

# Schools to Watch: School Transformation Network

A U.S. Department of Education  
Investing in Innovation (i3) Development Grant

Final Evaluation Report

Center for Prevention Research and Development  
University of Illinois

September 2015

This work has been supported by the U.S. Department of Education's Investing in Innovation program, through Grant Number U396C101182. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education.

## Contents

<b>Executive Summary</b> .....	3
<b>i3 STW Project Intervention</b> .....	6
<b>i3 STW Project Components and Standards for Implementation</b> .....	8
<b>Evaluation Design</b> .....	10
School Characteristics and Demographics .....	11
Measurement Methods and Instrumentation .....	13
Impact Study .....	15
Implementation Study .....	17
Limitations .....	20
<b>Impact Study Findings</b> .....	20
Final Analytic Sample .....	20
Baseline Equivalence .....	21
Impact Study Estimates .....	21
<b>Implementation Study Findings</b> .....	23
Implementation Adaptations .....	23
Implementation Results by Programmatic Component .....	24
Challenges to Programmatic Component Implementation .....	25
Implementation Results by School .....	26
Impact of Level of Implementation .....	27
<b>Findings for Immediate and Intermediate Outcomes</b> .....	29
School Improvement Self-Study Teacher Survey Findings .....	30
School Improvement Self-Study Student Survey Findings .....	33
Student Subgroup Analysis .....	35
STW Rubric Findings .....	35

STW Designated Project Schools.....	36
Evolution of Coaching Services .....	40
Findings by STW Coach Turnover .....	44
Findings by Principal Turnover .....	46
Observations of Improvements at Project Schools.....	48
<b>Theory of Change .....</b>	<b>49</b>
<b>Unexpected Findings/Spillover Effects .....</b>	<b>52</b>
<b>Persistent Challenges .....</b>	<b>58</b>
<b>Sustainability and Growth .....</b>	<b>61</b>
<b>Lessons Learned .....</b>	<b>62</b>
<b>List of Project Reports, Papers, Presentations, and Products .....</b>	<b>66</b>
<b>References .....</b>	<b>69</b>
<b>Appendix A .....</b>	<b>71</b>
<b>Appendix B .....</b>	<b>73</b>

## Executive Summary

The Schools to Watch: School Transformation Network Project (i3 STW Project) is a whole school reform model designed to improve the educational practices, experiences, and outcomes of low-performing middle-grades schools. Developed by the National Forum to Accelerate Middle-Grades Reform, the four-year project was funded in 2010 by a U.S. Department of Education Investing in Innovation (i3) development grant. The framework for the project is the Forum's STW vision and criteria for high performing middle-grades schools, combined with a multi-layered system of support:

- Tools and data for assessment, goal setting, action planning, and monitoring;
- Technical assistance including a STW coach, a principal mentor, and a mentor STW school;
- Networking opportunities at the state and national levels;
- Implementing an early warning indicators system; and
- Engaging in focused professional development to build a learning community and address the needs of students at risk of educational failure.

The project was implemented in California, Illinois and North Carolina by the California League of Middle Schools, the Association of Illinois Middle-Grade Schools, and the North Carolina Association for Middle Level Education. The evaluation was conducted by the Center for Prevention Research and Development at the University of Illinois.

### *Evaluation Design*

The evaluation employed a quasi-experimental design which included formative evaluation to inform the intervention and summative evaluation to assess the impact on student achievement. Two student cohorts were tracked over four years at 34 schools (17 intervention and 17 comparison). The intervention sample was comprised of persistently low-performing middle-grades schools serving high need students. Comparison schools were selected using key demographics to match to intervention schools within each state. Comparison schools did not receive the intervention.

The outcome data for the impact study included student English and math achievement scores on annual standardized state assessments (California Standards Test, Illinois Standards Achievement Test, and North Carolina End-of-Grade Test). To examine achievement scores between intervention and comparison students, a series of 2-level models (students within schools) were run to assess 8<sup>th</sup> grade achievement (i.e., after students received all three years of the intervention). Several process and measurement tools for assessing implementation and intermediate outcomes were also used, including the STW rating rubric, teacher and student surveys, an online coach's log, and focus groups.

### *Achievement Findings*

The analyses showed no overall intervention effects on either English or math student achievement. In other words, the students that received the i3 STW Project intervention performed the same as the comparison students. Despite the fact that the overall impact analyses did not find an intervention effect on either English scores or math scores, it should be noted that significant results were found for the highest implemented schools, those project schools that achieved STW designation during the project. Additionally, other hypotheses addressing immediate and intermediate outcomes and levels of implementation are important to these findings since they include the realities and complexities of the

influence of associated factors, such as culture, climate, instructional practices, etc. on achievement. We hypothesize that achievement scores take the longest to improve and will not improve without corresponding deeply imbedded improvements in the teaching and learning environment.

### ***Implementation Findings***

The i3 STW Project was implemented with fidelity at the majority of project schools, although implementation in the first two years of the project was not as complete as the last two years due to startup challenges and the development of protocols for service delivery. There was some variability in implementation by school, with smaller schools having higher overall implementation. Challenges related to coach turnover, diversity of state requirements, and availability of school personnel resulted in both adaptations to a few programmatic components as well as several components being implemented less completely than others. Implementation strongly influenced programs and practices at project schools, however, in that schools with higher implementation had statistically higher practices and outcomes, including the STW criteria, collaboration practices, and instructional practices. These findings suggest the importance of monitoring implementation and having well-defined guidelines for implementation that support consistency.

### ***Culture, Collaboration, and Instruction Findings***

Evidence indicates that i3 STW Project schools improved their culture and climate, collaboration practices, leadership practices, STW criteria, and classroom instructional practices.

- **Culture and climate** – Improvements in work climate, classroom climate, teacher decision making opportunities, and collective teacher efficacy.
- **Collaboration practices** – Increased team practices, team decision making, and quality of collaborative interactions.
- **Leadership practices** – Improvements in administrators communicating with faculty, follow-through on decisions, problem-solving, and a proactive approach to improvement.
- **STW criteria** – Higher implementation of strategies to support academic excellence, developmental responsiveness, social equity, and organizational structures.
- **Classroom instructional practices** – Increased recommended practices such as small group instruction, integration and interdisciplinary practices, and critical thinking practices.

Analyses to explore these contextual changes within different settings showed that project schools with a consistent STW coach throughout the four-year project had higher STW criteria, collaboration practices, and instructional practices.

### ***Impact on STW Designated Project Schools***

Three i3 STW Project schools were successful in making substantial improvements in both programs and practices (STW criteria, collaboration, instructional practices) and math achievement during the grant period. Therefore, they applied and were designated as a STW school by the National Forum before the end of the grant. Project stakeholders hypothesized that the reason for their rapid turnaround was threefold. First, they had a high degree of readiness for improvement among the principal and faculty that was unique among project schools and allowed them to gain quick teacher buy-in for the project. Second, they had a strong principal leader who was not only the curriculum

leader of the school but also the model for the teachers. Finally, stakeholders pointed to the strong coaching services team (STW coach, principal mentor, instructional coach) that was paired with each of these schools, saying “I think it was the right combination of coaches and their abilities.” A potential hypothesis is that the remaining project schools are on the pathway to these same improvements; however, because these three schools started the project at a higher level of readiness, the remaining project schools may need more time investment to achieve the same outcomes.

### **Challenges**

The i3 STW Project, like any complex and multi-faceted endeavor, faced a variety of challenges during implementation. The most persistent challenges included:

- Disruption to implementation from unexpected turnover of coaches, principals, and district staff.
- Balancing structured implementation with the complexities and realities of each state’s policies.
- Programmatic components that needed to be re-defined to meet the needs of schools.
- Changing district requirements that often took precedence over the project.
- Changing state standardized achievement tests to align with the common core standards.

### **Lessons Learned**

There are a myriad of lessons learned from the i3 STW Project about how to improve low-performing middle grades schools, including:

- The importance of coaching as the factor most impactful to improvement.
- Collaborative leadership which empowered teachers to actively participate and own the project.
- The STW criteria and rubric as a guiding vision, common language, and framework.
- The importance of a continuous improvement model to inform, set goals, reflect, and evaluate.
- The powerful impact of STW visits which allowed teachers to observe best practices.
- The value of participating in a network of schools to gain knowledge, resources, and support.
- The importance of cultivating school district involvement and support throughout the project.
- Project oversight and management practices for communication, collaboration, and refinement.

### **Conclusion**

The results of the evaluation provide unique insight into the reform process for i3 STW Project schools as well as other middle-grades schools that are struggling to improve. The multiple supports that i3 STW Project schools received combined with the guiding vision of the STW criteria and rubric have allowed these high need schools to improve. Culture and climate were the first areas to improve, followed by collaboration and leadership, then finally best practices within the STW framework and best teaching and learning practices. We cannot assume that this progression will be identical in all low-performing schools, however, many have similar challenges and therefore we hypothesize that these key elements will be part of the change process at low-performing schools. Although the theory of change does not yet connect these contextual improvements in culture and climate, collaboration, leadership, STW criteria, and classroom instruction with student achievement growth, it is encouraging to find these improvements in programs and practices because improvements in these areas are the building blocks to improvements in student achievement.

### i3 STW Project Intervention

The Schools to Watch (STW): School Transformation Network Project is a whole school reform model intended to improve the educational practices, experiences, and outcomes of low-performing middle-grades schools. Designed and administered by the National Forum to Accelerate Middle-Grades Reform, the four-year project was funded in 2010 by a U.S. Department of Education (USDE) Investing in Innovation (i3) development grant. This unique grant program is designed to develop, test, validate, and scale-up promising innovations to our country's educational challenges.

The STW: School Transformation Network Project (i3 STW Project) was implemented in 18 urban and rural middle-grades schools, serving students in grades 6 to 8, in California, Illinois, and North Carolina from October 2010 to September 2014. The framework for the project is the National Forum's STW criteria, a set of strategies and practices developed by the Forum which posits that high performing middle-grades schools are academically excellent, developmentally responsive, socially equitable, and establish organizational structures that provide students with high-quality teachers and supports. The project is guided by a logic model (Figure 1).

---

*The National Forum to Accelerate Middle-Grades Reform is an alliance of over 60 educators, researchers, national associations, and officers of professional organizations and foundations committed to promoting the academic performance and healthy development of young adolescents.*  
[www.middlegradesforum.org](http://www.middlegradesforum.org)

---

In addition to using the STW criteria as the vision for improvement, the project also provides project schools with an intensive multi-layered system of support throughout the four years of the project, including: providing tools and data for assessment, goal setting, action planning and monitoring; providing technical assistance including a STW coach, a principal mentor, and a mentor STW school; engaging in networking opportunities at the state and national levels; implementing an early warning indicators system; and engaging in focused professional development designed to build a learning community and address the needs of students at risk of educational failure. The project was extended to a fifth year with a no-cost extension approved by the USDE, allowing the team to continue to support project schools with limited services, such as a STW coach and the administration of the online STW rubric. The fifth year, however, was not part of the evaluation.

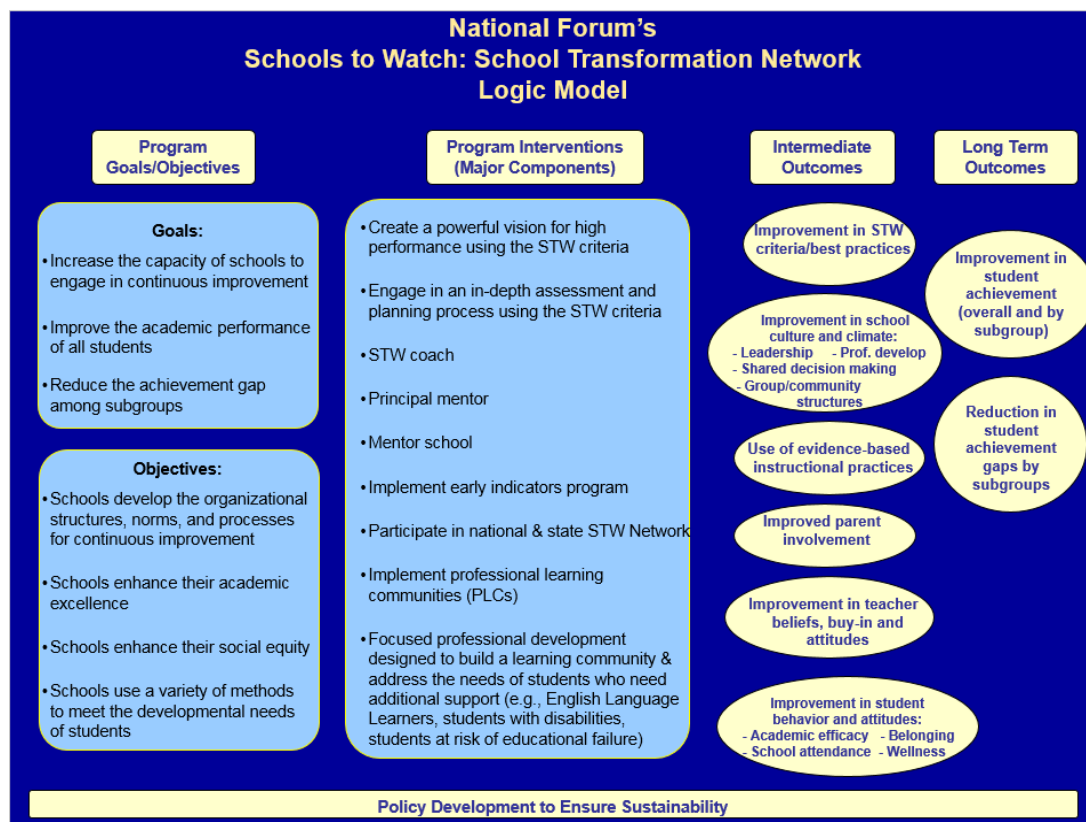
California, Illinois, and North Carolina, three states that are part of the National Forum's network, were chosen to serve as hubs for this project. These state hubs led and managed the implementation of the project at the schools in their state. The lead organization of each state hub are middle-grades leaders within their state and highly experienced in working with middle-grades schools. In California, the project hub is led by the California League of Middle Schools ([www.leagueofschools.org](http://www.leagueofschools.org)). In Illinois, the Association of Illinois Middle-Grade Schools ([www.aimsnetwork.org](http://www.aimsnetwork.org)) leads the project. And in North Carolina, the North Carolina Association for Middle Level Education ([www.ncmle.org](http://www.ncmle.org)) leads the work. These organizations delivered all project services to the schools in their state (6 schools per state) and managed the overall implementation of the project in their state, including hiring and training coaches

and mentors, providing professional development to schools, managing communication, and ensuring that the project is being implemented with high quality.

*The Schools to Watch (STW) Program operates in 18 states. It provides a process and set of criteria for identifying schools on a sustained, upward trajectory of growth and continuous improvement around academic excellence, developmental responsiveness, social equity, and organizational supports. To date, the National Forum has identified 382 schools as STW. ([www.middlegradesforum.org/what-is-school-to-watch](http://www.middlegradesforum.org/what-is-school-to-watch))*

The evaluation of the i3 STW Project was conducted by the Center for Prevention Research and Development (CPRD) at the University of Illinois under the leadership of Ms. Nancy Flowers and the CPRD research team. CPRD has a long history of conducting middle grades evaluation and research to understand whole school reforms in terms of structures, organization, instruction, process and outcomes (Flowers, Mertens, & Mulhall, 2003). CPRD also has a strong commitment to understanding and using data for decision making and continuous quality school improvement.

Figure 1. i3 STW Project Logic Model



Over time, project schools are expected to: 1) strengthen their structures, norms, and processes for continuous improvement; 2) increase their academic rigor; 3) promote equity for all students; and 4) develop an array of supports designed to meet the needs of young adolescents. Intermediate outcomes



expected included: improvement in STW criteria implementation, improvement in school culture and climate, use of evidence-based instructional practices, improved parent involvement, improvement in teacher beliefs, buy-in, and attitudes, and improvement in student behavior and attitudes. The long term outcomes of the i3 STW Project are to improve student achievement in mathematics and English/language arts/reading and achieve a reduction in student achievement gaps by subgroups (students with disabilities and English language learners).

### i3 STW Project Components and Standards for Implementation

The project leaders (i.e., the three state hubs, the National Forum, and CPRD) worked together at the beginning of the grant to identify and define the key programmatic components (i.e., program interventions) to be implemented at each project school as part of the i3 STW Project. Guided by the project's logic model, nine key programmatic components were identified, including:

1. Create a powerful vision for high performance using the STW criteria
2. Engage in an in-depth assessment and planning process using the STW criteria
3. STW coach
4. Principal mentor
5. Mentor school
6. Implement early indicators program
7. Participate in national and state STW network
8. Implement professional learning communities (PLCs)
9. Implement focused professional development to build learning communities and address the needs of students

In addition to identifying the programmatic components, the project leaders also developed the activities associated with their implementation, and a definition of what high implementation looks like (Table 1). State hub leaders then worked with their team of coaches and mentors to implement each component at every project school during each year of the four-year project.

Table 1. i3 STW Project Key Programmatic Components and Fidelity of Implementation Definitions

i3 STW Project Key Programmatic Components	Activities Associated With Key Programmatic Components	Definition of High Implementation
Create a powerful vision for high performance using the STW criteria	Planning and overview meeting for school and district staff; Teachers complete the self-rating STW Rubric; Review data results from the STW Rubric with school leadership team; Faculty supports the vision for the school.	A school is a high implementer if school and district staff attend a planning and overview meeting; at least 76% of teachers complete the self-rating STW Rubric; it reviews the data results from the STW Rubric with school leadership team; and at least 76% of faculty support the vision for the school.
Engage in an in-depth assessment and planning process using the STW criteria	Examine data and identify areas for improvement with school leadership team; Develop clear	A school is a high implementer if it examines data and identifies needs with school leadership team; develops clear

	and measurable goals with school leadership team; Implement an action plan with performance benchmarks.	and measurable goals with school leadership team; and implements an action plan with performance benchmarks.
STW coach	Planning and overview meeting for STW coaches; STW coaches work with schools; STW coaches participate in ongoing training.	A school is a high implementer if its STW coach attends a planning meeting; its STW coach makes 9 or more visits to the school to work them in Year 1 and 16 or more visits per year in Years 2-4; and its STW coach participates in ongoing training.
Principal mentor	Planning and overview meeting for principal mentors; Principal mentors work with principals; Principal mentors participate in ongoing training.	A school is a high implementer if its principal mentor attends a planning meeting; its principal mentor makes 5 or more contacts with principal to work with him/her in Year 1, and 9 or more contacts per year in Years 2-4; and its principal mentor participates in ongoing training.
Mentor school	Mentor schools serve as role models and work with school faculty and/or leadership team.	A school is a high implementer if its school faculty or leadership team work with a mentor school and/or visit a mentor school at least twice per year.
Implement early indicators program	Preliminary training on early indicators (Year 1); Comprehensive tailored training on early indicators (Year 2); Implement a process for regularly compiling and examining student data; Implement a process for regularly identifying students who need additional support; Implement a 3-tiered intervention process to assist "off-track" students.	A school is a high implementer if it attends a preliminary training on early indicators for schools in Year 1; it attends a comprehensive training on early indicators in Year 2; it implements a process for regularly compiling and examining student data; it implements a process for regularly identifying students who need additional support; and it implements a 3-tiered intervention process to assist "off track" students.
Participate in national and state STW Network	Schools access peer support from the national and state STW network.	A school is a high implementer if it accesses peer support from the national and state STW network by attending at least 2 events per year.
Implement professional learning communities (PLCs)	Schools create PLC structures for teachers.	A school is a high implementer if it creates PLC structures for teachers.
Implement focused professional development to build learning communities and address the needs of students	Schools participate in PD on PLC structures and functioning; Schools participate in PD on meeting the needs of at-risk students; Schools participate in other PD focused on their identified needs.	A school is a high implementer if it participates in at least one PD session on PLC structures and functioning; it participates in at least one PD session on meeting the needs of students who are most at risk; and it participates in at least 2 other PD sessions focused on their identified needs each year.

## Evaluation Design

The evaluation employed a quasi-experimental design (QED) with mixed-methods data collection. This design allowed the evaluation team to conduct formative evaluation to inform and refine the intervention throughout the project (a key aspect of i3 development grants), and summative evaluation to assess change between intervention schools versus comparison schools in the long term outcome of the project and student achievement scores. Two cohorts of students were tracked over the four-year project period with unique identification numbers (Table 2).

“A quasi-experimental design compares outcomes for students, classrooms, or schools who had access to the intervention with those who did not but were similar on observable characteristics.”

