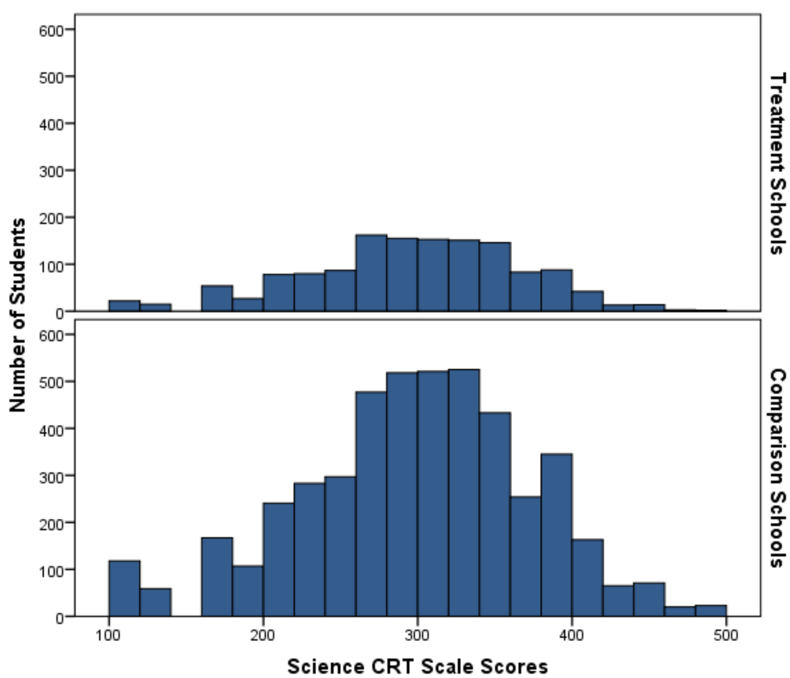


Exhibit 25. Scale Scores on the Science CRT for Treatment and Comparison Students in Grade 8



We conducted an exploratory set of subgroup analyses with the students in grades 6–8 to examine whether the impact of PSI on the PASS Assessment and science CRT scores differed for females and males, ELL and non-ELL students, Hispanic/Latino and non-Hispanic/Latino students, and

Black/African American and non-Black/African American students. As shown in Exhibit 26, with the exception of Hispanic/Latino students in grade 6, none of the subgroup analyses were statistically significant, indicating that the impact of the program was not different for these subgroups. The significant subgroup effect for Hispanic/Latino students in grade 6 showed that treatment Hispanic/Latino students (M = 16.89; SD = 5.40; n = 587) had higher PASS Assessment scores than comparison Hispanic/Latino students (M = 16.24; SD = 5.97; n = 1,399). Conversely, treatment non-Hispanic/Latino students (M = 16.25; SD = 5.89; n = 438) had lower PASS Assessment scores than comparison non-Hispanic/Latino students (M = 16.93; SD = 5.98; n = 780). Although the subgroup effect was statistically significant, the effect size for Hispanic/Latino students (ES = 0.11) and the effect size for non-Hispanic/Latino students (ES = -0.11) were both very small. The positive program effect for Hispanic/Latino students in grade 6 should be viewed with caution because of the small effect size and the fact that the pattern was not replicated for Hispanic/Latino students in grades 7 and 8.

Exhibit 26. Statistical Significance (*p* values) of the Subgroup Analyses with the PASS Assessment and Science CRT

Subgroup	Grade 6	Grade 7	Grade 8
Gender	.06	.18	.88
ELL status	.08	.71	.08
Black/African American	.77	.73	.37
Hispanic/Latino	< .001	.80	.48

Note: ELL = English language learner. The analyses for grades 6 and 7 used the PASS Assessment, and the analyses for grade 8 used the science CRT.

Conclusions

The findings from the implementation study did not reveal consistently high levels of implementation of PSI in the four participating middle schools. The schools had high levels of implementation on components dealing with the availability of classroom technology and participation in the STEM summer camp. However, none of the schools had high levels of implementation for teacher professional development, which is an extremely critical aspect of the intervention. Additionally, the schools generally had low levels of implementation in school-year-based extracurricular activities, including the weekly sessions with STEM professionals, STEM club, and science and math tutoring. Implementation measures, like the ones collected for the current study, can help evaluators determine whether an intervention is implemented with sufficient fidelity to produce an impact on the outcomes of interest (O'Donnell, 2008).