What Works Clearinghouse, 2014 (p. 10)

Table 2. i3 STW Project Student Cohorts Tracked

	Project Year 1 (2010/11)	Project Year 2 (2011/12)	Project Year 3 (2012/13)	Project Year 4 (2013/14)
Cohort 1	6 <sup>th</sup> Grade	7 <sup>th</sup> Grade	8 <sup>th</sup> Grade	--
Cohort 2	--	6 <sup>th</sup> Grade	7 <sup>th</sup> Grade	8 <sup>th</sup> Grade

The evaluation included 34 middle-grades schools (17 intervention and 17 comparison) from the three project states (California, Illinois, North Carolina). Due to district restructuring, one of the original 18 intervention schools in North Carolina was closed after the second year of project implementation, therefore the final analytic sample for the evaluation was reduced to 34 schools.

The intervention sample was comprised of persistently low-performing public schools serving high need students (i.e., high percentages of students in poverty and English learners) in either an urban or rural location. The intervention schools were selected prior to the application for funding being submitted to the USDE. The project leaders developed a set of criteria that each intervention school must meet in order to be considered for inclusion in the project, including: 1) Public school (no charter, private, or military schools); 2) Middle-grades school with a grade configuration of 6-8 or 5-8, or a middle-grades program within a K-8 school; 3) At least 40% of students participating in free/reduced lunch program; 4) Diverse student ethnicity; 5) Persistently low-performing school; and 6) Either an urban or a rural location. Project leaders used the criteria to identify eligible schools and contacted them to discuss the project and recruit their participation. It is considered a convenience sample with well-defined criteria.

The comparison schools were selected using key demographics to match to the intervention schools within each state. The schools in the comparison group did not receive the intervention. Students and teachers in comparison schools continued with their regular curriculum (business as usual). The matching was conducted at the school level within state. Schools were matched on the following:

- State
- District demographics (geographic location, size)

- School size
- Student demographics (minority, ELL, free/reduced lunch, and special education)
- Achievement performance of the schools on standardized tests
- No Child Left Behind (NCLB) Adequate Yearly Progress (AYP) status
- NCLB AYP status history
- Reason for failing NCLB AYP

When possible, comparison school matches were made within the same district as the intervention schools, however, the numbers of potential matches for intervention schools by district varied across states. For intervention schools in Chicago Public Schools, for example, matches were made within the district because the district was large enough. However, for the rural districts, we were unable to identify comparison schools within the same district due to their smaller size and therefore cross-district matching was necessary. In order to ensure that cross-district matching was done amongst comparable districts, the U.S. Census Bureau designation of size was used. The most similar schools were identified using the variables listed above. Schools were matched first on student achievement performance, allowing only a 5% margin of difference. Second, they were matched on the NCLB AYP status and NCLB AYP status history, again allowing only a 5% margin of difference. Finally, schools were matched on the remaining variables, allowing a 10% margin of difference.

### ***School Characteristics and Demographics***

The i3 STW Project was implemented in five school districts in California, Illinois, and North Carolina. Three of the districts, Compton Unified (California), Chicago Public Schools (Illinois), and Durham Public Schools (North Carolina) are large urban districts. The remaining two districts, Coalinga-Huron Unified (California) and Richmond County Schools (North Carolina) are located in rural communities. Eighteen schools serving the middle grades from these

districts participated in the project. In California and North Carolina, all participating schools were stand-alone middle grades schools containing grades 6<sup>th</sup> to 8<sup>th</sup>. In Illinois, the participating schools were middle grades programs serving grades 6<sup>th</sup> to 8<sup>th</sup> within a kindergarten to 8<sup>th</sup> grade building. All schools

**Low performing schools are marked by a limited capacity, a lack of high quality teaching, and frequent turnover of principals and other staff.**

The National Education Writers Association  
Research Brief on School Turnarounds (Brownstein, 2012)

were considered persistently low-performing at the beginning of the project because they were either classified with a status of improvement, corrective action, or restructuring by their state, or they lacked progress over multiple years on state achievement assessments. Prior to the beginning of the project, a significant percentage of students were not meeting expectations on state standardized tests (Table 3).

Overall, the i3 STW Project schools had an average of 85% free/reduced lunch students and contained 85% minority students (Table 4). Among the minority students, more than 50% are of Hispanic ethnicity (Figure 2). An average of 27% of students were identified as ELL by their school district. The student enrollment at project schools ranges from seven small schools (less than 300 students), six medium schools (between 301 to 700 students), and five large schools (more than 700 students). The overall average enrollment of all schools is 491 students. Twelve project schools are located in large urban settings. The remaining six schools are in rural locations.

Table 3. i3 STW Project School Pre-Intervention Achievement Levels

Test Subjects	2010* % of Students Meeting or Exceeding Standards		
	California Project Schools	Illinois Project Schools	North Carolina Project Schools
<b>English/Language Arts/Reading</b>			
6 <sup>th</sup> Grade	28%	57%	59%
7 <sup>th</sup> Grade	28%	56%	47%
8 <sup>th</sup> Grade	27%	76%	50%
<b>Mathematics</b>			
6 <sup>th</sup> Grade	28%	66%	64%
7 <sup>th</sup> Grade	32%	69%	64%
8 <sup>th</sup> Grade	21%	76%	67%

\* Data from state standardized achievement tests in 2010, prior to the start of the project.

The project schools were characterized at the start of the project as highly challenging schools. Observations by project leaders early in the project described the schools as having highly negative and dysfunctional climates, isolated teachers, lack of teacher trust in the administration, student behavior problems, lack of parent involvement, lack of focus, reactive instead of proactive, and very little instructional rigor. Administrators at the schools all expressed the will to improve, but needed the skills, resources, and vision to positively impact their educational practices, experiences, and outcomes.

*“At the beginning, the project schools were low performing. They were struggling. They did not have a strong level of inter-staff communication and collaboration...”*

i3 STW Project State Hub Leader

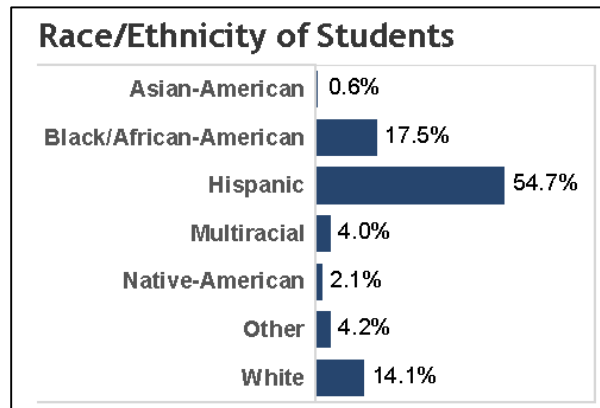
Table 4. i3 STW Project School Demographics

Districts	Schools	Grade Levels	Enrollment	Percent Free/Reduced Lunch	Percent Minority	Percent English Language Learners
Compton Unified Compton, CA (Urban)	Bunche MS	6-8	734	97%	100%	40%
	Davis MS	6-8	1,135	98%	99%	36%
	Vanguard Learning Center	6-8	348	78%	99%	22%
	Whaley MS	6-8	904	91%	98%	38%
Coalinga-Huron Unified Coalinga, CA (Rural)	Coalinga MS	6-8	665	73%	69%	34%
	Huron MS	6-8	365	100%	98%	77%
<b>California Totals</b>			4,151	90%	94%	41%
Chicago Public Schools	Corkery ES	K-8	172	98%	99%	39%

Chicago, IL (Urban)	Davis ES	K-8	283	98%	99%	49%
	Kanoon ES	K-8	184	97%	100%	55%
	Mason ES	K-8	157	99%	100%	0%
	Penn ES	K-8	109	92%	99%	12%
	Whitney ES	K-8	306	95%	99%	50%
<b>Illinois Totals</b>			<b>1,211</b>	<b>97%</b>	<b>99%</b>	<b>34%</b>
Durham Public Schools Durham, NC (Urban)	Githens MS	6-8	981	41%	79%	14%
	Lowe's Grove MS	6-8	656	77%	91%	12%
Richmond County Schools Hamlet, NC (Rural)	Ellerbe MS	6-8	237	79%	59%	11%
	Hamlet MS	6-8	568	70%	51%	2%
	Rockingham MS	6-8	751	67%	46%	2%
	Rohanen MS	6-8	297	78%	41%	2%
<b>North Carolina Totals</b>			<b>3,490</b>	<b>69%</b>	<b>61%</b>	<b>7%</b>
<b>Project Totals</b>			<b>8,852</b>	<b>85%</b>	<b>85%</b>	<b>27%</b>

Notes: Data from 2010/11 (Year 1). Enrollment for Illinois includes only grades served (6-8).

Figure 2. i3 STW Project Student Race/Ethnicity



### Measurement Methods and Instrumentation

The teacher and student-level data were collected via the School Improvement Self-Study Teacher and Student Surveys. These surveys are part of a data collection system of surveys for middle-grades schools, developed by CPRD in 1990. The surveys are grounded in research and have been used with more than a thousand schools, with results widely disseminated (Flowers & Mertens, 2003; Flowers, Mertens, & Mulhall, 1999, 2000a, 2000b, 2002, 2003, 2007; Mertens & Flowers, 2003, 2006; Mertens, Flowers, Hesson-McInnis, & Bishop, 2006, 2007). The teacher survey is comprised of numerous constructs related to the teaching and learning process including: interdisciplinary teaming practices, quality of team interactions, team decision making, work climate, collective teacher efficacy, teacher decision making, administrative leadership, and classroom instructional practices. The student survey is comprised of constructs that are focused on student attitudes about academics (academic efficacy and academic expectations), family involvement in the education, and school environment (belonging and

school climate). The items that make up each construct on the teacher and student surveys were combined and scale scores were calculated based on Cronbach's alpha (.76 to .96) See Appendix A.

The school-level data were collected via the STW Rating Rubric and the Coach's Log. The STW Rating Rubric is a tool developed by the Forum in 1998 that measures the implementation of the criteria associated with the four components of the STW Program. The rubric is used by middle grades schools to study and rate their practices as part of a continuing improvement process as well as part of a mandatory self-rating for schools interested in applying for a STW designation. The four STW program components measured by the rubric include:

1. Academic Excellence – Schools challenge students to use their minds well;
2. Developmental Responsiveness – Schools are sensitive to the unique developmental challenges of early adolescence;
3. Social Equity – Schools are democratic and fair, providing every student with high-quality teachers, resources, learning opportunities and supports; and
4. Organizational Structures and Processes – Schools establish norms, structures, and organizational arrangements to support and sustain their trajectory toward excellence.

For each of the four STW program components, teachers at intervention schools rated their school's level of implementation by responding to 37 general criteria and 100 concrete examples of excellence. Teachers used a metric ranging from one to four where: 4 = High quality, complete, mature, and coherent implementation – NEARLY PERFECT, LITTLE ROOM FOR IMPROVEMENT; 3 = Good quality, maturing but not fully implemented by all – GOOD QUALITY BUT STILL ROOM FOR REFINEMENT AND IMPROVEMENT; 2 = Fair quality, mixed implementation, immature practice, sporadic by some – SIGNIFICANT IMPROVEMENT NEEDED; and 1 = Poor quality, low level of implementation, new program – CONSIDERABLE PLANNING, CONSENSUS BUILDING AND IMPROVEMENT NEEDED.

The Coach's Log is an electronic data collection system for project coaches and principal mentors to complete after each visit or activity with a project school. The purpose of the log was to document the number of visits made to schools, the purpose of the visit (e.g., training, professional development, reviewing data, etc.) the outcome of the visit (e.g., goals set, action plan distributed, etc.), as well as to log the improvement progress of the school.

Several analytic methods were employed to examine the different data sources used for the exploratory analyses. Initially, data were analyzed using basic descriptive statistics to understand the patterns in the data. To test the hypothesis of whether there was any change in ratings on the measures or constructs across project years, 2-level modeling was used considering teachers or students nested within schools, multivariate analysis of variance (MANOVA) was used in order to handle those constructs with multiple dependent scales, and finally Univariate Analysis of Variance was used. Focus group data and coach's log data were analyzed using an inductive, constant-comparative method to identify categories for coding and common themes.

### **Impact Study**

The central research questions guiding the impact study were to explore whether the project had a positive impact on student achievement at intervention schools. The two confirmatory research questions for the impact study were:

1. Does the i3 STW Project have a positive impact on English/language arts/reading achievement as measured by state standardized tests for two cohorts of students starting in 6<sup>th</sup> grade in low-performing schools at the end of the three year intervention period?
2. Does the i3 STW Project have a positive impact on mathematics achievement as measured by state standardized tests for two cohorts of students starting in 6<sup>th</sup> grade in low-performing schools at the end of the three year intervention period?

The outcome data for the impact study, student English/language arts/reading achievement scores and student mathematics (math) achievement scores on annual standardized state assessments, were collected every year from the state departments of education or districts for intervention and comparison schools. These outcomes represent two separate domains. In California, the California Standards Test (CST) was used for English/language arts and math. In Illinois, the Illinois Standards Achievement Test (ISAT) was used for reading and math. In North Carolina, the North Carolina End-of-Grade (EOG) Test was used for reading and math. These state tests are standardized and we assume that they were constructed with high levels of validity and reliability. The outcomes align with the project's logic model. The English/language arts/reading and math assessments are broad enough to detect intervention effects, while at the same time, not overly aligned to the intervention.

Each test was administered annually in 2011, 2012, and 2013. In 2014, the tests were administered in Illinois and North Carolina, but California did not administer the CST nor any other standardized test, therefore intervention and comparison school data were not available in 2014 for California. There were also several other notable issues related to the achievement tests. In 2013, North Carolina implemented a new standard course of study in all subjects and grades, aligned to the common core. Paired with the instructional change was a new state student assessment. Since the North Carolina test was changed in 2013, correlational analyses were conducted between the scores from the new 2013 test and prior years' scores (i.e., 2011 and 2012) to determine whether the 2013 and 2014 data were appropriately aligned with prior years in order to use them in the analyses, and they were. The correlations were fairly strong (between .76 and .84). Also in 2013, Illinois revised some items on the ISAT in order to align to the new common core state standards, however, the majority of the test remained intact.

All three states administer the standardized state assessment tests in the spring by providing all districts in the state with a two to three week window to complete the testing. Since intervention schools were matched to comparison schools within the same state, timing and data collection procedures for the state assessments were the same for both groups (Table 5).



Table 5. i3 STW Project Outcomes for Impact Study

Outcome Domain	Instrument Name	Construction of Outcome Measure	Unit of Observation	Timing of Outcome Measure	Reliability
English/ Language Arts/ Reading	California Standards Test* Illinois Standards Achievement Test North Carolina End-of-Grade Test	Continuous	Student	Cohort 1: Spring 2011, 2012, 2013 (end of 6 <sup>th</sup> , 7 <sup>th</sup> , 8 <sup>th</sup> )  Cohort 2: Spring 2012, 2013, 2014 (end of 6 <sup>th</sup> , 7 <sup>th</sup> , 8 <sup>th</sup> )	Internal Consistency .89-.94
Math	California Standards Test* Illinois Standards Achievement Test North Carolina End-of-Grade Test	Continuous	Student	Cohort 1: Spring 2011, 2012, 2013 (end of 6 <sup>th</sup> , 7 <sup>th</sup> , 8 <sup>th</sup> )  Cohort 2: Spring 2012, 2013, 2014 (end of 6 <sup>th</sup> , 7 <sup>th</sup> , 8 <sup>th</sup> )	Internal Consistency .91-.94

\* California Standards Test was not administered in 2014 because the state suspended all standardized testing in 2013/14.

Baseline comparisons were conducted using 5<sup>th</sup> grade achievement scores (the year before the intervention took place) on the final analytic sample to determine student-level equivalence between the intervention group and the comparison group. Differences in baseline means were calculated using a 2-level model (students nested within schools). The 5<sup>th</sup> grade achievement scores were considered as the dependent variable whereas the school-level intervention indicator was used as the only covariate. The means for the intervention and comparison schools were the model adjusted means. When the standardized differences in baseline achievement scores were examined, if these differences would exceed .05 standard deviation units between intervention and comparison, we would make adjustments through the covariates. Although the final analytic sample didn't show any baseline difference, since the standard deviation units exceeded .05, adjustments were made in the analyses at both the student and school levels (see Impact Study Findings section).

For the impact study, the two cohorts of intervention students and comparison students starting in 6<sup>th</sup> grade in either the 2010/11 or 2011/12 school year were tracked over three years by matching individual students over time (Table 2). All students enrolled in intervention schools and comparison schools as 6<sup>th</sup> graders in 2010/11 or 2011/12 were part of the analytic sample. Students who entered the schools post baseline were excluded. Additionally, students with missing 5<sup>th</sup> grade pretest achievement scores or missing 8<sup>th</sup> grade achievement scores were also excluded.

Achievement test scaled scores for English/language arts/reading and math for both intervention students and comparison students were converted to z-scores, using the mean and standard deviation of the state test for each year per the test technical manuals, to allow comparison of scores over time and across different state tests. The z-score formula used was:  $Z = \frac{x - \mu}{SD}$  where  $x$  is the student's scale score,  $\mu$  is the corresponding state mean, and  $SD$  is the corresponding state standard deviation. Using this formula, a z-score was calculated for each individual student for each testing year.

To examine achievement scores between intervention students and comparison students, a series of 2-level models (students within schools) were run to assess the 8<sup>th</sup> grade achievement data. The reason for considering 8<sup>th</sup> grade data was the fact that by then, students received all three years of the intervention. The models controlled for the following covariates: student-level baseline achievement scores, ethnicity, and special education status; and school-level baseline achievement scores, urban/rural settings, state, percentage of free/reduced lunch students, and percentage of ELL students. STATA xtmixed was used to execute the models. The effect size was calculated using Hedges's *g* ES index. It is defined as the standardized mean difference, which is the difference between the mean outcome of the intervention group and the mean outcome of the comparison group, divided by the pooled within-group standard deviation (SD) on 8<sup>th</sup> grade achievement scores. Since the achievement scores are at the student level, the student-level standard deviation was used. The formula for computing standardized mean difference is as follows:  $g = (X_1 - X_2) / S_{pooled}$  where  $X_{intervention}$  and  $X_{control}$  are the means of the 8<sup>th</sup> grade achievement scores for the intervention and comparison group, respectively, and  $S_{pooled}$  is the pooled within-group SD of the achievement scores at the student level

and  $S_{pooled} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}}$  where  $n_1$  and  $n_2$  are the student sample sizes, and  $S_1$  and  $S_2$  are the student-level SDs for the intervention and comparison group, respectively.

Additional exploratory analyses to examine various key student sub-groups, ELL students and special education students, were also conducted in order to examine whether the intervention reduced the achievement gap between these students and the general population of students.

### **Implementation Study**

The impact study was relatively straightforward in terms of tracking and comparing achievement scores of different study cohorts. However, overall achievement effects need to be considered in the context of the implementation of the i3 STW Project framework with schools that have large disadvantaged student populations and higher diverse community settings. To that end, the evaluation team used four process measurement tools (Table 6) for assessing the implementation of the project and the adoption of key indicators. First, the Forum's STW Rating Rubric was administered annually at intervention schools to assess the implementation of the criteria associated with the four key components of the STW Program (organizational capacity, academic excellence, social equity, and developmental responsiveness). Second, CPRD's School Improvement Self-Study Surveys of teachers, students, and parents were administered every year at intervention schools to provide formative feedback on measures such as teacher reports of developmentally appropriate instructional practices and preparation (e.g., parent involvement, professional development, collaborative planning, organizational structures, authentic instruction and assessment, and academic focus, etc.), and summative information such as the intermediate outcomes of culture and climate, leadership, shared decision making, teacher attitudes, and student attitudes and behavior. Third, an online Coach's Log was completed on an ongoing basis by STW coaches and principal mentors. They reported their activities with intervention schools, including who they worked with, the purpose of the activity, decisions made, barriers, and any other pertinent information. Finally, qualitative data were collected via focus groups with project principals, STW coaches, and project mentors. The purpose of the focus groups was to provide formative results regarding the experiences of implementation and to assess the multi-layered

system of support (e.g., coaching, mentoring activities, networking, etc.). School-level, state-level, and national-level data reports from all four of these process measurement tools (STW Rubric, Self-Study Surveys, Coach's Log, and Focus Groups) were disseminated annually to project schools, state hub teams, and the project leadership team for use in monitoring improvement progress, setting goals, and refining the intervention.