Overall, the results from the student impact analyses revealed participation in PSI did not improve students' science achievement. Specifically, one year of participation in PSI did not impact 6th graders' science achievement and two years of participation in PSI did not impact 7th and 8th graders' science achievement. Additionally, the teacher impact analyses did not show that PSI had an effect on the science teachers' beliefs about science and attitudes toward STEM.

The current findings showing small differences between treatment and comparison students are consistent with trends in the field of education research. First, rigorous studies are not likely to identify positive program effects. For example, the Coalition for Evidence-Based Policy (2013) found that only 12% of the randomized controlled trials commissioned by the Institute of Education Sciences produced positive effects. Second, the lack of positive impacts with low levels of implementation is consistent with a review of intervention studies that found better outcomes were associated with greater fidelity of implementation (O'Donnell, 2008).

WestEd used a rigorous quasi-experimental design that identified comparison schools using scaled Euclidean-distance matching (Sekhon, 2011) to compare the science achievement of PSI participants with the achievement of their peers in non-participating schools. The treatment and comparison students were equivalent at baseline prior to the intervention and the evaluation utilized reliable measures to assess the students' achievement. As a result, it is likely that the student impact findings will receive a rating of *Meets WWC Group Design Standards with Reservations* from the WWC.

We do not know the generalizability of the current findings. The impact analyses relied on a small sample of middle schools in one school district. Additionally, the level of implementation in the four PSI schools was not high and the district reported that many of the comparison schools were implementing similar STEM activities during and after school. It may be that PSI could have a positive impact on students and teachers in a different setting and/or when the intervention is implemented with higher fidelity to the program model. In light of these possibilities, further research on PSI is warranted.

References

Abt Associates. (2013). *Studying implementation in the context of i3: W T Grant Foundation I3 Learning Community*. Retrieved from http://forumfyi.org/files/Abt%20i3%20Learning%20Community%20October%202013.pdf

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology, 44*(5), 427–445.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B Methodological, 57*(1), 289–300.

Coalition for Evidence-Based Policy. (2013). *Randomized controlled trials Commissioned by the Institute of Education Sciences since 2002: How many found positive versus weak or no effects*. Washington, DC: Coalition for Evidence-Based Policy.

Davis, D. R., Ellett, C. D., & Rugutt, J. K. (1999, November). *School level differences in teachers' perceptions of multiple dimensions of school culture*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Point Clear, AL.

Fraser, B. J. (1981). *Test of Science-Related Attitudes: Handbook*. Camberwell, Australia: Allanby Press.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M.D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. (NCSER 2013-3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 1*, 85–91.

McCoach, D. B., & Siegle, D. (2003). The school attitude assessment survey-revised: A new instrument to identify academically able students who underachieve. *Educational and Psychological Measurement, 63*(3), 414–429.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.

Nevada Department of Education. (2012). *Technical Report for the 2010–11 Nevada CRT and HSPE Assessments*. Dover, NH: Measured Progress.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research, 78*(1), 33-84.

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

Riggs, I. M., & Enochs, L. G. (1989, March). *Toward the development of an elementary teacher's science teaching efficacy belief instrument*. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching, San Francisco, CA.

Rogers, W. H. (1993). Regression standard errors in clustered samples. *Stata Technical Bulletin, 13*, 19–23.

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

Sekhon, J. S. (2011). Multivariate and propensity score matching software with automated balance optimization: The Matching Package for R. *Journal of Statistical Software, 42*, 1–52.