Table 6. i3 STW Project Data Components and Sample Sizes

Data Component	Year 1 (2010/11)	Year 2 (2011/12)	Year 3 (2012/13)	Year 4 (2013/14)
STW Rubric	361 staff	402 staff	478 staff	450 staff
Self-Study Surveys	275 teachers 5,315 students 1,820 parents	395 teachers 6,520 students 3,550 parents	439 teachers 6,189 students 3,653 parents	435 teachers 5,869 students 3,140 parents
Coach's Log	179 visits by coaches and mentors	660 visits by coaches and mentors	739 visits by coaches and mentors	616 visits by coaches and mentors
Focus Groups	10 coaches/mentors	9 coaches/mentors 8 principals	6 coaches/mentors 5 principals	5 coaches/mentors 7 principals

Additionally, a measurement tool to track fidelity of implementation of each of the nine key programmatic components (e.g., create a vision, engage in planning process, STW coach, networking, etc.) of the project was developed (Table 7 contains an abridged version and Appendix B contains the full version). Using the definition of each component, its associated activities, and what high implementation looks like, a data source for each was identified and a scoring system (high, medium, low) was developed to assess level of implementation. Data were collected every year of the project from intervention schools on their activities related to the implementation of each programmatic component. Using these data, an implementation score was calculated for each activity. The individual activity scores were then summed to calculate an implementation score for the programmatic component. The project leaders then defined what constitutes adequate implementation for the entire sample for each programmatic component (i.e., 70% of project schools must have high implementation of this component to meet fidelity). These scores and implementation thresholds allowed the evaluation team to assess fidelity of implementation of each programmatic component across the sample for each year of the project.

Table 7. i3 STW Project Fidelity of Implementation Data Collection and Scoring Thresholds

Key Programmatic Components	Data Source(s)	Implementation Score*	Threshold for Fidelity of Implementation for the Sample
Create a powerful vision for high performance using the STW criteria	Coach's log, Online STW Rubric data collection system, Self-Study Teacher Survey	Low = 0-8 Medium = 9-12 High = 13-16	At least 70% of schools: Years 1-3 have high implementation, Year 4 has medium or high implementation

Engage in an in-depth assessment and planning process using the STW criteria	Coach's log	Low = 0 Medium = 1-3 High = 4-6	At least 80% of schools have high implementation
STW coach	State hub activity reports, Coach's log	Low = 0-2 Medium = 3-5 High = 6-8	At least 70% of schools have high implementation
Principal mentor	State hub activity reports, Coach's log	Low = 0-2 Medium = 3-5 High = 6-8	At least 70% of schools have high implementation
Mentor school	State hub activity reports	Low = 0 Medium = 1 High = 2	At least 70% of schools have medium or high implementation
Implement early indicators program	State hub activity reports	Years 1-2: Low = 0-2 Medium = 3-8 High = 9-10 Years 3-4: Low = 0-1 Medium = 2-8 High = 9	At least 60% of schools have high implementation
Participate in national and state STW Network	State hub activity reports	Low = 0 Medium = 1 High = 2	At least 70% of schools have high implementation
Implement professional learning communities (PLCs)	State hub activity reports	Low = 0 High = 1	At least 60% of schools have high implementation
Implement focused professional (PD) development to build learning communities and address the needs of students	State hub activity reports	Years 1-3: Low = 0-2 Medium = 3-5 High = 6 Year 4: Low = 0-2 Medium = 3 High = 4-6	At least 80% of schools have high implementation

\* See Appendix B for a fidelity matrix containing a detailed explanation of the implementation scores.

In addition to assessing fidelity of implementation for each programmatic component for the sample, the implementation data collected about each project school were also used to calculate average implementation scores for each school and state. First an annual implementation score for each programmatic component was calculated for each school and state. Second an annual average implementation score across all components was calculated for each school and state. Third, an overall average implementation score for the project was calculated for each school and state. When calculating the annual implementation scores across components, a weighting system was applied to the scores. The project leaders, through numerous discussions, made the determination that several programmatic components were more impactful to the outcomes of schools than others, therefore, the most important components were weighted in the implementation scores for each school. The two most impactful programmatic components (STW coach and focused professional development) were

given a 20% weight each and the two second most impactful programmatic components (create a powerful vision for high performance using the STW criteria and mentor school) were given a 15% weight each. The rest of the five programmatic components were weighted 6% each. The annual weighted implementation scores for each school were analyzed to assess variability of implementation across schools. The overall average weighted implementation score for each school for the project was incorporated into the analyses of project school student achievement data to test the association between level of project implementation and student achievement.

### **Limitations**

The limitations of the evaluation include: 1) Teacher survey and STW rubric respondents were not matched at the individual level over time, therefore we must consider the possibility of dependence among respondents across project years. The variance estimates may not be stable, so the results of statistical significance might shift at some degree and a further exploration would be recommended; 2) For the 2-level models that integrated student survey and achievement data, the school sample size (level 2) was dropped to 17 schools. For any multilevel model, it is effective to have more schools for statistical power; and 3) For some 2-level models, we had a limited number of covariates that could minimize the within subject variances that occurs due to some factors that are not available.

### **Impact Study Findings**

As described in the Evaluation Design section, the impact study tracks two cohorts of intervention and comparison students starting in 6<sup>th</sup> grade in either the 2010/11 or 2011/12 school year over three years by matching individual students over time. All students enrolled in intervention schools and comparison schools as 6<sup>th</sup> graders in 2010/11 or 2011/12 were part of the final analytic sample. Students who entered the schools post baseline were excluded. Additionally, students with missing 5<sup>th</sup> grade pretest achievement scores or missing 8<sup>th</sup> grade achievement scores were also excluded. Since California did not administer a statewide achievement test in 2013/14 (cohort 2, 8<sup>th</sup> grade), all California intervention and comparison students from cohort 2 were excluded. Achievement test scaled scores for English/language arts/reading and mathematics for both intervention students and comparison students were converted to z-scores. The goal of the impact study was to examine whether there was any intervention effect on student achievement.

### **Final Analytic Sample**

The final analytic sample for English/language arts/reading included 2,710 intervention students and 2,897 comparison students. The final analytic sample for math included 2,721 intervention students and 2,929 comparison students. Students were from a total of 34 schools (17 intervention and 17 comparison). The analysis for the impact study combined the two cohorts of students. Table 8 shows the 8<sup>th</sup> grade unadjusted means and standard deviations for both groups for each achievement outcome.

Table 8. i3 STW Project Impact Study Final Analytic Sample

Outcome Domain	Intervention Group			Comparison Group		
	Sample Size	Unadjusted Mean	Unadjusted Standard Deviation	Sample Size	Unadjusted Mean	Unadjusted Standard Deviation
English/ language arts/reading	2,710	-0.4386	0.9112	2,897	-0.3713	0.8791
Math	2,721	-0.2952	0.8711	2,929	-0.2484	0.9094

### Baseline Equivalence

Using students' 5<sup>th</sup> grade achievement scores (the year before the intervention) from the final analytic sample, baseline equivalence of the intervention students and the comparison students was tested. Differences in baseline means were calculated using a 2-level model (students within schools). There were no significant differences in baseline means (Table 9). In other words, the intervention students and the comparison students were equivalent on achievement test scores prior to the start of the intervention. Therefore, any differences in achievement between the groups after the intervention can be considered a result of the intervention and not because the groups began at different levels of achievement. Although the final analytic sample didn't show any baseline difference, since the standard deviation units exceeded .05, adjustments were made in the analyses at both the student and school levels.

Table 9. i3 STW Project Impact Study Baseline Equivalence

Outcome Domain	Intervention Group		Comparison Group		Model Adjusted Mean Difference	Standardized Difference	P value
	Sample Size	Unadjusted Standard Deviation	Sample Size	Unadjusted Standard Deviation			
English/ language arts/ reading	2,710	.9520	2,897	.9223	-.0834	-.089	.314
Math	2,721	.9585	2,929	.9248	-.0680	-.072	.508

### Impact Study Estimates

To examine achievement scores between intervention students and comparison students, a series of 2-level models (students within schools) were executed to assess 8<sup>th</sup> grade achievement data (in 2013 and 2014) between intervention students and comparison students. Treating students within schools examines the behavior of student's achievement outcome as a function of both students (level 1) and schools (level 2) predictors. The analysis began with two unconditional means models examining variations in English/language arts/reading and math across schools. The unconditional means models indicated a 5% clustering of English/language arts/reading scores and a 16% clustering of math scores within schools. The variance components between schools were five to twenty times higher in math and English/language arts/reading compared to within-schools variation. To evaluate the intervention

effects on English/language arts/reading and math scores, the models controlled for the following available covariates: student-level baseline achievement scores, ethnicity, and special education status; and school-level baseline achievement scores, urban/rural settings, state, percentage of free/reduced lunch students, and percentage of English language learner students. STATA xtmixed was used to execute the models. The effect size was calculated using Hedges's *g* ES index. It is defined as the standardized mean difference, which is the difference between the mean outcome of the intervention group and the mean outcome of the comparison group, divided by the pooled within-group standard deviation (SD) on 8<sup>th</sup> grade achievement scores. Since the achievement scores are at the student level, the student-level standard deviation was used.

Table 10. i3 STW Project Impact Study Results

	English/Language Arts/Reading			Math		
	Slope of Intervention	Effect Size	P value	Slope of Intervention	Effect Size	P value
<b>Combined Cohort 1 and Cohort 2</b>	-0.053	-0.059	.125	-0.129	-0.144	.110

**Notes:**

- **Slope of Intervention** is the amount of change in average achievement scores due to intervention group effect.
- **Effect Size** is a measure that describes the magnitude of the difference between the Intervention group and the comparison group.
- **P value** is for the Slope of Intervention. If the P value is  $\leq 0.05$  with a positive Slope of Intervention, we may conclude that there is a significant intervention effect on the achievement scores. If the P Value is  $> 0.05$ , we may conclude that there is no intervention effect.

The final models, adjusting for any baseline differences at the student and school levels, suggested no significant intervention effects on either English/language arts/reading scores or math scores (Table 10). In other words, the students that received the STW intervention performed the same as the comparison students after three years on both the English/language arts/reading test and the math test. For both English/language arts/reading and math, intervention schools had a negative slope (downward), but the P values are not considered to be statistically significant and the magnitude of the slopes were fairly small. The English model P value is .125 and the math model P value is .110, neither of which is statistically significant (i.e.,  $\leq 0.05$ ). Additionally, the effect sizes are small (the English model effect size is -0.059 and math effect size is -0.144), which in educational research, is generally .20 or less. Therefore, we may conclude that there was no intervention effect minimizing any known variations.

The fact that the impact analysis did not find an overall intervention effect on either English/language arts/reading scores or math scores should not be interpreted to mean that the i3 STW Project intervention was not effective, but that project schools need additional time to fully implement a complex, whole school reform model with middle grades students to advance achievement scores. We examined other hypotheses that are helpful for the interpretation of these results, including the following: 1) The impact study results are illustrative of the challenges of positively impacting student achievement in a relatively short time frame among low-performing schools, therefore we can hypothesize that it may be too soon to detect changes in achievement among intervention schools;

2) We must consider the complexities of the influence of associated factors, such as culture, climate, and instructional practices, etc. on achievement because achievement does not improve without corresponding improvements in the teaching and learning environment; and 3) The level of implementation of the intervention must be factored into the analysis of outcomes because whole school reform models such as this one are strongly influenced by both external and internal factors that affect successful implementation. The subsequent sections of this report explore these hypotheses.

## Implementation Study Findings

The i3 STW Project is based on nine key programmatic components (see i3 STW Project Components and Standards for Implementation section). The three state hubs were responsible for planning and managing the implementation of each programmatic component at the project schools in their state. This included hiring and training coaches and principal mentors for their schools, providing or arranging for professional development for their schools, managing communication with coaching staff, and ensuring that the project was being implemented with high quality by monitoring and documenting progress. All of these activities were done in consultation with the project director. Additionally, state hub leaders participated in bi-monthly conference calls throughout the project period where they had the opportunity to discuss implementation progress, challenges, and adaptations with other state leaders.

### Implementation Adaptations

As mentioned previously, the i3 STW Project was implemented in five districts in three different states. Each state had its own unique state department of education requirements and state laws, and each district had its own local policies. As a result of these different settings, adaptations to the implementation of several key programmatic components were necessary in order to accommodate the realities of each setting.

The first adaptation was to the “STW coach” programmatic component. In Illinois, STW coaches were paid for the time they spent planning and preparing for their visits to project schools. A total of seven hours per month of planning time, in addition to the time for visits, was built into the Illinois budget. In California and North Carolina, their budgets were not structured in this manner, but rather they paid coaches for school visits.

The second adaptation was to the “principal mentor” programmatic component. In Illinois and North Carolina, principal mentors were located in close enough proximity to their project principals to visit them at their schools when they worked together. In California, however, principal mentors were geographically located larger distances from their project principals so were unable to visit them every time they worked together, but rather did some of the work by phone or email. Therefore, Illinois and North Carolina principal mentor contacts were logged and counted as visits and in California, they were logged and counted as visits, phone calls, or emails.

The third adaptation to programmatic components was to the “mentor school” component. Due to the large size of several schools in California and North Carolina, rather than most of the faculty visiting their mentor school, the leadership team visited the mentor school and presented their observations and findings to the faculty upon their return. In Illinois, smaller schools allowed most faculty to visit the



mentor school. Additionally, in Illinois the close geographic proximity of mentor schools to project schools allowed a one-to-one assignment that remained in place throughout the project. In California and North Carolina, however, where mentor schools were not located close to project schools, a cadre of three to eight schools served as mentor schools, chosen for visits based on the particular improvement issue the project school was focused on at the time.

The fourth adaptation was to the “focused professional development” programmatic component. Similar to the mentor school component, again due to the large size of several schools in California and North Carolina, rather than most of the faculty attending professional development activities, the leadership team attended and presented the information to the faculty upon their return. In Illinois, smaller schools allowed most faculty to attend the professional development activities.

The last adaptation was related to additional services, in the form of instructional coaching that were provided to Illinois and North Carolina project schools. In Illinois and North Carolina, STW coaches and principal mentors provided the regular coaching services to their schools, same as California. In addition, however, Illinois project schools also had instructional coaches who worked regularly with them on literacy and math instruction. In Years 3 and 4, Illinois also had one of their STW coaches serve as a math coach for all six project schools. These additional services in Illinois were supported by funding from the Fry Foundation in Chicago. Similarly, in North Carolina, project schools also had Instructional coaches focused on technology, math, and literacy who were called in to work with project schools by the STW coaches and principal mentors as needed. This additional coaching in North Carolina was supported by funds from either the district or the state for low-performing schools.

### ***Implementation Results by Programmatic Component***

An examination of implementation fidelity across the entire sample for each of the nine key programmatic components highlighted several key findings related to variability over time and among components (Table 11). It is clear from the data that implementation in the first two years of the project was not as complete as the last two years. In Years 1 and 2, four or five of the nine components were implemented with fidelity. By Years 3 and 4, this increased to seven of the nine components being implemented fully. In interviews with the project director and state hub leaders, they commented that the work of the grant started too soon after funding came in and they should have taken more time to plan, define services, develop materials, and implement a more strategic rollout of the services to project schools. Additionally, they indicated delays in service delivery in the early years because of the learning curve of working with each district, each school, the support structures already in place at the school, and getting everyone onboard with the project.

**“We should have taken the time to plan before we got started and really talk about how we would start the implementation and get a more clarity on what we wanted to have happen.”**

i3 STW Project Director

Table 11. i3 STW Project Yearly Implementation with Fidelity by Component

Key Programmatic Components	Threshold Goal	Implemented with Fidelity for the Sample			
		Year 1	Year 2	Year 3	Year 4
		% of Sample Yes/No	% of Sample Yes/No	% of Sample Yes/No	% of Sample Yes/No
Create a powerful vision for high performance using the STW criteria	70%	71% Yes	94% Yes	88% Yes	76% Yes
Engage in an in-depth assessment and planning process using the STW criteria	80%	35% No	100% Yes	100% Yes	94% Yes
STW coach	70%	24% No	59% No	82% Yes	70% Yes
Principal mentor	70%	0% No	41% No	41% No	35% No
Mentor school	70%	70% Yes	47% No	88% Yes	88% Yes
Implement early indicators program	60%	0% No	0% No	30% No	76% Yes
Participate in national and state STW Network	70%	71% Yes	100% Yes	100% Yes	100% Yes
Implement professional learning communities (PLCs)	60%	65% Yes	76% Yes	100% Yes	76% Yes
Implement focused professional development to build learning communities and address the needs of students	80%	35% No	100% Yes	100% Yes	71% No

It is also clear from the implementation data that certain programmatic components were implemented more fully over time than others. “Creating a powerful vision,” “participate in networks,” and “implement PLCs” were all implemented with fidelity every year of the project. “Engage in assessment and planning” and “mentor school” were implemented fully for three of the four years. The components of “STW coach,” “implement early indicators,” and “focused professional development” were fully implemented for only one or two years. In the case of the “principal mentor” component, it was never fully implemented. For several of the components with lower implementation, ongoing challenges to their implementation were regularly discussed by the project director and state hub leaders, some of which were external and beyond the control of the team (i.e., project staff turnover, district mandates) and some were internal challenges (i.e., lack of defined implementation plan). The next section explicates the challenges of implementing these components.

### **Challenges to Programmatic Component Implementation**

Challenges were encountered by the i3 STW Project leaders and state hub teams that significantly impacted the implementation of three of the nine key programmatic components, as described below:

STW coach programmatic component – The primary challenge to the implementation of the STW coach component was related to the external factor of turnover in coaches. It is important to note that this issue did not impact the implementation of this component at all schools. In fact 10 of the 17 schools had the same STW coach for all four years of the project. STW coach turnover, however, did have an impact on implementation at seven project schools. Specifically, six schools had two different STW coaches during the project (three schools got a new coach in Year 2 and three got a new coach in Year 3) and one school had three different STW coaches during the project (a new coach in Year 3 and in Year 4). The impact of STW coach turnover at these schools were twofold: 1) Coaching services were at a lower frequency during the changeover; and 2) There was a delay in work because the new coach had to build relationships with the faculty and principal and establish trust before the grant work continued.

Principal mentor programmatic component – The implementation of the principal mentor component proved more difficult than anticipated due to both external and internal factors, and resulted in it never being fully implemented across the sample during any of the four project years. There were three primary barriers to providing a mentor to regularly work with project principals. First, the startup was slow because it was both challenging to choose and recruit mentors in light of the fact that their role and responsibilities were not specifically defined at the beginning of the project. Second, there was turnover in project mentor staff. Only 5 schools had the same principal mentor for all four years of the project. The third barrier was lack of access to the principal at project schools, sometimes due to the principal’s busy schedule and sometimes due to district restrictions on the principal’s time. The impact of these challenges was that principal mentors were often not able to establish the partner relationships with principals as intended and certainly not able to spend the amount of time with principals that was originally intended.

“We’ve lost coaches due to illness or personal situations, and then we have to find new coaches. This type of change is hard because they have to catch up and they have to re-convince people they’re working with to stay with it, and that’s sometimes difficult.”

i3 STW State Hub Leader

Implement early indicators programmatic component – The early indicators component, coordinated by FHI 360 who had a subcontract from the Forum to conduct the work, was considered a significant challenge by the project director, state hub leaders, and project schools throughout the project. The state hubs and project schools expressed frustration with the timing and content of the trainings, the access to materials, and the need for more assistance with identifying and designing tiered intervention strategies. The project director modified the plan for this component in Year 3 to better reflect project needs, so that implementation could continue. The impact of these challenges, however, resulted in significant delays in implementation at project schools, with at least 4 schools never fully implementing this component while others who achieved full implementation did not do so until the final year of the project.

### **Implementation Results by School**

An examination of implementation results by school and by state highlight additional key findings related to the variability of implementation. Table 12 shows the annual weighted average

implementation scores for each school and state, as well as across years. It is clear from the overall weighted average across years that implementation was fairly high at project schools, with a score of 74.1% overall. Implementation by school was lowest in Year 1 of the project, it increased in both Years 2 and 3, and experienced a drop in Year 4. The largest drop in implementation in Year 4 occurred in North Carolina schools. Among the three states, Illinois had higher overall implementation than California or North Carolina. Additionally, Illinois had the least variability in implementation among their schools. We hypothesize that implementation in Illinois schools may be higher because the Illinois schools were smaller in size and therefore implementation across the school was more manageable than in larger schools. Although overall implementation was moderately high, variability across schools and states led us to examine the impact of implementation on other data collected as part of the project.