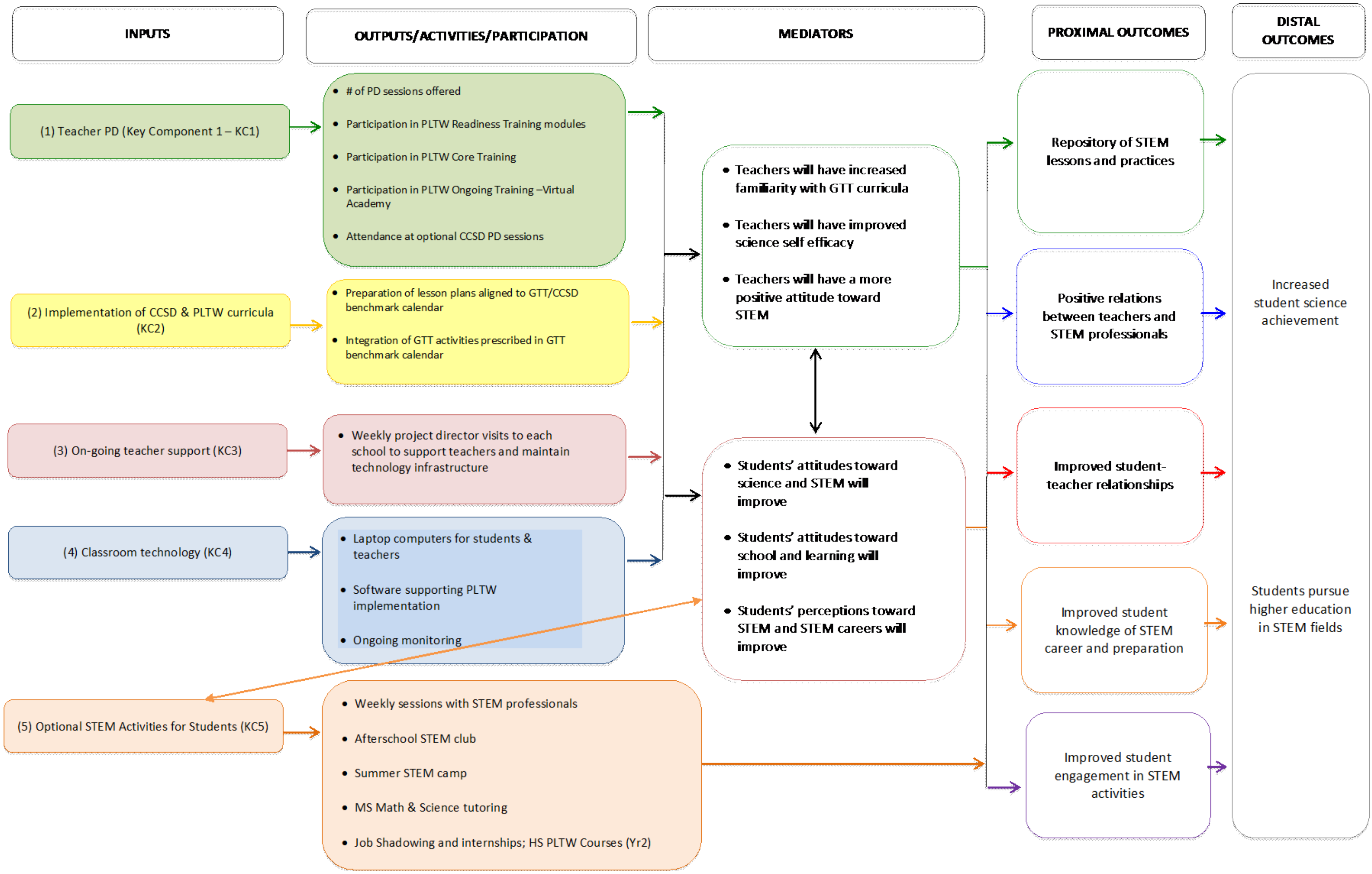
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

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

Tyler-Wood, T., Knezek, G., & Christensen. R. (2010). Instruments for assessing interest in STEM content and careers. *Journal of Technology and Teacher Education, 18*(2), 341–363.

U.S. Department of Education (2014). *What Works Clearinghouse: Procedures and standards handbook version 3.0*. Retrieved from http://ies.ed.gov/ncee/wwc/

Appendix A: Pathways to STEM Initiative (PSI) Logic Model



Appendix B: Comparison School Selection

WestEd implemented scaled Euclidean-distance matching using R's Matching package (Sekhon, 2011) to select the comparison schools. Given the importance of including all variables known to be associated with science achievement and participation in PSI (Rubin & Thomas, 1996), we utilized the following school-level measures in the matching procedure: 1) the mean scale score for the grade 8 science CRT; 2) the mean scale score for the grade 8 reading CRT; 3) the mean scale score for the grade 8 math CRT; 4) the student transiency rate; 5) the percentage of students qualifying for FRL; and 6) the percentage of White, Black/African American, and Hispanic/Latino students. Although it would have been optimal to use school-level measures from 2012–13 (i.e., the year directly before the start of PSI), we needed to notify the schools that were selected as comparison sites before the 2012–13 data were ready. As such, we used the aforementioned school-level measures from 2011–12.