Table 12. i3 STW Project Average Yearly and Overall Weighted Implementation by School and State

States	Schools	Overall Weighted Implementation Score				
		Year 1	Year 2	Year 3	Year 4	Overall Average
CA	School 1	37.8%	65.7%	76.4%	65.0%	61.2%
	School 2	69.5%	90.0%	86.1%	94.0%	84.9%
	School 3	59.0%	73.2%	83.7%	65.0%	70.2%
	School 4	69.5%	87.2%	88.9%	94.0%	84.9%
	School 5	61.8%	65.7%	67.0%	57.5%	63.0%
	School 6	59.0%	76.0%	86.5%	65.0%	71.6%
	<b>CA Average</b>	<b>59.5%</b>	<b>76.3%</b>	<b>81.4%</b>	<b>73.4%</b>	<b>72.7%</b>
IL	School 7	73.3%	76.4%	96.4%	92.4%	84.6%
	School 8	65.8%	68.1%	94.2%	84.9%	78.2%
	School 9	65.8%	76.4%	94.2%	92.4%	82.2%
	School 10	70.5%	70.9%	88.9%	82.1%	78.1%
	School 11	73.3%	69.0%	91.9%	87.9%	80.5%
	School 12	68.6%	75.4%	94.2%	89.6%	81.9%
	<b>IL Average</b>	<b>69.6%</b>	<b>75.4%</b>	<b>94.2%</b>	<b>89.6%</b>	<b>81.9%</b>
NC	School 13	40.6%	71.2%	100.0%	72.8%	71.2%
	School 14	51.4%	79.4%	77.5%	65.2%	68.4%
	School 15	30.0%	71.4%	69.6%	69.4%	60.1%
	School 16	49.0%	77.8%	97.8%	80.2%	76.2%
	School 17	43.9%	79.4%	85.0%	40.5%	62.2%
	<b>NC Average</b>	<b>43.0%</b>	<b>75.8%</b>	<b>86.0%</b>	<b>65.6%</b>	<b>67.6%</b>
<b>Overall Average</b>		<b>58.2%</b>	<b>74.9%</b>	<b>87.0%</b>	<b>76.4%</b>	<b>74.1%</b>

### **Impact of Level of Implementation**

Due to the variability evident in implementation levels across years, states, and schools, implementation results were analyzed in combination with other process and outcome data collected at

project schools with compelling results. Using the overall average weighted implementation score across years, project schools were categorized as higher implementers (74% or above) or lower implementers (below 74%). Project schools with higher implementation had statistically higher ratings on the STW Rubric (Figure 3), statistically higher collaboration opportunities and outcomes (Figure 4), and statistically higher “best” middle-grades instructional practices (Figure 5).

Figure 3. STW Rubric Results by Level of Overall Project Implementation

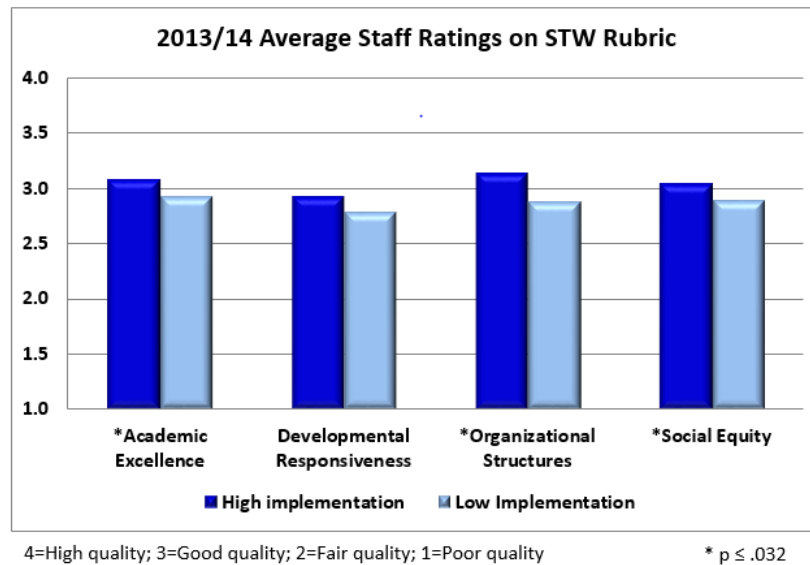
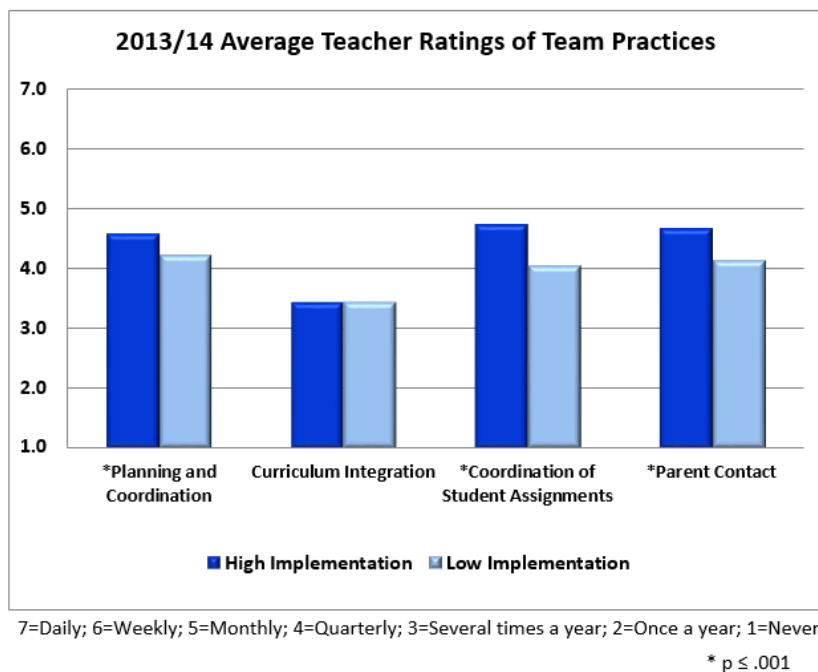


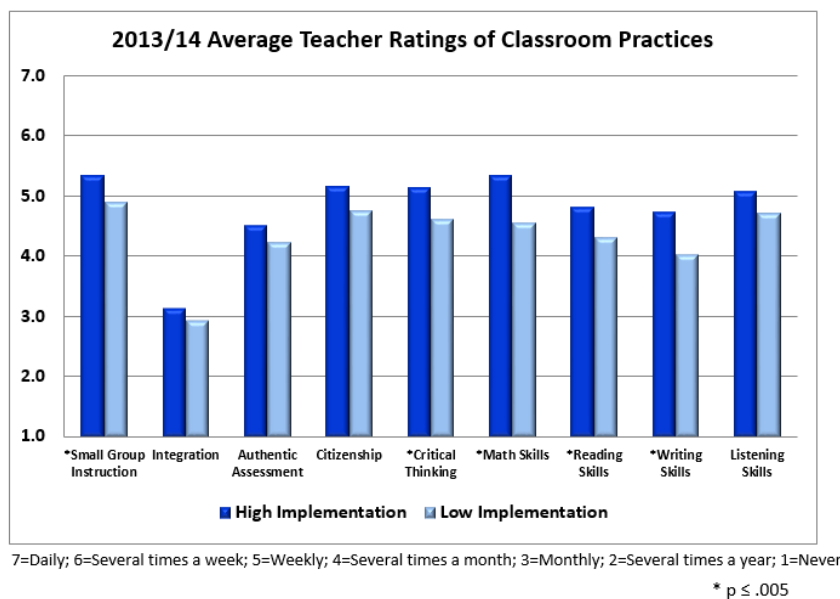
Figure 4. Teacher Ratings of Team Practices by Level of Overall Project Implementation



These findings demonstrate that project implementation is an important influence on the level of programs and practices at project schools. Higher levels of successful contexts were reported among the

project schools that implemented the project components with higher levels of implementation. Schools implementing the i3 STW Project as well as other reform models can see the importance of adhering to high levels of implementation. For project developers and implementers, it suggests that a higher level of monitoring of implementation may be required to ensure high implementation. Additionally, well-defined guidelines for implementation that are supportive of consistent implementation may also assist with tighter implementation across schools. Although these results do not extend to achievement (i.e., there were no statistically significant differences in achievement between high implemented schools versus lower implemented schools), it is encouraging that higher implemented schools have higher levels of the practices that have been shown to be associated with student achievement in prior research (Allen, Pianta, Gregory, Mikami, & Lun, 2011; Mertens & Flowers, 2003).

Figure 5. Teacher Ratings of Classroom Instructional Practices by Level of Overall Project Implementation



## Findings for Immediate and Intermediate Outcomes

As described above, the evaluation results show positive results by level of implementation of school and teacher practices. To delve deeper into more proximal outcomes, the evaluation team conducted exploratory analyses of the evaluation data in order to investigate and understand the immediate and intermediate outcomes of the project as well as the contextual changes occurring at project schools. At the student and teacher levels, data were collected to examine changes in attitudes and behavior, changes in instructional practices, shared decision making, and changes in school climate and culture. At the school level, data were collected to measure the implementation of the criteria associated with the four components of the STW Program. At the project level, data were collected to track project coaching and professional development services as well as observations of improvements at project schools. These exploratory analyses are helpful in understanding the building blocks needed in order for outcomes, such as student achievement progress to occur.

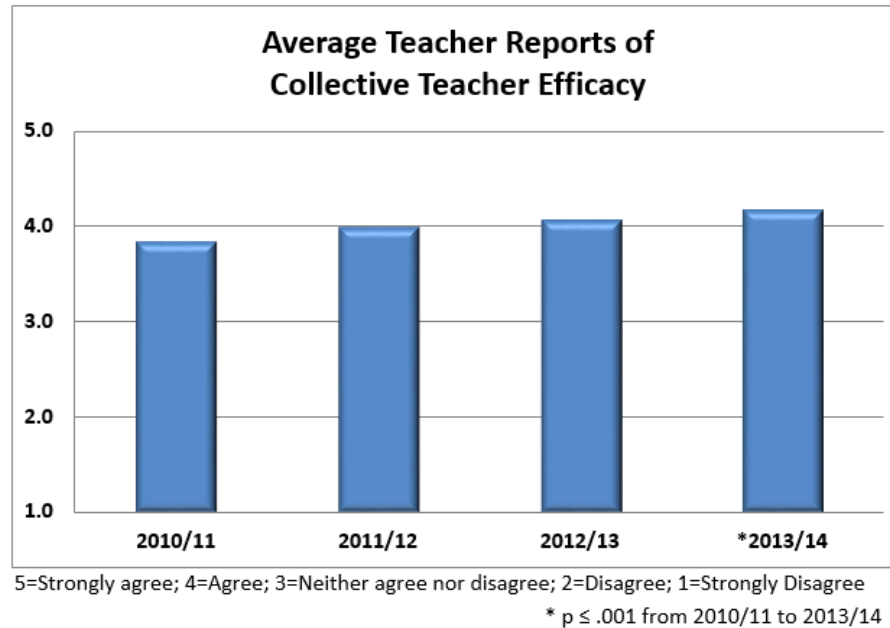
### ***School Improvement Self-Study Teacher Survey Findings***

The School Improvement Self-Study Teacher Survey was administered annually at all project schools and school-specific, state-specific, and national reports on the results were disseminated. At an aggregate level, several analytic methods, including Univariate Analysis of Variance, 2-level modeling (teachers within schools), and Multivariate Analysis of Variance techniques were used to examine the data. Overall, teachers reported statistically significant improvements in numerous constructs related to the focus of the project such as climate, collaboration, leadership, and classroom instructional practices.

Teachers reported statistically significant ( $p \leq .021$ ) improvements in the culture and climate (i.e., work climate, classroom climate, teacher decision making, collective teacher efficacy) of their school from the first year of the project to the last year of the project:

- **Work Climate Improved** – Teachers reported higher levels of commitment to their school, including pushing themselves to do their best work, taking pride in the school, and agreeing that there is group spirit among the staff. Teachers also reported stronger feelings of being recognized for their contributions by the school administration. For example, teachers more often felt the administration recognized staff who did well, stood up for staff, and encouraged staff to make recommendations and suggestions. Teacher ratings of both commitment to their school and their feelings of being recognized improved in Year 2, Year 3, and Year 4.
- **Classroom Climate Improved** – Teachers reported improved climate in their classrooms in Year 3 and Year 4. Teachers felt that students were working more productively and respectfully with each other by Year 3 and Year 4 (e.g., enjoy working together, respect each other's viewpoints). Teachers also reported that student behavior in class was less disruptive (i.e., less restless, less likely to call out answers, and fewer instances of students arguing with each other). Finally, teachers also reported more frequent positive interactions with students, such as students sharing their concerns and asking for support when needed.
- **Teacher Decision Making Increased** – Teachers reported an increase in opportunities to participate in decisions in Year 3 and Year 4 of the project. Additionally, teachers reported more freedom to autonomously make decisions by Year 4. Both of these areas are correlated to work climate, which according to teachers, improved prior to improvements in participation and autonomy in making decisions.
- **Collective Teacher Efficacy Improved** – Teachers reported an improved collective commitment to the success of their students each year of the project in Year 2, Year 3, and Year 4 (Figure 6). In other words teachers reported stronger beliefs that teachers at their school have what it takes to get the children to learn, are able to get through to difficult students, are confident they will be able to motivate their students, and believe that if a child doesn't learn something the first time, teachers will try another way.

Figure 6. School Improvement Self-Study Teacher Survey: Collective Teacher Efficacy by Year



Teacher ratings of the leadership practices of the administration at their school increased significantly ( $p \leq .001$ ) in Year 3 and in Year 4. Teachers reported that the administrators in their school were doing better in areas such as more frequent communication with the faculty, follow-through on decisions and next steps, problem-solving strategies, handling conflict in the building, using a proactive approach to school improvement and bettering the school climate, and an increase in collaborative leadership opportunities offered to faculty.

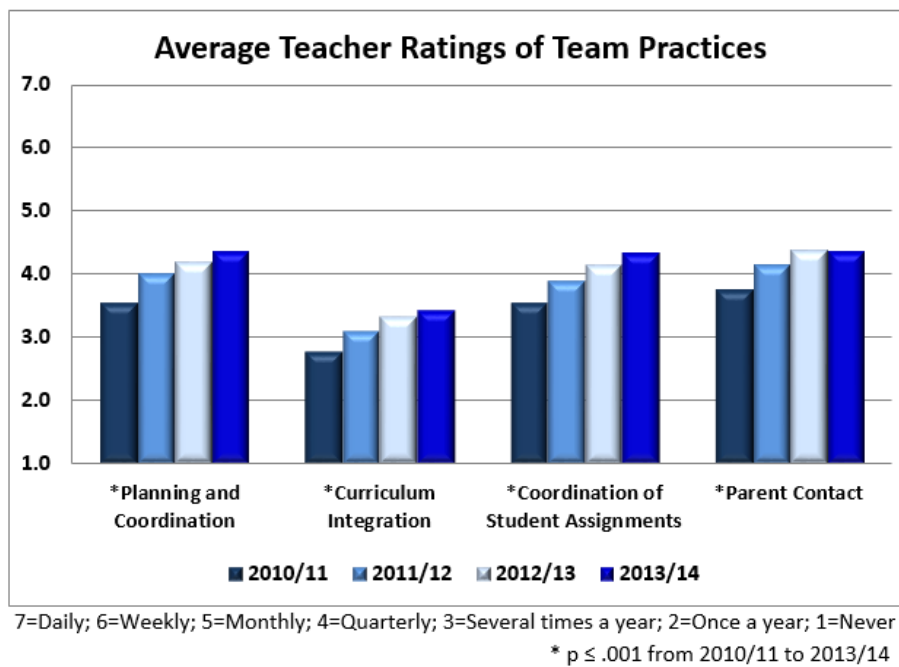
Teachers reported statistically significant ( $p \leq .001$ ) improvements in collaboration opportunities and outcomes (i.e., team practices, team decision making, quality of team interactions) from the first year of the project to the last year of the project:

- Team Practices Increased** – Teachers reported more frequent collaborative communication, practices, and coordination with members of their interdisciplinary teams or grade level teachers in both Year 3 and Year 4 of the project (Figure 7). First, teachers reported engaging more often in planning and coordination activities with their fellow teachers, such as setting goals and objectives related to student learning, planning special team projects and activities, and using block or flexible schedules to achieve instructional goals. Teachers also reported a higher level of coordination of student assignments and assessments with each other, including coordinating homework across subjects, coordinating test schedules across subjects, matching instruction to student learning styles, and discussing student needs. Curriculum integration practices also increased during Year 3 and Year 4 of the project with teachers reporting they engaged more often in working together to co-present lessons, teach interdisciplinary units, and coordinate curricula across subjects. Finally, contact with parents increased in Year 3 and held that level in Year 4 with teachers implementing strategies to increase parent involvement and informing parents about the progress of their child. Planning and coordination activities, coordinating assistance for students, and parent contact increased



from 2 to 3 times a year in Year 1 to between quarterly and monthly in Year 4. Curriculum integration increased from once a year in Year 1 to occurring 2 to 3 times a year in Year 4.

Figure 7. School Improvement Self-Study Teacher Survey: Interdisciplinary Team Practices by Year

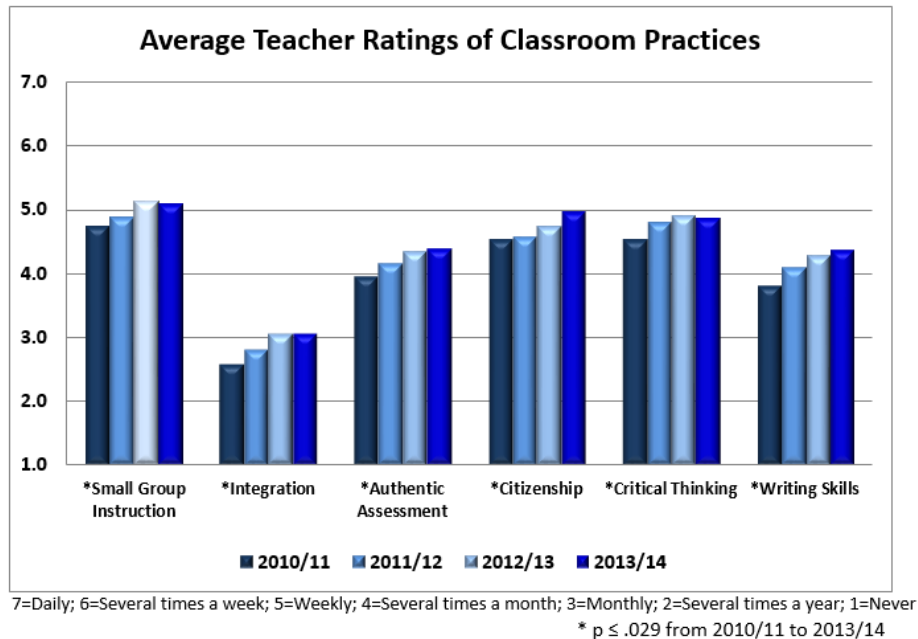


- Team Decision Making Increased – Teachers reported increased decision making opportunities among their interdisciplinary teams or grade levels in Years 2, 3, and 4 regarding team practices (i.e., kind and amount of homework students receive), school-wide policies and practices (i.e., discipline policies), and student performance and assessments (i.e., evaluation of student progress). Each of these three constructs went from occurring a “moderate amount” in Year 1 to occurring “much” in Year 4. In other words, as the project went on, teachers had more decision-making authority and opportunities around issues that were directly related to the students they taught (i.e., students on their teams or in their grade).
- Quality of Team Interactions Improved – One of the constructs on the teacher survey measures the quality of interactions between teachers when they work collaboratively together. Teachers reported, that over time, cohesion and harmony among teachers working together improved (i.e., work well together, respect each other ideas), teachers felt more prepared to work together in a collaborative way, and teachers agreed that they not only addressed student needs (i.e., get to know students as individuals) but had consistently high expectations for students (i.e., communicate high expectations for achievement to all students). Increases could be seen by Year 3 with teachers indicating full consensus that these four areas functioned better by Year 4 of the project.

Teachers reported statistically significant ( $p \leq .029$ ) increases in the use of “best” middle-grades instructional practices. Teachers report engaging more frequently in the following six instructional practices among the nine measured from Year 1 to Year 4 (Figure 8):

- Small group active instruction (weekly by Year 4)
- Integration and interdisciplinary practices (monthly by Year 4)
- Authentic instruction and assessment (several times a month by Year 4)
- Citizenship and social competence instruction (weekly by Year 4)
- Critical thinking practices (approaching weekly by Year 4)
- Writing skill practices (several times a month by Year 4)

Figure 8. School Improvement Self-Study Teacher Survey: Classroom Instructional Practices by Year



A deeper analysis of the improvement in classroom instructional practices showed that Illinois schools contributed the most to the changes in instructional practices over time. This may have been due to the fact that Illinois schools had the highest level of coaching by instructional coaches, suggesting that instructional coaching was impactful. Another interesting observation about instructional practices is that while some of the other constructs measured with the teacher survey had improvements by Year 2 of the project, the improvements in classroom instructional practices took longer to manifest. In many ways, this makes logical sense because changing instruction requires prior planning and collaboration work before being tested and perfected in the classroom. For example, a climate of collaboration and trust must exist in order for teachers to work together on improving lesson plans and instructional strategies. Additionally, teachers must have adequate time to work together as well as the autonomy to make decisions about their instruction. In many ways, instructional practices are the final ingredient in the reform process. And since changes in instructional practices didn't occur until Year 3 or Year 4, students may not have experienced them long enough to positively impact their achievement.