The four PSI schools used the SpringBoard curriculum in 2013–14. SpringBoard is a reading and math curriculum that the district began using in 2012–13 in 31 middle schools. Because SpringBoard has the potential to impact the students' science achievement, we selected comparison schools from only the pool of non-PSI middle schools in the district also using SpringBoard. In addition, we removed one school that was using the GTT curriculum, which is a component of PSI, from the pool of potential comparison schools. This final pool of potential comparison sites included 26 schools. WestEd's strategy is consistent with Stuart's (2010) notion of combining exact matching with an additional matching strategy, such as scaled Euclidean-distance matching.

WestEd used a one-to-many matching strategy in order to improve the statistical power of the analyses (Shadish et al., 2002). Due to the amount of resources needed to administer a survey and a science assessment to large numbers of students, we selected two non-PSI schools for each PSI school and administered student surveys and the PASS Assessment (for grades 6 and 7) in these 12 schools. For the analyses with the grade 8 science CRT, we selected two additional comparison schools for each PSI school so that the analyses included a total of four PSI schools and 16 non-PSI schools.

WestEd used the Weight.matrix argument in the Matching package to weight the variables used in the matching process. The science CRT was weighted twice as heavily as the reading CRT, math CRT, and the FRL percentage. Additionally, the science CRT was weighted 10 times as much as the percentage of White, Black/African American, and Hispanic/Latino students and 10 times as much as the student transiency rate. WestEd used nearest neighbor matching without replacement (Stuart, 2010). That is, the non-treatment schools with the closest scaled Euclidean-distance scores to the PSI schools were selected as matches. The two schools with the closest scaled Euclidean-distance scores to each PSI school were the matches that were used for the survey and science assessment. By matching without replacement, each school was removed from the pool of potential comparison schools, ensuring that a comparison school did not match with multiple PSI schools. Since the order in which the matches are identified can impact which schools are selected as matches, the schools were sorted based on a random number generator prior to the matching.

Following the matching, we calculated the effect size or the standardized difference in the means (i.e., the mean difference between the treatment and comparison groups divided by the student-level standard deviation) for each of the achievement measures used in the matching process. This type of numerical balance diagnostic helped us determine the quality of the matches. We used the estimate of the student-level standard deviations based on the school-level standard deviations from the 30 schools included in the matching process using the following formula presented in the What Works Clearinghouse handbook (U.S. Department of Education, 2014):

SDcluster = SDstudent * sqrt (intra-class correlation)

We used an estimate of 0.20 for the intra-class correlation (U.S. Department of Education, 2014). The results showed that the mean differences between the treatment and comparison schools were all less than 0.15 standard deviations. We additionally calculated the mean differences between the treatment and comparison groups based on the student transiency rate; the percentage of students qualifying for FRL; and the percentage of White, Black/African American, and Hispanic/Latino students. There was less than a nine percentage-point difference between the treatment and comparison schools for each of these five non-achievement measures.

Appendix C: Baseline Balance Testing

WestEd conducted analyses to assess the baseline equivalence of the six survey scales contained in the fall 2014 teacher survey. The scales came from the STEBI (Riggs & Enochs, 1989) and the STEM-T (Tyler-Wood et al., 2010). Our original evaluation plan called for an HLM analysis with teachers nested within schools. However, the HLM models with the six scales did not consistently produce admissible solutions, which is consistent with the research showing that HLM analyses are problematic when the number of groups (i.e., schools) is below 30 (Maas & Hox, 2005). As a result, the baseline equivalence with the teacher survey data was tested using single-level regression models. In order to appropriately account for the multi-level structure of the data, we used Stata 13.1's cluster option in the standard regression package to calculate accurate *p* values (Rogers, 1993). Consistent with the HLM models used for the student analyses, the regression models included the treatment status (treatment teacher = 1; comparison teacher = 0) and three dummy-coded variables representing the four matching blocks as predictors. The analyses included all teachers who had complete baseline, outcome, and demographic data.