### **School Improvement Self-Study Student Survey Findings**

The School Improvement Self-Study Student Survey, like the teacher survey, was administered and reported annually at all project schools. The same analytic methods used to examine the aggregate teacher survey data were also used for the student survey data, including Univariate Analysis of Variance, 2-level modeling (students within schools), and Multivariate Analysis of Variance techniques.

Overall, there was no significant growth over time on the constructs measured with the student survey among project students. In other words, student overall reports of academic efficacy, academic expectations, belonging, family involvement, and school climate did not change significantly across the entire sample during the project period. There were several individual schools that showed improvement in some constructs over time, but a consistent pattern was not apparent. As with classroom instructional practices, we can hypothesize that student experiences may be one of the final ingredients in the reform process as well, and thus provides an explanation for lack of change over time. We may also hypothesize that some of the student survey constructs, student expectations and aspirations, are more deeply ingrained values and harder to impact.

Since improvements in student constructs over time were not apparent, further analyses were conducted to explore what factors from the student survey constructs (i.e., academic efficacy, academic expectations, etc.) are related to student achievement. This information would be helpful to the project implementers and schools because it would assist them in understanding the attitudes and experiences of students that are most likely to support their learning and achievement growth. It is important to note that while merging the student survey data with achievement data, a considerable portion of the sample for cohort 2 were excluded because of unavailable California achievement data. The sample contained 1,529 students with both achievement data and survey data from cohorts 1 and 2 (a reduction from approximately 2,700 students in the final analytic sample for the impact study).

The results of 2-level modeling showed that academic efficacy (i.e., a student's belief in their own ability to be academically successful) had a positive impact on math achievement. Although this finding did not hold for English/language arts/reading achievement, it is important for math teachers to recognize that strengthening students' attitudes and mindsets about their abilities can positively impact their achievement. Students with higher self-ratings of academic efficacy tended to have higher math achievement scores. Correlational analyses also support this finding in that there was a positive correlation (i.e., association) between academic efficacy and math achievement. Additionally, there was also a positive correlation between academic efficacy and English/language arts/reading achievement.

Several other interesting findings emerged when examining factors related to student achievement. First, results of 2-level modeling showed that student reports of their family's involvement in their education (i.e., a student's family talks to them about the importance of school and encourages them to do well at school) had a negative effect on their English/language arts/reading achievement and math achievement. In other words, students who reported higher levels of family involvement in their education tended to have lower achievement scores. One hypothesis to explain this finding is that the parents of students who are not doing well in school tend to become more involved in order to encourage and support their child to do better, thus family involvement is higher for those students. A similar finding was also evident for academic expectations (i.e., a student's short and long-term academic expectations for themselves, as well as their perceptions of the expectations that their parents and teachers have for them). Students who reported higher academic expectations also tended to have lower English/language arts/reading and math achievement. Researchers studying the aspiration attainment gap among high school students found similar results in that students' aspirations for academic attainment may be high, but it does not necessarily mean that their goals match their current achievement. Particularly among high need students, a gap exists between aspirations and actual

attainment (Roderick, 2006). Finally, correlational analyses showed a statistically significant association between English/language arts/reading achievement and math achievement. Students with higher English scores tended to have higher math scores, affirming the findings of other studies regarding the predictive power of English proficiency on math scores (Henry, Nistor, & Baltes, 2014; Larwin, 2010).

### ***Student Subgroup Analysis***

To examine whether the i3 STW Project reduced the achievement gap between subgroups of students, both ELL students and special education students, additional models were run for English/language arts/reading and math. The models were run separately on the samples of ELL students, special education students, and all other students, and included confidence intervals for all effects so that overlap between the samples could be examined. The results showed that there was overlap in the confidence intervals suggesting that the impact on achievement was not larger in the subgroups than in the main group. Therefore, there is no evidence that the project reduced the achievement gap between the groups. Project leaders can use this information to assess and review the specific programmatic focus and activities directed to these student subgroups for refinement and potential strengthening.

### ***STW Rubric Findings***

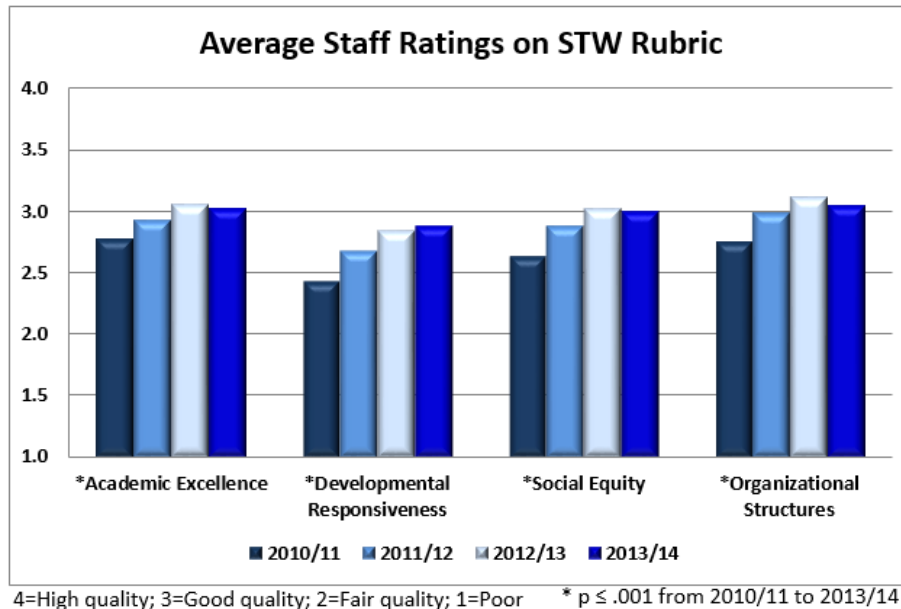
In addition to reporting the annual school-level, state-level, and national results of the STW Rubric, analyses to assess change in aggregate teacher reports of the implementation of the four key components of the STW criteria across years was conducted. 2-level modeling was used to examine the nested data (teachers within schools). The analysis showed that there was significant improvement in teacher ratings of their implementation of each of the four key STW Rubric criteria (Figure 9). Overall, staff reported significantly higher implementation of the STW criteria each year ( $p \leq .001$ ) with a relatively smaller change from Year 2 to Year 3. Improvement was measured among all four criteria from Year 1 to Year 2, from Year 1 to Year 3, from Year 1 to Year 4, and from Year 2 to Year 3. The level of implementation of the STW Rubric components was maintained from Year 3 to Year 4.

Examining each STW component individually, we see that for academic excellence, there is overall improvement in Year 2 and Year 3. While there was no improvement in Year 4, the level of implementation was maintained. Developmental responsiveness had overall improvement in Year 2 and Year 3. The change from Year 3 to Year 4 was negligible. For social equity, there was overall improvement in Year 2 and Year 3, and like academic excellence, the level was maintained in Year 4. Organizational structures had gains in Year 2 and Year 3 as well, with a slight decline in Year 4. Overall, the largest improvements were from Year 1 to Year 2, followed by Year 2 to Year 3. While Year 4 did not show continued growth across the components, the level of implementation of the STW Rubric components was maintained in Year 4.

When state was incorporated as a fixed effect in the statistical model, an interaction between state and yearly change became significant as well ( $p < .001$ ) for all four STW Rubric areas. These interactions indicated that the changes over time in academic excellence, developmental responsiveness, social equity, and organizational structures, varied by state. From the implementation study results, we know that the implementation of the project was somewhat different among states but more similar within states. Therefore, the analysis of the STW Rubric data was also performed separately for each state in

order to understand the longitudinal changes. For all four key STW Rubric components, North Carolina contributed the most ( $p < .001$ ) in all project years followed by Illinois ( $p < .001$ ). California contributed in all key STW Rubric components except for academic excellence ( $p = .147$ ). It should be noted that the California project schools had the largest student enrollments and the largest number of teachers, thus suggesting that it may take longer to implement both the project and the STW criteria in larger schools.

Figure 9. STW Rubric Results by Year



The overall growth in the four key components of the STW criteria across project years is an important accomplishment and indicates that project schools improved their programs and practices during the project period. Since the STW criteria is the framework of the i3 STW Project, it is compelling to see that project schools showed significant growth in these areas over time. Schools that use the STW criteria to apply to be designated as a STW by the Forum are expected to be consistently averaging scores between 3.4 and 4.0 (out of 4) in all sections and components of the criteria. Although the project schools are not quite at that level of implementation, they have shown marked improvement since the beginning of the project.

### **STW Designated Project Schools**

As described previously and mentioned in the prior section, the framework for the project is the Forum's STW criteria. At a national level, the Forum uses the STW criteria for identifying middle-grades schools across the country that are on an upward trajectory of growth and continuous improvement. Schools that are interested in being designated as a STW by the Forum engage in an application process with the Forum's affiliate in their state that is comprised of several steps, including: complete an online application form; submit data on student practices and performance; complete the online STW rubric; and host a visit to their school by a STW review team that includes interviews with the administrators, teachers, students, and parents; observations of classrooms; and team meetings. Typically, each of the 18 STW state affiliates designates 3 to 5 schools per year as STW schools.

Although being designated as a STW was not articulated as an outcome for project schools by the end of the grant, it was understood that the project would lay the foundation for schools to be designated in the future. It turned out, however, that three project schools (Hamlet Middle School in North Carolina, Nathan S. Davis Elementary School in Illinois, and Eli Whitney Elementary School in Illinois) made substantial improvements in practices and outcomes during the grant period and so applied and were designated as a STW school before the end of the grant. Hamlet Middle School was designated in spring 2013, Nathan S. Davis Elementary School in spring 2014, and Eli Whitney Elementary School in spring 2015. They join more than 300 other schools across the county that are STW schools and they serve as success stories for the i3 STW Project.

In an effort to understand how these three project schools succeeded in becoming STW schools during the grant period, project stakeholders (state hub leaders and the project director) were asked in interviews at the end of the project to reflect on why these three schools were able to accomplish this achievement. Three themes emerged as reasons for the rapid turnaround and success of these project schools, including readiness for improvement, principal leadership, and strong coaching services support.

Stakeholders identified a level of readiness among the principals and faculties of these three schools that was unique among all project schools. All three schools already had a mindset of improvement when they joined the grant, they gained quick teacher buy-in for the project, and they had teachers who were willing to become leaders. One stakeholder said “This was a school that was ready for the next thing. They just didn’t know what their next thing was to go to the next level.” The second theme that was described as an ingredient for the success of these three project schools was the importance of a strong principal leader. Stakeholders described the principals at the three schools as being not only the “curriculum leader of the school,” but also the “model for the teachers.” Additionally, these principals trusted their teachers, providing them with the freedom to try things their way, and providing them with the necessary supports to be successful. The principals also strongly believed in the STW criteria and as a result, they promoted it and advocated for the STW approach with their faculty. Finally, stakeholders pointed to the strong coaching services team (STW coach, principal mentor, instructional coach) that was paired with each of these schools, saying “I think it was the right combination of coaches and their abilities.” The coaching team members “piggybacked” support when they saw a need and thus strengthened the impact.

**“The principals had a really strong, passionate drive to continue to raise their school’s profile.”**

i3 State Hub Leader

In addition to examining stakeholder interviews, process and outcome data were analyzed to explore whether these three project schools who were designated as STW during the grant were performing at a higher level than the other project schools. The results of the analyses were very positive in that they showed that the three project schools designated as STW had statistically ( $p \leq .001$ ) higher ratings on the STW Rubric (Figure 10), statistically higher collaboration opportunities and outcomes (Figure 11), and statistically higher “best” middle-grades instructional practices (Figure 12). These three schools were successful in making a more rapid implementation of the STW criteria as well

as programs and practices facilitated and supported by the coaching services of the grant than other project schools.

Figure 10. STW Rubric Results by STW Designated Project Schools

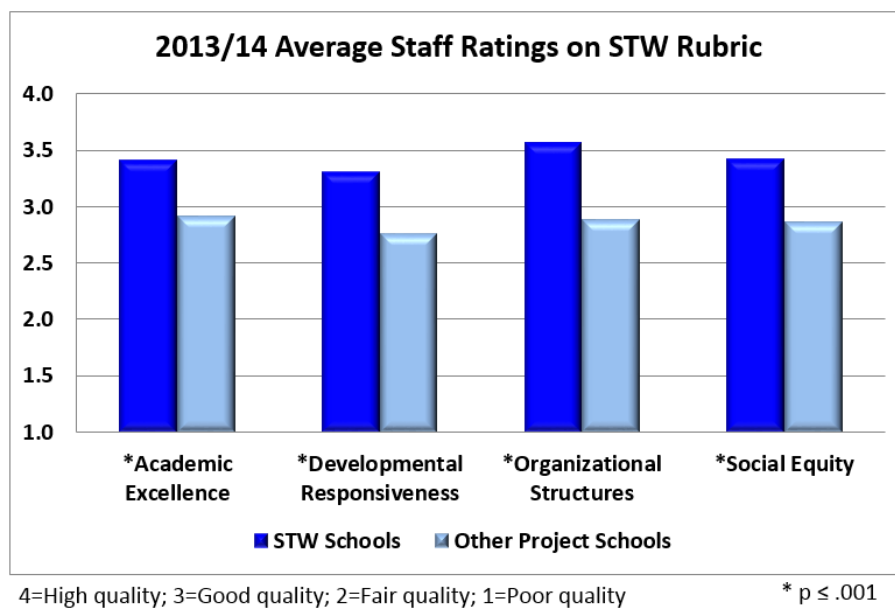


Figure 11. Teacher Ratings of Team Practices by STW Designated Project Schools

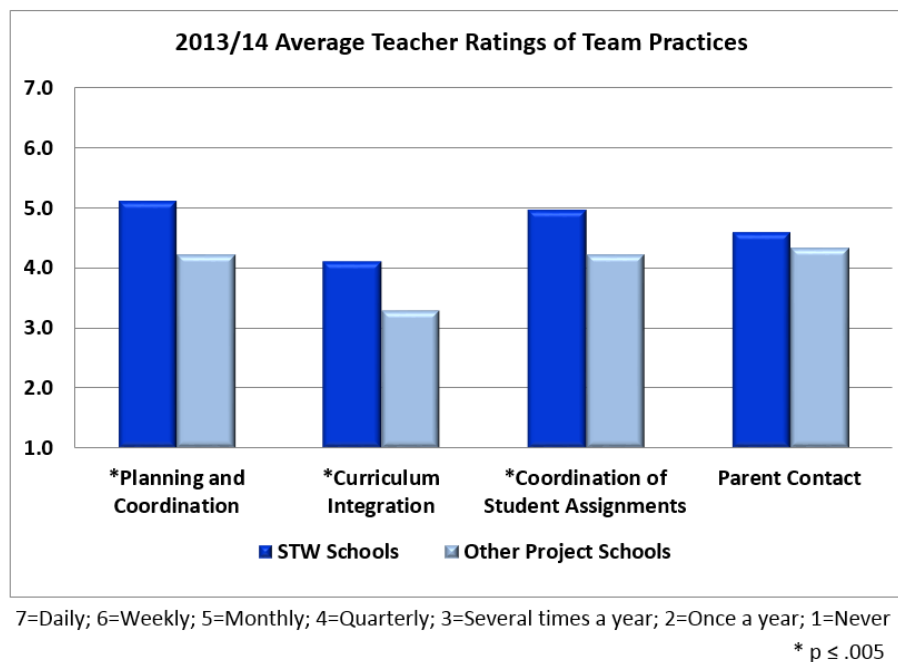


Figure 12. Teacher Ratings of Classroom Instructional Practices by STW Designated Project Schools

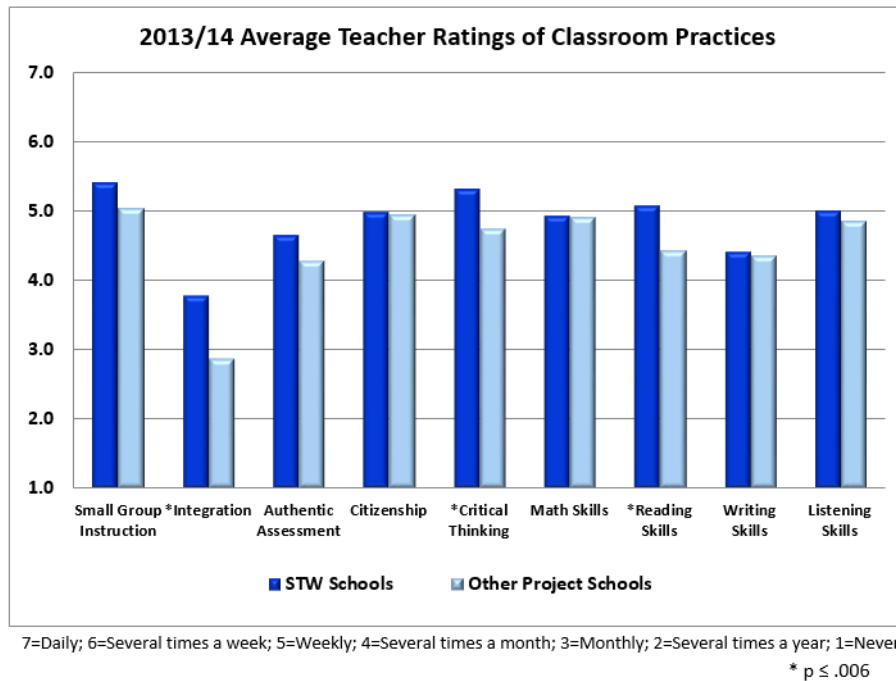
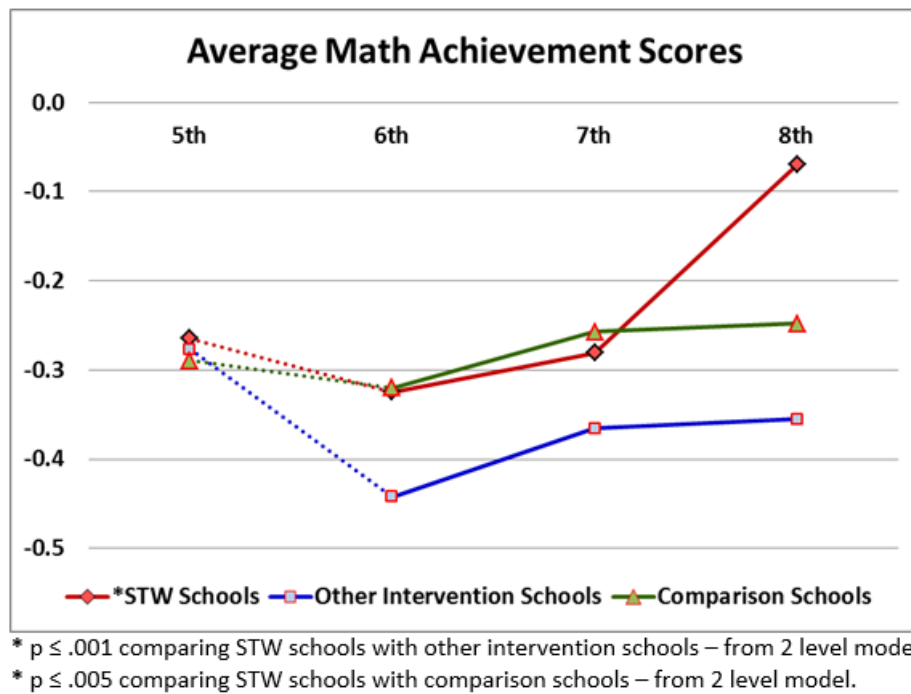


Figure 13. Math Achievement Scores by STW Designated Schools



The achievement scores of students attending the three project schools that were designated as STW during the project were also examined and compared with the achievement of students at other project schools as well as at comparison schools with compelling results for math achievement (Figure 13). When analyzed with 2-level models, excluding comparison schools, the math scores of STW schools



turned out to be statistically significantly higher than other project schools (positive slope=.536 and  $p \leq .001$ ). Additionally, STW schools also performed significantly higher in math scores than the comparison schools (positive slope=.354 and  $p \leq .005$ ). There was no statistical difference, however, in English/language arts/reading achievement scores between the groups. When hypothesizing why math was higher but not English/language arts/reading, it may be because the three STW schools focused more strongly on math. This finding, however, further adds to the evidence that these three schools were successful in not only making more rapid implementation of the STW criteria and programs and practices, but also positively impacted their math achievement. A potential hypothesis is that although the remaining project schools are on the pathway to these same improvements, because they started the project at a slightly higher level of readiness, the remaining schools may need more time investment than these three schools.

### **Evolution of Coaching Services**

Coaching services were provided to project schools by a combination of STW coaches, principal mentors, and instructional coaches. Their work was logged and tracked in both the coach's log and via annual focus groups. These data were reported annually to schools, coaches, mentors, state hub teams, and the project leadership team to track and assess the work. In California, STW coaches and principal mentors provided regular coaching services. In Illinois and North Carolina, in addition to STW coaches and principal mentors, instructional coaches were also used. Although instructional coaches were not part of the original programmatic plan for the project, it emerged as a need among project schools. Instructional coaches in Illinois worked regularly with schools and focused on literacy and math instructional practices. In North Carolina, instructional coaches focused on technology, mathematics, and literacy and were called in to work with project schools by the STW coaches and principal mentors as needed. Two schools in California also had instructional coaches in Year 3 and early in Year 4 via district support, but their activities were not recorded in the coach's log.

Table 13. i3 STW Project Coaching Services Delivered by Year

Annual Contacts and Activities	Coaching Services*				
	Year 1	Year 2	Year 3	Year 4	Total
Number of schools served	16	17	17	17	17
Number of contacts	179	660	739	616	2,194
Number of activities	398	1,372	1,585	1,382	4,737
Average contacts per school	11	39	43	36	129

\* Coaching services includes STW coaches, principal mentors, and instructional coaches.

Overall, coaching services were a very intensive and ongoing aspect of the intervention service delivery for all i3 STW Project schools. Throughout the four-year project, STW coaches, principal mentors, and instructional coaches made over 2,000 contacts with project schools consisting of close to

5,000 different activities (Table 13). This amounted to an average of 129 contacts per project school. The majority of contacts and activities (over 50%) at all schools were delivered by STW coaches. Instructional coaches provided a great deal of contacts as well in Illinois and North Carolina project schools (approximately 25% in Illinois and approximately 15% in North Carolina). Principal mentor contacts were highest in North Carolina (28%), followed by California (17%), and Illinois (14%).

Principal mentors worked specifically with the project principals and supported them in leadership development and overall reform and improvement. Instructional coaches worked with individual teachers to improve lessons, instruction, and classroom management in literacy, math, or technology. The STW coach, however, had a much broader scope of work, which was guided by the STW vision, but often driven by the specific needs of the school. It included vision and goals, curriculum and instruction, culture and climate, collaboration, and more. Over time, STW coaches defined their different coaching roles with their schools in the following three ways: 1) Reform coaching – working on global cultural and environment, and building collaborative leadership; 2) Instructional coaching – working one-to-one with teachers on instructional practices, and providing small group and whole school professional development and resources; and 3) Responsive coaching – mentoring and responding to the changing needs of the principal, individual teachers, and teams of teachers.

The focus of coaching services evolved over the four years of the project as the schools progressed and their needs changed. Although coach's log data in Year 1 was limited for a third of the schools due to low completion rate of the log, we were able to augment it with focus group data, and report that the key coaching activities during Year 1 that were identified as prerequisites to future work addressing specific improvement needs included:

- 1) Building trusting relationships with the principal, leadership team, and teachers;
- 2) Learning about the school by observing, asking questions, and listening; and
- 3) Establishing a culture and climate of collaboration at the school.

**“Our STW coach has provided for us a real strong vision and an extreme amount of knowledge so that she’s able to make the connections for us.”**

i3 STW Project Principal

In Year 2, there was turnover among coaching staff at several schools, thus necessitating revisiting some of the Year 1 activities for those coaches. Overall, however, an examination of coaching activities during Year 2 showed growth in coaching content and included the following key activities:

- 1) Using data (e.g., STW Rubric, Self-Study Surveys, benchmarks) to identify the specific needs of each school and working with them to establish an action plan;
- 2) Providing resources and facilitating professional development sessions on new programs or strategies to address needs (e.g., common core, advisory programs, technology); and
- 3) Working with small groups of teachers (PLCs, teams, grades) or individual teachers to observe classroom instruction and model lessons in order to improve the culture of the classroom, innovate lessons, and implement effective instructional practices.

Coaching activities continued to evolve in Year 3 and were clearly tailored to the specific needs of each school, and driven by their school improvement plans and action plans. Commonalities in coaching in Year 3 included:

- 1) Continuing to improve and augment collaboration as well as leadership opportunities for teachers by strengthening PLCs, teams, small groups, and leadership teams (e.g., common understanding of roles and responsibilities, establishing protocols for implementation of action plan, collaboration norms);
- 2) Providing a myriad of resources to the administration, teachers, and staff in a variety of areas including scheduling, classroom management, instructional strategies, subject-specific content, best practices, common core, and ELL;
- 3) Improving classroom instruction particularly in the areas of literacy, language arts, mathematics, the use of technology in the classroom; and increasing the overall rigor of academics; and
- 4) Addressing individual student needs which for some schools was addressing student behavior, attendance, classroom management, discipline practices and for others was analyzing data and early intervention information (student grades, behavior, and absences) and working to design interventions for individual students.

*“Our coaches have stressed to us the idea of having more people involved in leadership roles.”*

i3 STW Project Principal

---

### ***Key Ingredients of Successful Coaching in the i3 STW Project:***

*Individualize the coaching approach for each school.*

*Establish one-to-one relationships with the administration and teachers.*

*Network and share with other coaches.*

*Remain focused on the central STW vision and goals.*

---

In Year 4, overall coaching continued to be tailored to each school’s individual needs, school improvement plan, and the district’s direction. For example, some schools were focused on implementing an advisory program, others on academic improvement, still others on improving instruction. Some schools were working on student engagement while others were preparing their application to be designated as a STW. Others were working on school environment. Commonalities in coaching content for Year 4 included:

- 1) Refining classroom instruction with more rigor, integrated units, the use of technology, and for subgroups of students (ELL, special education). Coaches worked with the leadership teams and small groups of teachers or PLCs to plan and set goals. Some coaches also worked with individual teachers to model lessons, provide instructional resources, observe instruction, and provide feedback;
- 2) Addressing student behavior, attendance, classroom management, and discipline practices (e.g., student interventions, advisory programs), which for some meant refining the analysis of early intervention data (student grades, behavior, and absences) and designing interventions for small groups of students and individual students who are struggling; and
- 3) Improving communication and collaboration with district representatives in order to support a holistic and integrated school improvement plan.

*“What’s been great about the coaches that I’ve had is it’s been a side-by-side coaching relationship, and that was very important. Very important not to have someone come into my school and just say, ‘You should do this.’ And it hasn’t been that at all.”*

i3 STW Project Principal

STW coaches, principal mentors, and instructional coaches faced a variety of challenges throughout the project as they worked with their schools. Some challenges were isolated to a particular state, district, or school. For example: a 7-day teacher strike in September 2012 in Chicago Public Schools; the closure of a project school in Durham Public Schools after Year 2 of the project; and two Illinois project schools facing potential closure along with nearly 50 other schools in the district. Other challenges were common for most of the coaching staff. The typical challenges encountered throughout the project included:

- 1) Principal turnover – Nine of the 17 project schools had at least one principal change during the project period; several had two principal changes. Thus, for coaches, orienting and working with the new principals meant returning to an introduction of the STW criteria and the i3 grant, establishing trusting relationships between teachers and the new principal, and ensuring that the leadership teams were functioning collaboratively. It also necessitated a well-developed plan of communication and orientation. It meant starting the project over again in many ways, however, it did not take as long to get up to speed due to other teachers and leadership team members who had been involved in the project in prior years.
- 2) Coach turnover – Seven of the 17 project schools had turnover in their STW coach during the project period; several had two STW coach changes. For the state hubs, this often meant recruiting and hiring a new coach to take over the duties. It also meant a training period for the new coach to learn about the project as well as learn about their school. This resulted in a gap in coaching services for project schools.
- 3) District mandates, changes, and demands – The project schools are from five different school districts; three large urban districts and two smaller rural districts. Coaches had to navigate many different district requirements throughout the project including: union driven board decisions related to resources and teacher time (coaches had to be creative to find the time to work with teachers and principals); district demands that sometimes took precedence and attention away from the project focus; district accountability programs that involved teacher evaluation, compliance, and high stakes academic evaluation which created an atmosphere of pressure; and reductions in budgets and resources available to the school (e.g., loss of assistant principal position, loss of district academic coaches). All of these situations required coaches to be creative, flexible, and adaptive to the changing needs and focus of the school.
- 4) Assisting schools to continue to push forward in their refinement and implementation – It became challenging in some schools to work with them to continue drilling down to deeper implementation such as instituting varied interventions for specific students, keeping the focus on the whole child, ensuring buy-in of all teachers, and continuing to improve the rigor of instruction. This challenge necessitated perseverance by the coaches as well as using a variety of resources (e.g., STW visits, mentor schools) to continue to inspire and engage the school.

Despite these challenges, however, coaches overwhelmingly reported improvements at project schools as the project went along, and certainly by the end of the grant. Coaches reported:

*“It is amazing to see how much the teachers have improved over the past three years. They are beginning to experiment with more student-focused and interactive teaching strategies, and as a result students are more engaged in the classroom.”*

"I am observing a climate shift. Student behavior has improved since the start of the project. The principal is setting a higher standard of teacher and student accountability."

"This school has shown such positive growth in the last year and half. The building looks better, cleaner, more student friendly, and the principal is highly visible and spends a lot of time in the classroom."

"When you enter a classroom now, there is a 'vibrancy' that truly wasn't there three years ago. The teachers are very aware of the 'ideals' established in the STW Rubric."

"The principal believes that this school year has been the most successful one since he became principal of the school. He noted that the hard work of the past three years is now beginning to show some significant positive results."

### **Findings by STW Coach Turnover**

Coaching services, delivered most frequently throughout the project by STW coaches, were clearly the central vehicle for delivering the intervention to project schools. Principals were effusive in praising their coaches, with many calling them a "blessing" to the school. They recounted the positive ways coaches had impacted their schools, beginning with how they had cultivated a relationship with everyone, resulting in teachers valuing their time working together and viewing them as an integral part of the school as opposed to being an outsider. Principals also reported that the coach's efforts to promote collaborative relationships ensured a safe environment for listening, communicating, and sharing ideas. Principals also recognized the value of the i3 STW Project assigning coaches for the duration of the project, when feasible, to promote a strong relationship and consistency of services between the coach and the school.

"I appreciate that we have coaches that remain in our school that are the same individuals. In my short tenure as principal, I have had three different supervisors. So it's nice to have consistency. ...It is really important that if their work is going to flourish in our school that the trust is embedded in the coaching."

i3 STW Project Principal

Given the critical nature of the role of STW coaches and the challenges when there was turnover among STW coaches, process and outcome data collected at project schools were analyzed by whether the school worked consistently with the same STW coach throughout the project or whether there was turnover in their STW coach. As mentioned previously, 7 of the 17 project schools had turnover in their STW coach during the project period. Therefore, 10 schools worked with the same STW coach for all four years of the project. Project schools were categorized as either having a consistent STW coach (10 schools) or experiencing turnover in their STW coach (7 schools). Results of the analysis of these two groups of schools revealed striking differences between the groups. Project schools with a consistent STW coach for the entire project had statistically significant higher ratings on the STW Rubric (Figure 14), significantly higher collaboration opportunities and outcomes (Figure 15), and significantly higher "best" middle-grades instructional practices (Figure 16).

Figure 14. STW Rubric Results by STW Coach Turnover

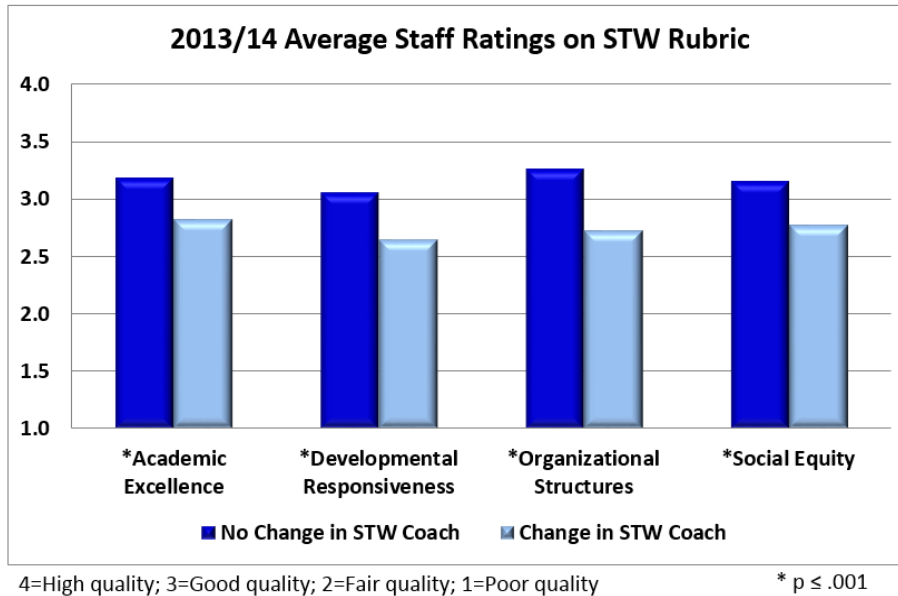
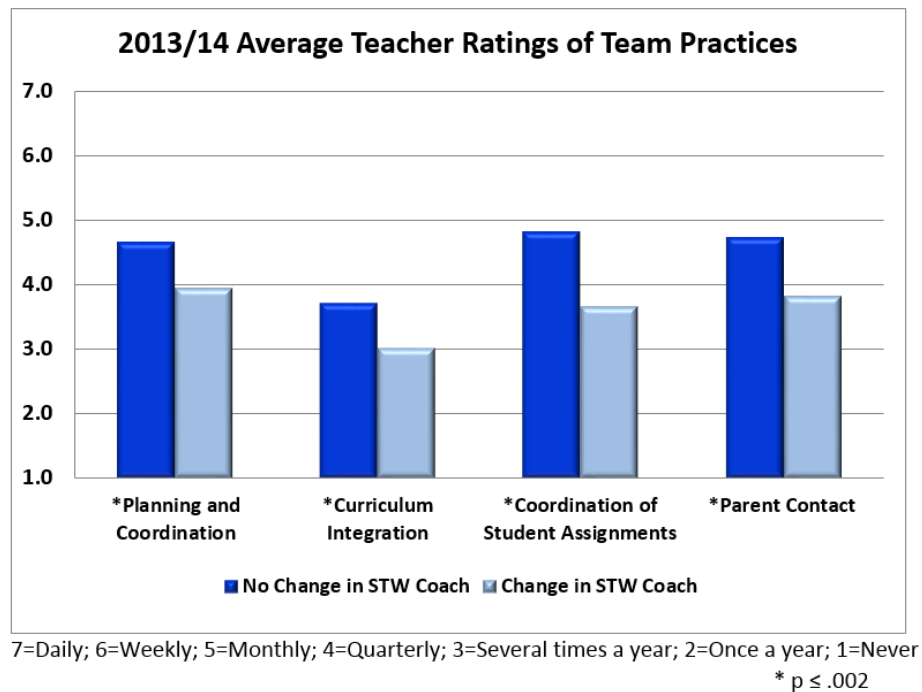


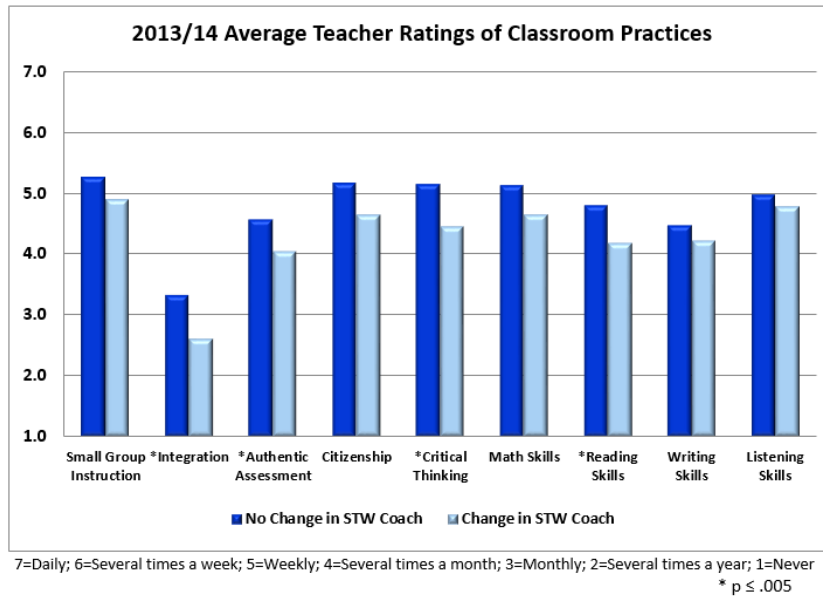
Figure 15. Teacher Ratings of Team Practices by STW Coach Turnover



The fact that the schools with a consistent STW coach for the entire project had significantly higher STW Rubric ratings and higher programs and practices than schools who experienced turnover in the STW coach is an important finding for the i3 STW Project. First, it clearly highlights the central role that STW coaches played during the reform process. Any efforts to both retain STW coaches at project schools as well as careful assignment of STW coaches to project schools so that the fit is right are important to consider and plan for. Second, consistency of services to schools is an important factor to

ensure. While it is certainly not always possible to retain a STW coach for the duration of the project, well-developed protocols and procedures for a smooth transition to a new coach, as well as a well-articulated implementation plan to follow, are necessary to have on hand before turnover occurs so that consistency in services is supported.

Figure 16. Teacher Ratings of Classroom Instructional Practices by STW Coach Turnover



### Findings by Principal Turnover

As previously mentioned, nine of the 17 project schools had at least one principal change during the project period and several had two changes. Since principal turnover, like STW coach turnover, created challenges for consistency of support and leadership at project schools and for the implementation of the intervention, process and outcome data collected at project schools were analyzed by whether the school had the same principal throughout the project or whether there was turnover in the principal. Project schools were categorized as either having a consistent principal (8 schools) or experiencing turnover in the principal (9 schools). The results were consistent with the results of the STW coach turnover analysis, however, not as compelling in the strength of the findings. Project schools with a consistent principal for the entire project had slightly higher ratings on the STW Rubric (Figure 17) and some higher collaboration opportunities and outcomes (i.e., interdisciplinary team practices) (Figure 18), however, there was no differences in “best” middle-grades instructional practices. Although schools with a consistent principal had slightly higher practices than those with principal turnover, the differences between the groups were not as dramatic as the differences between schools with a consistent STW coach versus coach turnover.

It is interesting, but will not be too surprising to the Forum and the state hub leaders, to see that differences in practices between schools with principal turnover and those without were not as great as were found in the STW coach turnover analysis. Because the STW program and thus the i3 STW Project values collaborative leadership in whatever form is best suited to the school (e.g., leadership team, teacher leaders, interdisciplinary teams, PLCs), as well as a continuous improvement model centered on

the STW criteria and rubric, project schools involve a deep cadre of teachers in the reform process. Therefore, some are able to keep the momentum going after a change in principal. The Forum believes this is one of the reasons why so many STW schools are re-designated multiple times despite changes in leadership. By involving many faculty in the STW and improvement process, it deepens the infrastructure and enables the school to stay on course in the face of change. Project schools, although not as fully implemented as STW schools, have begun to build these reserves and we hypothesize that this is the reason that the principal turnover analysis isn't as striking as the STW coach analysis.