WestEd also conducted analyses to examine the equivalence of the treatment and comparison students in grade 6 in Year 2 on the science, math, and reading CRTs from 2014; the treatment and comparison students in grade 7 in Year 2 on the science, math, and reading CRTs from 2013; and the treatment and comparison students in grade 8 in Year 2 on the math and reading CRTs from 2013. The baseline equivalence was tested separately by grade level. The baseline balance testing included only students with complete baseline, outcome, and demographic data.

The baseline equivalence was tested using two-level HLM models (i.e., students nested within schools) so that the models for the baseline balance testing had the same structural components as the impact models. The HLM models did not include any level-1 predictors but included the treatment status and the three dummy-coded variables representing the four matching blocks as level-2 predictors. We utilized the standardized mean difference to calculate effect sizes using the baseline measures. The effect sizes were calculated by dividing the differences between the mean baseline scores for the PSI group and the mean baseline scores for the comparison group by the pooled standard deviation of the baseline measures. The treatment group means were adjusted for the covariates in the HLM models (i.e., the three dummy-coded variables representing the matching blocks). In addition, to judge whether the baseline differences were statistically significant, we utilized the Benjamini-Hochberg correction for contrasts with statistically significant baseline measures. The model for assessing baseline equivalence is outlined by the equations below using Raudenbush and Bryk's (2002) terminology and notations.

Level-1 model:

$$Y_{ij} = \beta_{0j} + r_{ij}$$

Where

Y_{ij} is the average baseline science, math, or reading CRT score for student i in school j;

β_{0j} is the average baseline science, math, or reading CRT score for students within school j;

r_{ij} is the residual (i.e., a level-1 random effect) associated with student i's baseline score in school j using the level-1 model.

Level-2 model:

$\beta_{0j} = \gamma_{00} + \gamma_{01}$(Intervention Status)$j$ + γ_{02}(Blocking Variable 1)j + γ_{03}(Blocking Variable 2)j + γ_{04}(Blocking Variable 3)j + u_{0j}

Where

γ_{00} is the average baseline science, math, or reading CRT score for the comparison group after accounting for the covariates;

Intervention status (PSI = 1; non-PSI = 0) was a dummy-coded variable that contrasted the schools participating in the intervention with the comparison schools. γ_{01} is the level-2 coefficient describing the strength and direction of the association between the intervention status and the baseline scores. Values greater than zero for γ_{01} indicated the PSI students had higher baseline scores while values below zero indicated the comparison students had higher baseline scores. γ_{02}, γ_{03}, γ_{04} are additional level-2 coefficients that represented the four matching blocks. Finally, u_{0j} is the school random effect corresponding to the deviation of school j's level-1 intercept, β_{0j}, from its predicted value using the school-level model.

Appendix D: Impact Analyses

To conduct the analyses that examined the impact of PSI on the students' science achievement in 2015, WestEd employed two-level HLM models (Raudenbush & Bryk, 2002) that accounted for the clustering of students (i.e., level-1) within schools (i.e., level-2). Our original study plan additionally included the evaluation of the impact of PSI on students' math and reading achievement in 2015. However, due to technical problems with the SBAC testing in 2015, students in CCSD did not complete the state math and reading assessments that we proposed to use for the outcome measures in 2015. As a result, we could not examine the impact of PSI on reading and math achievement in 2015. The analysis with students in grade 6 assessed the impact of PSI after the students participated in the program for one year. The analyses with the students in grades 7 and 8 assessed the impact of two years of PSI participation. The students' scores on the science (for grades 6 and 7 only), math, and reading CRTs were used as covariates. The student-level demographic variables and the school-level variables included in the models as covariates were consistent across the different grade levels.

The two-level model that we used to examine the impact of the intervention on the students' science achievement is outlined by the equations below using Raudenbush and Bryk's (2002) terminology and notations.

Level-1 model:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{Baseline Achievement Measure 1})_{ij} + \ldots + \beta_{Qj}(\text{Baseline Achievement Measure Q})_{ij} + \beta_{3j}(\text{Demographic Covariate 1})_{ij} + \ldots + \beta_{Zj}(\text{Demographic Covariate Z})_{ij} + r_{ij}$$

Where

Y_{ij} is the average science achievement score for student i in school j;

β_{0j} is the average science achievement score for students within school j;

β_{1j} to β_{Qj} are level-1 coefficients that describe the strength and direction of the associations between the baseline scores for science, math, and reading achievement and the science achievement outcome measures. For example, the analysis that examined the impact of PSI participation on 6th grade students' science achievement as measured by the 2015 PASS Assessment included the students' 5th grade science, math, and reading CRT scores from 2014 as the baseline measures. The analysis that examined the impact of PSI participation on 7th grade students' 2015 PASS Assessment scores included the students' science, math, and reading CRT scores from 2013 when the students were in 5th grade. Additionally, the analysis that examined the impact of PSI participation on 8th grade students' 2015 science CRT scores included the students' math and reading CRT scores from 2013 when the students were in 6th grade. The use of pre-test scores from 2013 for the students in grades 7 and 8 allowed us to assess the impact of two years of PSI participation.

β_{3j} to β_{Zj} are level-1 coefficients that describe the strength and direction of the associations between the student demographic characteristics and the science scores. The student-level control variables included in all analyses were race/ethnicity, ELL status, IEP status, and gender. We dummy coded the student demographic characteristics. For race/ethnicity, Hispanic/Latino was the reference group. Finally, r_{ij} is the residual (i.e., a level-1 random effect) associated with student i's science score in school j using the level-1 model.

Level-2 model:

$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Intervention Status})j + \gamma_{02}(\text{Blocking Variable 1})j + \gamma_{03}(\text{Blocking Variable 2})j + \gamma_{04}(\text{Blocking Variable 3})j + u_{0j}$

Where

γ_{00} is the average science achievement score for the comparison group after accounting for the covariates;

Intervention status (PSI = 1; non-PSI = 0) was a dummy-coded variable that contrasted the schools participating in the intervention with the comparison schools. γ_{01} was the level-2 coefficient that described the strength and direction of the association between the intervention status and the science scores. Values greater than zero for γ_{01} indicated the PSI students had higher outcome scores while values below zero indicated the non-PSI students had higher outcome scores. γ_{02}, γ_{03}, γ_{04} were additional level-2 coefficients that represented the four matching blocks. One block (i.e., one PSI school and the comparison schools that were identified as its matches) was selected as the reference group and the three dummy-coded variables contrasted the other three blocks with the reference group. Finally, u_{0j} was the school random effect corresponding to the deviation of school j's level-1 intercept, β_{0j}, from its predicted value using the school-level model.

We conducted exploratory subgroup analyses with the science achievement scores to investigate whether the impact of PSI differed for: 1) females and males; 2) ELL and non-ELL students; 3) Hispanic/Latino students; and 4) Black/African American students. The subgroup analyses utilized the same HLM models as the impact analyses but included one additional variable that tested whether the program effect differed for the subgroups of interest. Specifically, we created variables indexing the interaction between treatment status and the demographic characteristics and included these variables as additional predictors in the model. For the analyses examining Hispanic/Latino and Black/African American students, we created dummy codes contrasting these two groups with all of the other racial/ethnic groups and used these variables in place of the original dummy-coded variables for race/ethnicity. We then created the interaction terms using these two dummy codes. All of the interaction analyses were conducted separately by demographic group and grade level.

To conduct the analyses that assessed the impact of Year 2 of PSI on the constructs assessed by the teacher surveys, WestEd employed single-level regression models using Stata 13.1's cluster option in the standard regression package to calculate accurate p values (Rogers, 1993). The initial plan called for HLM but the models did not consistently produce admissible solutions because of the small sample size and the cluster option was used to produce estimates that accounted for the multi-level structure of the data. The model used to assess the impact on the teacher survey scales from spring 2015 is outlined by the equation below.