Figure 17. STW Rubric Results by Principal Turnover

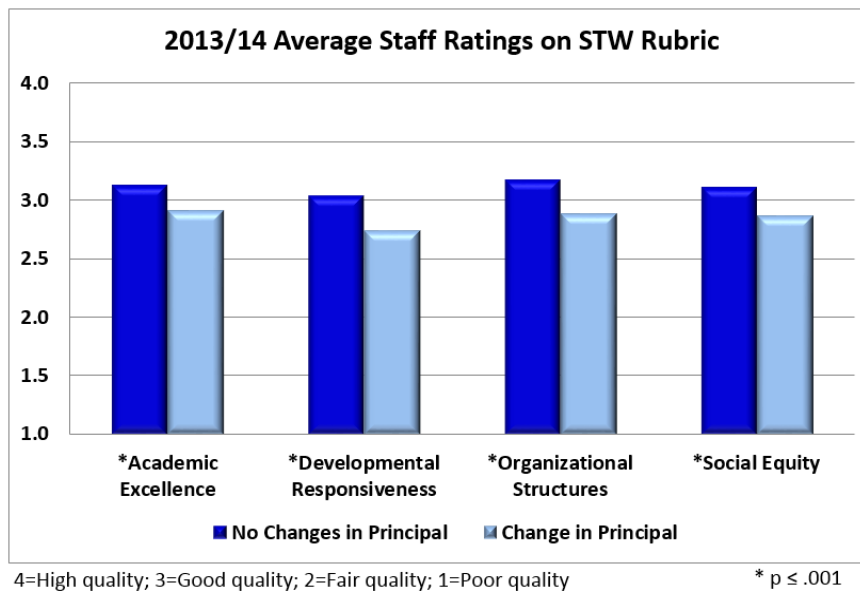
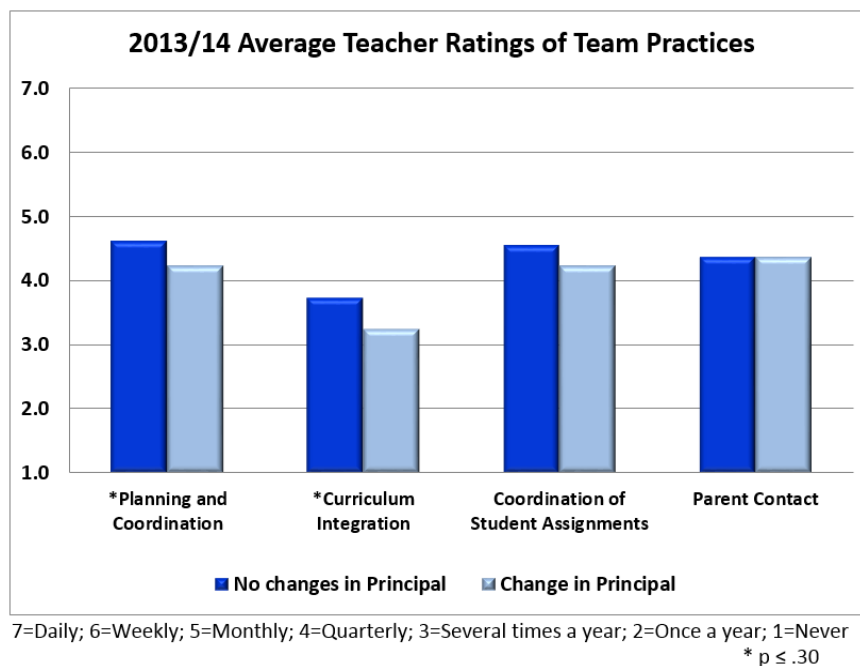


Figure 18. Teacher Ratings of Team Practices by Principal Turnover





### Observations of Improvements at Project Schools

Through coach and mentor observations, state hub leader visits, and statements gathered via the online coach's log, quarterly state reports, and focus groups, a picture of the transformation of the culture, climate, and practices of the project schools emerged (Table 14). While anecdotal, this information is powerful in its consistency across states and schools, across reporters, and with other data sources such as the STW Rubric and the teacher surveys.

Table 14. Observed Characteristics of Project Schools Before and After the i3 STW Project Intervention

Observed* Characteristics of Project Schools	
Before the i3 STW Project:	After the i3 STW Project:
Negative and often dysfunctional climate	Culture and spirit of collaboration
Isolation of teachers	Teachers sharing, meeting, and discussing
Teacher lack of trust in the administration	Support of the administration
Student behavior problems	Better classroom management and student behavior
Lack of parent involvement	Improved parent involvement
Reluctance to engage in anything	Collaborative leadership and teacher leaders
Lack of focus	Strong vision
Reactive instead of proactive	Highly visible principal
Lack of instruction rigor	Improved instructional strategies
	Comfortable using data
	Student centered focus
	Higher standard of teacher and student accountability

\*Observations were made by coaches, mentors, and state hub leaders.

According to STW coaches, principal mentors, and principals, the first observable changes in schools by the end of the second year of the i3 STW Project included:

- Schools now had a vision, goals, common language, and structure that enabled change.
- Schools had powerful examples of “best” practices through visits to STW designated schools where they observed the possibility of change.
- Schools achieved a higher level of awareness of their practices by using the STW Rubric, that involved questioning their practices and thinking more strategically about improving.
- School began going beyond just raising levels of awareness of their practices to starting to implement “best” practices.
- Principals moved toward a more collaborative leadership style, empowering their leadership teams and teachers, by giving them opportunities to collaborate and have input into the decision making process about implementation and improvement.
- School culture shifted towards a greater sense of cohesiveness among staff, as well as between staff, students and parents.

In stakeholder interviews conducted at the end of the project, state hub leaders and the project director described the overall key improvements at i3 STW Project schools in the following ways:

- Although everyone acknowledged the lack of student achievement success yet, they felt that the schools were on the brink of impacting student achievement, particularly since the changes in culture, climate, and practices were evident.

- Schools have developed collaborative relations within their schools (shared leadership, PLCs), across schools, with mentor schools, and across the states.
- Schools have made distinct and lasting collaborative leadership structures and practices.
- Development of many teacher leaders at schools.
- Principals and teachers have a newfound sense of confidence in themselves.
- “Real, legitimate change” in school climate, culture, and practices.
- The student experience is so much better now than 4 years ago and the impact will be sustained.

“Part of the success is – are these schools different than they were when we started with them and I think they are. Are they there yet? No. Have they learned enough to get themselves there? I think in some cases they are on the brink of that, particularly to affect student outcomes. So I think there’s a lot to celebrate with moving the culture, climate, and practices in these schools forward.”

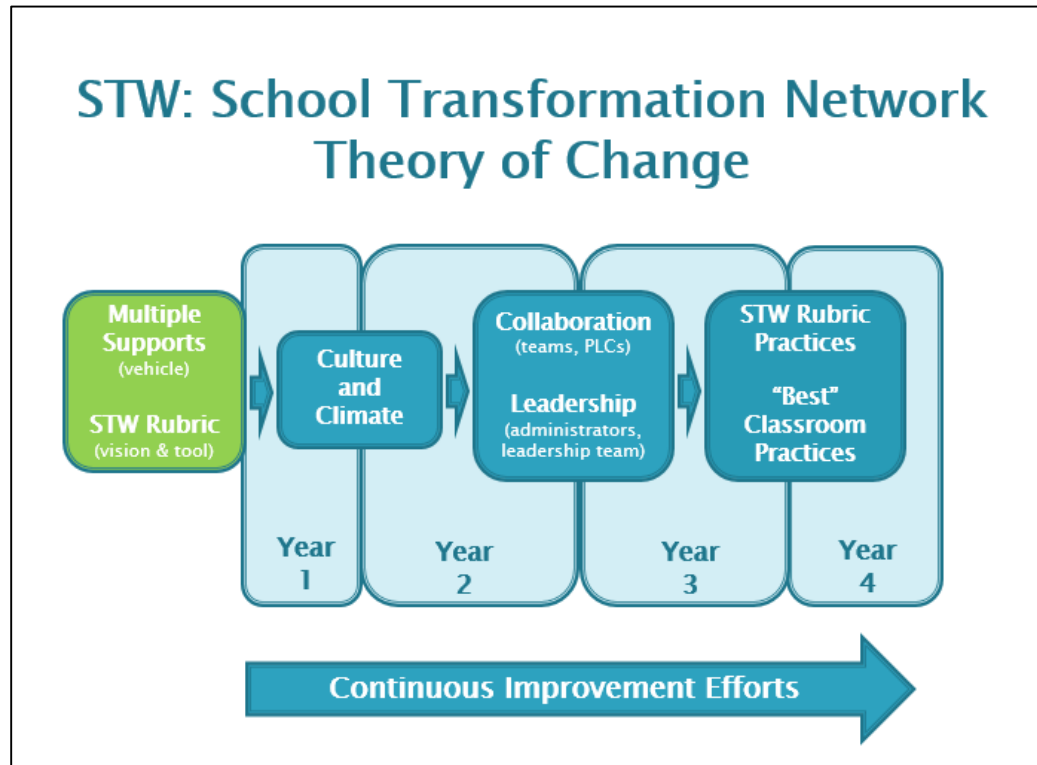
i3 STW Project Director

## Theory of Change

The results of the exploratory data analyses presented above are positive in that they highlight numerous improvements at i3 STW Project schools. The results also provide unique insight into the reform process for the i3 STW Project leaders as well as middle-grades schools that are struggling to improve their educational practices, experiences, and outcomes. The figure below (Figure 19) graphically depicts the exploratory results and what has been learned about how the low-performing middle-grades schools that were part of the i3 STW Project changed and improved, as well as the pathways in the change process. The multiple supports that i3 STW Project schools received combined with the guiding vision of the STW criteria and rubric have allowed these high needs schools to positively impact their culture and climate, their collaboration, leadership practices, STW rubric practices; and their implementation of “best” classroom teaching and learning practices (Flowers, Begum, Carpenter, Mulhall, & Poes, 2014).

As described previously, the multiple supports and services that project schools received included: tools for assessment, goal setting, action planning and monitoring; technical assistance from a STW coach, principal mentor, and a mentor STW school; networking opportunities at the state and national levels; implementing an early warning indicators system; and focused professional development designed to build a learning community and address the needs of students at risk of educational failure. The STW coaches and the principal mentors were considered to be the “vehicles of change” by coordinating and managing the multiple supports and always focusing schools on the bigger picture of the STW vision by connecting everyday work into a larger context. They also facilitated the development of collaborative leadership, provided professional development and modeling “best” classroom instructional practices, and facilitated the use of data for continuous improvement. The multiple supports were delivered to project schools within the framework of the STW guiding vision and criteria. The STW rubric was used as a tool within the continuous improvement process to support facilitated reflection, goal setting, and the examination of data.

Figure 19. i3 STW Project Theory of Change



The findings from the exploratory analyses of the evaluation data showed that project schools receiving these multiple supports demonstrate improvements in their culture and climate by Year 2 of the project. This includes improvements such as the establishment of trusting relationships among teachers, collective support for improvement, improved work climate, improved collective teacher efficacy, and increased teacher involvement in decision making. The data also showed that in Years 2 and 3, staff at project schools began experiencing improvements in collaboration practices, leadership practices, and STW Rubric practices. In terms of collaboration, teachers reported working together more often (i.e., team practices, team decision making) and experiencing improvements in the quality of interactions when working together. For leadership practices, teachers reported that the building leadership at their school was more effective, particularly in areas around communication with the faculty, follow-through on decisions and next steps, problem-solving strategies, handling conflict in the building, and a proactive approach to school improvement and bettering the school climate. Teachers also reported implementation of higher levels of STW recommended practices beginning in Year 2, but culminating to their highest level by Year 3 and maintained in Year 4. Finally, the data showed that the implementation of “best” classroom instructional teaching and learning practices improved in Year 3 and continued to improve into Year 4.

The progression of change over time in the project schools is noteworthy for the insights it provides into the key elements or building blocks that change first in low-performing school. With a longitudinal study such as this, the time frame in which improvements occur is helpful to understand. In the case of the i3 STW Project schools, culture and climate were the first areas to improve, followed by collaboration and leadership, then finally best practices within the STW framework and best teaching

and learning practices. We cannot assume that this progression will be identical in all low-performing schools; however, many have similar challenges and therefore we hypothesize that these key elements will be part of the change process at low-performing schools.

It is also interesting to note that these results showed some variability in improvements by state, suggesting that implementation variances by state may be impacting school progress. We know that the schools in Illinois had higher overall implementation rates. It may have been due to the fact that they had smaller enrollments and thus fewer teachers, so the change process may have moved more rapidly. It may also be related to the instructional coaches that Illinois used. We also know that the project schools in rural settings in California and North Carolina often had less administrative bureaucracy than the larger urban districts, which may have impacted their rate of implementation of project objectives. We can say with certainty, however, that consistent implementation and high levels of implementation support improved practices, programs, and outcomes.

It can be seen from the theory of change graphic that the progression of change over time is supported by continuous improvement efforts. This is depicted as a foundational aspect of the project because all project schools used a continuous improvement model to set goals and review progress. Through the project's evaluation, schools received numerous data reports on their school and state (STW Rubric, Self-Study Surveys, Coach's Log, focus groups) to inform their planning, set goals, reflect on and evaluate their progress, and refine implementation. This structured approach to ongoing improvement was critical to keep project schools focused on their goals and continuing to progress.

Although the theory of change does not yet connect the contextual changes in culture and climate, collaboration, leadership practices, and the implementation of "best" classroom instructional practices with student achievement growth, it is encouraging to find these improvements in programs and practices because improvements in these areas are the building blocks to improvements in student achievement. Student achievement cannot be improved without first changing and improving the teaching and learning process. The fact that instructional practices were the last area to improve, beginning in Year 3 and continuing in Year 4, may mean that these more rigorous instructional practices were only experienced by students toward the end of the project and therefore did not have enough time to impact their learning and thus their achievement scores. It is also encouraging to find factors that are positively associated with student achievement, such as student feelings of academic efficacy. These findings assist in understanding the factors that influence student achievement. Finally, the fact that the three project schools who were designated as STW during the project had both higher programs and practices, as well as higher math achievement is also encouraging and leads to the hypothesis that the outcomes are feasible and other project schools may need more time to get to this level (Flowers, Begum, Carpenter, & Mulhall, 2015).

The educational importance of these findings are in rigorously evaluating an innovative approach that provides multiple supports to struggling middle-grades schools, in assisting them in implementing best practices and in learning from these lessons and applying them to refining the intervention for replication and application in other settings. The results have the potential to assist middle-grades schools to understand the culture and practices that build the foundation for a healthy and collaborative environment that supports and challenges students to succeed both socially and academically.

## Unexpected Findings/Spillover Effects

While the i3 STW Project logic model (Figure 1) clearly illustrates the anticipated intermediate and long-term outcomes expected from the implementation of the nine programmatic components, in practice there were additional, unexpected outcomes as a result of the spillover effect that benefitted the project schools, state hubs, and the Forum. This section highlights the most significant findings.

### Project Schools

Increased value of the importance of data for evaluation and improvement – Working with coaches, project schools evolved from looking at a single source of data as an isolated activity to developing protocols to utilize multiple sources of data for data-based decision making and school improvement planning with the following features:

- Utilizing existing groups to examine data (e.g., leadership team, PLCs)
- Structuring time on a regular basis for looking at data;
- Focusing on individual student data;
- Keeping the improvement vision and data visible; and
- Using the data to validate existing practices, move beyond the plateau, and empower teachers.

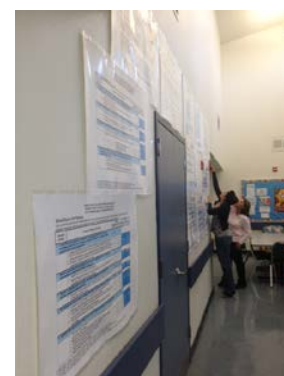
*“Are we where we want to be? Every year, you want to improve. You know, our scores aren’t what we want them to be. And we want to get better. Every year, we identify new goals through the STW rubric and our own improvement. I don’t think any principal will ever say, ‘Yes, I’m done.’”*

i3 STW Project Principal

The use of data transformed schools and became part of a reflective, collaborative culture for a faculty that was no longer “scared of using data” but was anxious to get their results back from the evaluators to measure progress and adjust targeted school improvement goals. School improvement planning no longer had an end date but became an ongoing process with schools continuously raising the bar for what they wanted to achieve for their students and their school. Schools converted meeting rooms and teacher lounges into data rooms and papered the walls with data to facilitate discussions and as a visual reminder of their commitments.

Reflective practices at the school level trickled down to individual teachers to help them “teach smarter” and focus on individual student needs. Teachers reported that the i3 STW Project helped them look more critically at their teaching practices to become better teachers through self-evaluation, looking at standards, and using data to focus early indicator work and to prepare students for the state achievement test.

Collaboration among project schools – Initially there was no communication, sharing of ideas, or collaboration between project schools within the same school district because they believed they were in competition with one another. This changed when the i3 STW Project began scheduling meetings before or during professional development sessions, as well as at the



STW Rubric results blown up and posted in the Teacher’s Lounge

state and national STW conferences. Inter-school comradery stemmed from principals having frank conversations about how they were implementing the i3 STW Project at their own schools. Principals realized the benefits of collaboration across schools and began to consult with one another and meet together outside of project-arranged events.

“I think for me the biggest piece was just to know that you’re not alone and that everyone else has had the same struggles you have. And so when you talk to other principals, ‘Oh, you went through that? How did you come through it and become successful?’”

i3 STW Project Principal

Project principals in one school district formed a PLC where they learned from and supported one another, in part by making arrangements for collaborations between their teachers (e.g., teacher exchanges across schools to provide professional development on best practices and core content; and grade-level core teachers planning lessons with their counterparts at another school where each school only had one grade-level team). Principals also participated in mock STW visits at each other’s schools where they provided feedback to the school visited, and also brought back ideas to implement in their schools. Instead of anticipating that they would feel vulnerable in opening up their schools for a visit, principals were empowered to make changes based on the feedback of their district cohort.

Collaborations also occurred between principals across school districts based on networking at the national STW Conference. Afterwards principals phoned and emailed one another to share ideas. Both across districts and within, mentoring partnerships developed. Overall, principals talked about how fortunate they were because they were “growing together” through the project.

Increased project school attendance and deeper involvement with their state hub/middle school association – Initially principals reported their appreciation for being invited to attend state conferences and professional development opportunities. Schools had rarely had this kind of exposure to quality professional development, which was a strength of the grant. Not only did school attendance increase at these events, but schools exhibited a deeper level of involvement with their state middle school association. One of the state hub leaders characterized this as, “We’ve brought them deeper into the fold and we’re planning on continuing those relationships through the middle school association after the project ends.” Additionally, some principals and teachers moved from spectators to leaders, such as principals serving on the state board or principals and teachers making presentations based on their experiences with the project. Coaches reported that all presenters were empowered by their experiences at the state-level conferences and with a newfound confidence they looked to present, or did present, at a national conference, such as AMLE or the STW Conference. These presenters also inspired their project colleagues in the audience with the belief they could also lead in delivering professional development, thereby generating more

“We invited teachers to come and present [at the state conference]. And it built their self-esteem because they believed presenters were only being asked because they were from good schools, the high quality STW schools. And it really empowered these teachers to want to present at the state level. And we’ve even had them say, ‘Do you think that we can go present at nationals?’ And to us, they have come such a far way from them being looked at and directed, to ‘We can do this. We want to show them how we did it and what we can share with them.’”

i3 STW Project Principal Mentor

presentations. One positive impact of principal turnover was that former i3 STW Project principals not only continued their involvement with the middle school association through attendance at events, but brought their new staff with them, thereby increasing the reach and impact of the middle school association.

One of the state hubs collaborated with coaches and teachers to develop a professional development series where they utilized teachers from i3 STW Project schools to provide training to other project schools and non-project schools in the district. This professional development series impacted schools in the following ways: 1) making connections across schools so they can continue to grow; 2) building teaching repertoires through professional development; 3) reinforcing knowledge when it is communicated to someone else; and 4) confidence as they are “honored for their expertise.”

“One of our i3 schools presented at the state conference and it turned the whole school around. They came back and they were so proud, so excited, so energized that they immediately applied to present at the AMLE conference and were accepted! And that whole community was like, ‘Whoa, one of our schools is doing this?’ I mean it’s been amazing!”

i3 STW Project STW Coach

---

### ***Three i3 Project Schools Were Designated as Schools to Watch:***

Hamlet Middle School, Richmond County, North Carolina was designated as a STW school in 2013.

Nathan S. Davis Elementary School, Chicago, Illinois was designated as a STW school in 2014.

Eli Whitney Elementary School, Chicago, Illinois was designated as a STW school in 2015.

---

From i3 STW Project school to being designated as a STW – While there was an expectation that the i3 STW Project would lay the foundation for project schools to become STW schools in the future, it was unexpected that schools after 2 years of project services (Hamlet Middle School), 3 years of project services (Nathan S. Davis Elementary School) or even 4 years of project services (Eli Whitney Elementary School) would be able to turn around from being persistently low-performing schools to receiving a National Forum designation of STW. One of the initial challenges in matching STW mentor schools with project schools was identifying STW schools with similar demographics given the high percentage of at-risk students within the project schools. These three schools are now included in the cadre of STW mentor schools that will be disseminating their lessons learned from their experiences in the i3 STW Project statewide to schools that are struggling to improve, and the project expects the designation of several more schools in the upcoming years.