$Y = \beta_0 + \beta_1$(Intervention Status) + β_2(Blocking Variable 1) + β_3(Blocking Variable 2) + β_4(Blocking Variable 3) + β_5(Baseline Measure of Science Attitudes/Science Teaching Efficacy) + β_6(Number of Years Teaching) + ε

Where

Y is the average spring teacher science attitudes or science teaching efficacy score for teacher i in school j;

β_0 is the intercept in the model;

β_1 is the coefficient describing the strength and direction of the association between the intervention status and the spring scores. Values greater than zero for β_1 indicated the PSI teachers had higher spring scores while values below zero indicated the non-PSI teachers had higher spring scores. β_2, β_3, and β_4 represent the four matching blocks. β_5 is the coefficient that describes the strength and direction of the associations between the teacher science attitudes or science teaching efficacy scores from fall 2014 and the spring scores. β_6 is the coefficient that describes the strength and direction of the association between the number of years teaching in CCSD and the spring scores. Finally, ε is the residual or error term.

Appendix E: Implementation Findings from Year 1

Exhibit E-1. Year 1 Implementation Levels for Component 1: Teacher Professional Development

School	Full Component	Subcomponent 1-1: Participation in PLTW Readiness Training Modules	Subcomponent 1-2: Participation in PLTW Core Training	Subcomponent 1-3: Participation in PLTW Ongoing Training - Virtual Academy	Subcomponent 1-4: Attendance at Optional CCSD PD Sessions
Findlay	Emerging implementation	Less than 34% of teachers completed 4-5 Readiness Modules	Less than 34% of teachers participated in Core Training for more than 4 days a week	51% or more teachers registered for Virtual Academy	Less than 34% of teachers attended two optional CCSD PD sessions
Garside	Moderate implementation	Greater than 66% of teachers completed 4-5 Readiness Modules	Greater than 66% of teachers participated in Core Training for more than 4 days a week	51% or more teachers registered for Virtual Academy	Less than 34% of teachers attended two optional CCSD PD sessions
Gibson	High implementation	Greater than 66% of teachers completed 4-5 Readiness Modules	Greater than 66% of teachers participated in Core Training for more than 4 days a week	51% or more teachers registered for Virtual Academy	34-66% of teachers attended two optional CCSD PD sessions
Johnston	High implementation	Greater than 66% of teachers completed 4-5 Readiness Modules	Greater than 66% of teachers participated in Core Training for more than 4 days a week	51% or more teachers registered for Virtual Academy	Greater than 66% of teachers attended two optional CCSD PD sessions

Exhibit E-2. Year 1 Implementation Levels for Component 2: Implementation of CCSD and PLTW Curricula

School	Full Component	Subcomponent 2-1: Lesson Plans Aligned to GTT/CCSD Benchmark Calendar	Subcomponent 2-2: Classroom Observations of Integration of GTT Activities
Findlay	Moderate implementation	Greater than 66% of teachers had GTT objectives included in greater than 25% of lesson plans	34-66% of teachers of teachers had GTT activities observed in greater than 25% of lessons
Garside	Moderate implementation	34-66% of teachers had GTT objectives included in greater than 25% of lesson plans	Greater than 66% of teachers had GTT activities observed in greater than 25% of lessons
Gibson	Emerging implementation	Less than 34% of teachers had GTT objectives included in greater than 25% of lesson plans	Less than 34% of teachers had GTT activities observed in greater than 25% of lessons
Johnston	High implementation	Greater than 66% of teachers had GTT objectives included in greater than 25% of lesson plans	Greater than 66% of teachers had GTT activities observed in greater than 25% of lessons

Note: A greater number of classroom observations were conducted during quarters when more GTT activities were planned based on CCSD's schedule for science lessons.

Exhibit E-3. Year 1 Implementation Levels for Component 3: Ongoing Teacher Support

School	Full Component	Subcomponent 3-1: Weekly On-Site Support from PSI Project Coordinator
Findlay	Moderate implementation	50-70% of weekly visits occurred
Garside	Moderate implementation	50-70% of weekly visits occurred
Gibson	Moderate implementation	50-70% of weekly visits occurred
Johnston	Moderate implementation	50-70% of weekly visits occurred

Exhibit E-4. Year 1 Implementation Levels for Component 4: Classroom Technology

School	Full Component	Subcomponent 4-1: Laptops for Students and Teachers
Findlay	High implementation	Greater than 75% of the computers were functioning
Garside	High implementation	Greater than 75% of the computers were functioning
Gibson	High implementation	Greater than 75% of the computers were functioning
Johnston	High implementation	Greater than 75% of the computers were functioning

Exhibit E-5. Year 1 Implementation Levels for Component 5: Weekly Sessions with STEM Professionals

School	Full Component	Subcomponent 5-1: Teacher Participation in Weekly Sessions	Subcomponent 5-2: STEM Professional Participation in Weekly Sessions	Subcomponent 5-3: Student Participation in Weekly Sessions
Findlay	Emerging implementation	One teacher participated in 26-85% of weekly sessions	Less than 60% of weekly sessions were supported by a STEM Professional	An average of less than 11 students per weekly session
Garside	Moderate implementation	One teacher participated in greater than 85% of weekly sessions	Less than 60% of weekly sessions were supported by a STEM Professional	An average of 11-20 students per weekly session
Gibson	Moderate implementation	One teacher participated in greater than 85% of weekly sessions	60-75% of weekly sessions were supported by a STEM Professional	An average of less than 11 students per weekly session
Johnston	Emerging implementation	One teacher participated in 26-85% of weekly sessions	Less than 60% of weekly sessions were supported by a STEM Professional	An average of 11-20 students per weekly session

Exhibit E-6. Year 1 Implementation Levels for Component 6: STEM Club

School	Full Component	Subcomponent 6-1: Teacher Participation in STEM Club	Subcomponent 6-2: Student Participation in STEM Club
Findlay	High implementation	One teacher participated in greater than 75% of STEM Club activities	An average of more than 30 students per STEM Club activity
Garside	Emerging implementation	One teacher participated in less than 60% of STEM Club activities	An average of less than 21 students per STEM Club activity
Gibson	Moderate implementation	One teacher participated in greater than 75% of STEM Club activities	An average of less than 21 students per STEM Club activity
Johnston	Emerging implementation	One teacher participated in 60%-75% of STEM Club activities	An average of less than 21 students per STEM Club activity

Exhibit E-7. Year 1 Implementation Levels for Component 7: STEM Summer Camp

School	Full Component	Subcomponent 7-1: Implementation of Planned Activities	Subcomponent 7-2: Teacher Participation in Summer Camp	Subcomponent 7-3: Student Participation in Summer Camp
Findlay	High implementation	Greater than 80% of planned activities were implemented	3 or more teachers per school participated in the summer camp	Less than 30 students per school attended the summer camp
Garside	High implementation	Greater than 80% of planned activities were implemented	3 or more teachers per school participated in the summer camp	Less than 30 students per school attended the summer camp
Gibson	High implementation	Greater than 80% of planned activities were implemented	3 or more teachers per school participated in the summer camp	Less than 30 students per school attended the summer camp
Johnston	High implementation	Greater than 80% of planned activities were implemented	3 or more teachers per school participated in the summer camp	Less than 30 students per school attended the summer camp

Exhibit E-8. Year 1 Implementation Levels for Component 8: Science and Math Tutoring

School	Full Component	Subcomponent 8-1: Teacher Participation in Science Tutoring	Subcomponent 8-2: Teacher Participation in Math Tutoring	Subcomponent 8-3: Student Participation in Science Tutoring	Subcomponent 8-4: Student Participation in Math Tutoring
Findlay	Emerging implementation	One teacher participated in less than 26% of tutoring sessions	One teacher participated in 26-85% of tutoring sessions	An average of less than 6 students per tutoring session	An average of less than 6 students per tutoring session
Garside	Moderate implementation	One teacher participated in 26-85% of tutoring sessions	One teacher participated in greater than 85% of tutoring sessions	An average of less than 6 students per tutoring session	An average of more than 10 students per tutoring session
Gibson	High implementation	One teacher participated in greater than 85% of tutoring sessions	One teacher participated in greater than 85% of tutoring sessions	An average of 6-10 students per tutoring session	An average of 6-10 students per tutoring session
Johnston	High implementation	One teacher participated in greater than 85% of tutoring sessions	One teacher participated in 26-85% of tutoring sessions	An average of more than 10 students per tutoring session	An average of 6-10 students per tutoring session

WestEd.org

730 Harrison Street
San Francisco, California 94107-1242