### ***State Hubs***

State hubs developed a cadre of expertise in their states in the form of their project leadership team, coaches, mentors, and professional development providers – The state hubs developed effective leadership teams where each team member might play a different part, but through communication, collaboration, and a high level of leadership, the leadership team was successful in guiding and managing the project. Critical to this success was follow-through and persistence “no matter what the changing circumstances.”

Each state leadership team hired coaches with significant experience in implementing best middle-grades practices. While coaches were initially hired to provide reform coaching, i.e., working with schools on global culture and environment, building collaborative leadership etc., the needs of their schools necessitated that coaches additionally provide instructional coaching and “responsive” coaching. As a result of these adaptations, coaches deepened their skill sets and coaching improved over time.

The state leadership teams, with the help from the coaches, developed targeted professional development through a powerful combination of a good keynote speakers to get teachers motivated, and breakout sessions to provide them with the necessary tools to take action. Project schools received consistent and responsive professional development both through statewide events and those brought *directly* to their school.

“I think the state leadership teams have all learned a lot about helping schools, delivering services, beginning to understand what you have control over and how change happens, and what you can do to move schools to the next level of change.”

i3 STW Project Director

Additional professional development and mentoring came from effective use of the existing state STW networks as the state leadership teams and coaches reflected on the needs of each project school, matched them with the best STW school, and provided guidance on visits to facilitate lasting relationships between schools. This active guidance also built capacity in the STW schools on how to better serve as mentor schools.

Strong collegial relationships developed between state leadership teams and school staff that will endure beyond the grant – Whenever principals or teachers were asked what factor was most responsible for their school’s improvement since the i3 STW Project began, coaching (and more specifically their STW Coach) was the first thing mentioned. Principals and teachers praised their coaches for: developing relationships with administrators and teachers; being supportive, passionate, and committed to helping; providing a student-centered focus; and helping the school stay on track and focused on school improvement. In essence, coaches were viewed as a true partner with the school.

Both principals and coaches talked about a “strong state organization” that kept project schools informed and connected about the i3 STW Project and professional development opportunities. State leadership teams reported intentionally establishing collaborative and proactive relationships with principals and schools that they are committed to continuing beyond the grant. During the no-cost

“My teachers are willing to stay after school, come in on Saturdays, to meet with these coaches. That means something’s going on that’s valuable to them. I also like the fact the coaches...they’ll come into my school [and say] ‘What can I help you with?’ And my teachers feel like they are special because they have this coach that meets with them.”

i3 STW Project Principal

extension year, state leadership teams provided limited services and continued checking in with project schools on their progress and needs, as well as connecting them with additional supports to continue their work. State leadership teams are looking for future opportunities and funding to involve schools when they can, but also believe that schools will solicit their help in the future when applying to become a STW.



Inroads into districts by the state hubs – State hubs reported that they positively impacted the school districts in the project by creating a greater awareness of the unique needs of the middle-grades and knowledge of tools and resources, such as the STW Rubric and STW visits, available through the project to assist schools. State hubs envisioned that some of their districts, now familiar with these resources, would train new principals by giving them the school’s STW Rubric results as an introduction to the school and inviting them to visit a neighboring STW school to see how change can be implemented. State hubs also believed that the school districts now had a better understanding of the responsibilities of principals as both administrative and instructional leaders as a result of their work with the project.

Like their principals and teachers, project superintendents were excited to attend the national STW Conference in Washington DC to learn how STW schools had implemented the changes that made them successful, and to visit Capitol Hill to talk with their elected representative about the work that they were doing with the i3 STW Project to improve their schools. These two opportunities inspired and motivated superintendents to envision, and in some districts expect, all of the middle-grades schools in their district to become STW. As with the project schools, the state hubs intend to maintain and grow their relationships beyond the project with current and former school district personnel.

Strong collegial relationships developed between state hubs that will endure beyond the project – Although members of the state hubs had known each other professionally through the Forum, the i3 STW Project was the first time that they worked together collaboratively. Constant communication through bi-monthly phone calls, quarterly face-to-face meetings, and frequent emails enabled their relationship to evolve and allow for frank, open discussions where members could “say what was on their minds without fear of repercussions of how anyone would take it” and they believed it made the project stronger that different opinions could be expressed and respected.

“We began to form our own PLC among the three states, and the leaders of the three states were able to share not only the positive things going on, but the frustrations as well, and learning to really – all three – work together, and talk to each other, and listen to each other.”

i3 STW Project State Hub Director

The regularly scheduled phone calls and meetings were also helpful in understanding how each state was implementing the program in their schools, the challenges they were facing, and the successes that they were having. State hub leaders realized that they did not have all of the answers individually and that they needed to listen and learn from one another. The framework of the model, the STW criteria, served as the unifying principle that connected the three different state efforts. Even when approach varied due to starting point of the school or state organizational structure, they were able to fall back on a shared vision and goals to guide their work.

In stakeholder interviews at the end of the project, state hub members reported that they believe their collaborative relationships with each other will continue after the project as they view each other not only as national partners, but as friends. They are already calling on one another for help with work outside of the i3 STW Project and applying for grants to fund future work together

State hubs developed expertise and willingness to share lessons learned and build capacity among other STW states – As a result of their work on the i3 STW Project, state hubs have a broader understanding of the needs of struggling schools, greater capacity and creativity to address school needs, and more confidence in how they go about helping schools to improve. The state hubs have grown as organizations, are sharing their lessons learned, and beginning to mentor other states on how to work with high need schools. Other STW Directors are beginning to engage in this work, and the goal is to encourage them to be more entrepreneurial in thinking about what they can do to provide additional services, such as a summer institute at STW sites for struggling schools. In addition to providing technical assistance and mentoring, the state hubs, in collaboration with the Forum developed a handbook for STW Directors, coaches, and mentors for replicating the i3 STW Project’s work with schools. Other resources they have to offer include: protocols for schools to make the most out of their visits to STW schools; project-developed early warning indicator strategies and tools; case studies that were written about project schools utilizing the framework of the STW Rubric; as well as successes with grant-writing experiences, including leveraging the i3 STW grant to receive funding to support additional services such as instructional coaching.

“How can we help you learn to do it your way but still stay true to the program? And if you don’t have tons of people or resources, what can you do?’ We can inform other STW state hubs about what needs to happen, what they can find in terms of funding, promoting it with their state departments or regional funders, etc. We can begin to raise up other states.”

i3 STW Project Director

Increased appreciation and value of the evaluation data – The state hubs believed that the high level of data reported back at the individual school, state, and project levels from the evaluators was a “strength of this grant.” State hubs recognized the leadership of the evaluators to guide them through understanding the evaluation and the importance of using the data to guide improvement plans. They believed the evaluators made the data “meaningful, digestible, and understandable” not only to them, but to their schools, which facilitated the work that state hubs and coaches were doing with schools to develop continuous school improvement planning.

The state hubs appreciated that the evaluators collected a variety of data sources (STW Rating Rubric, School Improvement Self-Study Surveys, Coach’s Log, focus groups, and achievement data) for a comprehensive and exploratory analysis of the project implementation and outcomes. State hubs valued the process of evaluators leading the group through the reflective exercise of creating the fidelity of implementation matrix based on the logic model, and then revisiting it annually in order to refine program interventions. The state hubs also appreciated the evaluators’ willingness and responsiveness in analyzing the data to explore their theories and questions regarding why a particular intervention did or did not work, which arose through data conversations with the evaluators.

### **National Forum**

The project elevated the National Forum’s visibility and credibility – Through their work on the i3 STW Project, the Forum contributed to the middle-grades field by: presenting at the first i3 Project briefing by the USDE; presenting at research (AERA) and practitioner-based (AMLE, Learning

Forward) conferences and the i3 Project Directors Meeting; and participating in the i3 Learning Network meetings sponsored by the Spencer and W.T. Grant Foundations. This visibility led to further opportunities such as NCEES asking the Forum to serve on an expert panel to advise them on the development of the national Middle Grades Longitudinal Study, support from the Raikes Foundation to attend their Deeper Learning Conference, etc. Overall there is enormous pride in the National Forum that the organization successfully received and implemented an i3 grant.

Applying lessons learned from the i3 STW Project –The National Forum is already applying lessons learned from the i3 STW grant to inform its 2013 i3 Middle-Grades Leadership Development grant, such as building in time upfront to get a “clear visualization of our management plan” and utilizing “backward mapping” to prepare for implementation across multiple states, seeking greater involvement from school districts by requiring MOUs and a school district representative on the school’s leadership team, etc. Recently submitted grant proposals by the Forum have built upon the findings of the i3 STW Project, particularly the recognition that both culture and climate change as well as instructional interventions are needed to rapidly affect high need schools.

“We’ve managed a multiple million dollar grant... we’ve been able to show that we have the capacity through the state hubs to pull that off. I think it’s really been successful for us [National Forum], too, in that it’s put us in situations where we’ve gotten increased visibility and there’s been a sort of a sense of legitimacy brought to our work because of it and I think that’s important, as well.”

i3 STW Project Director

## Persistent Challenges

With any complex project such as the i3 STW Project – whole school reform model, multiple states, multiple districts, nine unique programmatic components, etc. – a variety of challenges are bound to occur. Many of challenges faced by the i3 STW Project director, state hub leaders, coaches, principal mentors, and principals were monitored, discussed, and resolved. Other challenges, however, were more persistent and reoccurring throughout the project period. Data collected on the project as well as minutes from project leader phone calls and meetings highlight the most persistent challenges encountered by the project leaders and project state hub teams.

### *Disruption from Unexpected Personnel Turnover*

One of the ongoing challenges to the implementation of the i3 STW Project was the all too frequent, and unexpected, turnover of STW coaches (7 out of 17 schools), principal mentors (12 out of 17 schools), and principals (9 out of 17 schools). In fact, only two out of the 17 schools maintained the same principal, STW coach, and principal mentor throughout the four-year project. Turnover occurred for a variety of reasons such as not having a good fit between the coaching staff and the school, illness, school district decisions, and career opportunities. The turnover of STW coaches and principal mentors resulted in an interruption of project services provided to the school during the changeover period and then a further delay in the project work while the new person built trusting relationships with the principal, and in the case of the STW coach, with the whole faculty. Principal turnover resulted in the disruption of leadership (and sometimes school culture) during the changeover period followed by the

need for the principal to learn and understand the i3 STW Project services being provided, expectations for the school's participation, and again building a trusting relationship with the coaching staff. Turnover also occurred with school district personnel, such as the superintendent, assistant superintendent, or curriculum director who had previously supported the i3 STW Project in the school thus requiring the state hub team, STW coach, and principal mentor to re-establish relationships and re-sell the project to garner continued support. In hindsight, it would have been beneficial to the ongoing work of the project to have established protocols for all of these turnovers at the beginning of the project to smooth transitions when they arose.

### ***Balancing Structured Implementation with Realities in Each State***

Despite the i3 STW Project's shared vision, common goals, logic model, and fidelity of implementation description, adaptations to programmatic components occurred within states due to the complexities and realities of implementing the project in each state. Differences by state in such factors as school size, geographic location, and added resources from external sources or district sources, resulted in implementation variability across states. For example, in order to accommodate the large size of schools in California and North Carolina, a "train the trainer" strategy of sending the leadership team (instead of the entire faculty) to attend focused professional development and visit mentor schools was employed. Another example of variability was in California where distant, geographic locations presented challenges for schools to work with principal mentors, thus necessitating principal mentors to engage in project work by phone and email rather than through personal visits. In California and North Carolina, mentor schools were not located close enough to project schools to allow for a one-to-one assignment with ongoing visits by all faculty throughout the project. Instead project schools selected mentor schools to visit based on a particular need or intervention that the school was looking to implement and sent the most relevant staff for this purpose. Finally, the last key adaptation by state was the additional coaching services, instructional coaches, which were delivered to schools in North Carolina and Illinois, providing another layer of support not given to California schools. The North Carolina instructional coaches were brought in by the district/state and the Illinois instructional coaches through a grant from the Fry Foundation.

These variabilities to the structured implementation plan for the i3 STW Project often resulted in conversations about whether the variabilities were being appropriately balanced with the need to consistently implement the project across states and schools. It was often challenging to address these state-specific complexities and realities in such a way that the project leaders agreed that the variabilities were within an acceptable adaptation range from the original intent of the project. The state hub leaders discussed variances with the entire project leadership team for input, feedback, and final decisions about how to proceed. Additionally, all variations to the original project plan were documented. The implication of this variability, however, results in some varied experiences for schools.

### ***Project Components that Did Not Work Out as Planned***

There were two programmatic components of the i3 STW Project that were never fully implemented as intended. Both the principal mentor component and the early warning indicators component had a variety of persistent challenges which fundamentally changed the intention of these components. Principal mentors faced three barriers for implementation: their role and responsibilities were not specifically defined at the beginning of the project, making it challenging and time consuming to select

and match mentors with schools; a high mentor turnover rate, which interrupted the delivery of services; and limited access to project principals due to busy schedules or district restrictions on the principal's time. In the end, principal mentors were not always able to establish the intended principal partnership or spend the prescribed amount of time with principals with the result that this intervention was never implemented with fidelity during the four years of the project.

The design and plan for implementing the early indicator work was ambitious and in retrospect, the project leaders agreed that it should have been staged for implementation differently. Additionally, the roll-out plan for the more advanced trainings on identifying and designing specific intervention strategies for tiered support were not effective and left some projects schools struggling to make the work meaningful and overwhelmed with the sheer number of students requiring services. Another challenge for schools was trying to create an in-house data management system where teachers could input and then query their collected data in order to target student services directly. As a result the early warning indicators were not implemented as intended or at all in the first two years of the project. In response to the need to provide these services to students, the project director restructured this component to better meet the needs of schools. Also, some individual schools and coaches began to develop their own strategies and tools for use with their school, which they then shared with colleagues. Although one coach developed strategies and a protocol that was shared throughout the state in California, the overall result of this organic development of the early indicator work was that the intervention was not implemented in a uniform way across schools and states.

### ***Changing District Requirements***

Unfortunately, district demands often took precedence, and sometimes attention, away from the i3 STW Project focus. In particular, principals were often laden with district changes, issues, and requirements that pulled them frequently from their buildings and project work. For example, a 7-day teacher strike in September 2012 in Chicago Public Schools; the closure of a project school in Durham Public Schools after Year 2 of the project; two Illinois project schools facing potential closure along with nearly 50 other schools in the district; and in Year 4, Compton Unified School District began working with their principals three to four days a week for most of the year allowing no time for principals to work with principal mentors. In addition, districts had their own professional development opportunities and requirements for schools to implement in order to improve schools that were chronically low-performing. These good intentions for improvement were sometimes unfocused, creating goals that were in competition with one another. School culture was further impacted as districts fired or reassigned principals interrupting the collaborative leadership begun by the project. One of the ways that the i3 STW Project staff responded was to have coaches and mentors align these competing services and manage the many changes by making connections with the school's school improvement plan and the i3 STW Project work. This assisted many schools with staying focused on their vision for improvement and not becoming overwhelmed nor pulled in too many different directions.

### ***Changing State Achievement Tests***

The impact study utilized student English/language arts/reading and math achievement scores from standardized state assessments to explore whether the project had a positive impact on student outcomes. With state and national adoption of the common core standards, this led to instructional and curricular changes as well as changes to standardized assessments in all three project states. In 2012/13,

North Carolina implemented a new standard course of study in all subjects and grade levels, aligned to the common core, along with a new state assessment. Fortunately correlational analyses conducted between scores from the new 2013 test and prior test scores determined that the new 2013 test was comparable and appropriate for use for in the analysis. Also in 2012/13, Illinois revised some of its test and performance expectations in order to align to the new common core state standards. In California the state suspended all standardized testing and reporting in 2013/14 to allow school districts to prepare for and transition to the new California Measure of Academic Performance and Progress assessments, slated for administration statewide in 2015. The result of these shifts in standards and testing created a state of flux in all three project states and had an impact on the evaluation team's access to data as well as likely had an overall impact on student achievement during these transitions.

## **Sustainability and Growth**

The project director, state hub leaders, coaches, mentors, and principals involved in the i3 STW Project have expressed a belief that project schools will not only sustain the improvements they have made, but continue to grow and improve. Sustainability by definition is the “capacity to endure” and although the evaluation cannot predict whether project schools will sustain their accomplishments and continue to grow, data collected on the project highlight several key structures and practices that are likely to support sustainability and continued growth.

### ***STW Program and Network***

i3 STW Project schools are now immersed in the Forum's STW program and network of schools, both at the state and the national levels. This is an established network of more than 300 schools across the country. Project schools can tap into the knowledge, resources, and support of the STW program and networks in order to continue to advance the improvements at their schools. Additionally, a fundamental aspect of the STW program itself is focused on school improvement efforts that are characterized by a continuous trajectory toward success. Use of the STW rubric supports this approach, allowing schools to continually use the criteria to rate themselves, assess where they are, set goals, and continue to improve. Project schools now have this mindset of continuous improvement as part of their cultures, as well as structures to support it. In fact, at least three project schools have indicated their intention to apply for STW designation in the next one to two years.

### ***School Improvement Practices and Structures***

All of the i3 STW Project schools have adopted ongoing practices in their buildings which support their continued improvement. These practices, all fundamental to the i3 STW Project, include collaborative leadership, continuous improvement mindsets, a strong vision of excellence, reflective and collaborative cultures, and regular use of multiple sources of data. Additionally, project schools have established structures in their schools to support these practices such as leadership teams, PLCs, and interdisciplinary teams. The combination of practices and structures to support them that project schools have developed as part of this work, can support the sustainability and continued growth of these schools. As stated by one i3 STW Project principal, “We have created this model where improvement is the goal, instead of continually trying things and throwing them out. Now we feel like we really have implemented this continuous improvement model and really owned it.”

### ***Middle School Associations***

As a result of the i3 STW Project, schools have not only been introduced to their state-level middle school association (California League of Middle Schools, Association of Illinois Middle-Grade Schools, and North Carolina Association for Middle Level Education) and national middle school association (Association for Middle Level Education), but they have also participated in conferences and many have presented at conferences. This connection, particularly to their state associations, is a resource that is easily accessible and will continue to be available to project schools to support them as they work to sustain their growth and continue to improve. Project schools can attend conferences, participate in professional development sessions, connect with other middle-grades schools, access resources, and have an outlet to present their accomplishments.

### **Lessons Learned**

There are a myriad of lessons learned from the i3 STW Project that are important to document as part of the evaluation of the implementation and outcomes of the project. These lessons provide valuable information for future implementation of this project and serve to inform other middle-grades schools embarking on whole school reform. In fact, many of these lessons have already informed the planning and implementation of the National Forum's 2013 i3 grant (Middle-Grades Leadership Development Project). Although there were many differences between the seventeen project schools such as their location (urban/rural) across the three states, enrollment, building grade configuration, ethnic composition, and district and state requirements, they shared many of the same challenges and struggles to improve their educational practices, experiences, and outcomes. The lessons presented below reflect findings from the data collected as part of the evaluation from all stakeholders in the project including principals, leadership teams, teachers, coaches, principal mentors, state hub leaders, and the project director.

#### ***The Importance of Coaching***

i3 STW Project schools were unanimous in declaring coaching as the factor most responsible for their school's improvement throughout the project period. Three types of coaching were necessary for successful improvement: reform coaching (i.e., working on global culture and environment, building collaborative leadership), instructional coaching (i.e., working with teachers on instructional practices and interventions), and responsive coaching (i.e., responding to the changing needs of their school). Instructional coaching was not initially incorporated in the project as a unique and intentional aspect, however, it was recognized by project leaders as an important feature to address, and they agreed it would be incorporated earlier in future implementation of the project.

By cultivating a relationship with everyone in the school and achieving buy-in, coaches were able to provide valuable resources, supports, and focus. Coaches kept the school on-track and focused by helping them to: utilize the STW Rubric, both as a vision and as a tool for continuous school improvement; develop collaborative leadership; and increase academic rigor through improved classroom instructional practices and strategies. Coaches also helped schools to develop and maintain a student-centered focus, and served as a liaison to the principal as well as an advocate for what the teachers needed. Coaches were instrumental in setting up and leading visits to STW schools matched to

their school's needs and in facilitating professional development and networking with other i3 STW Project schools.

### ***Empowering Teachers through Collaborative Leadership***

The i3 STW Project provided the structure and support to schools to facilitate the development of collaborative leadership. Project school principals now utilize a collaborative leadership style that empowered their teachers and provided them with opportunities for collaboration, reflection, and shared leadership through participation in the leadership team, interdisciplinary teams, PLCs, etc. More importantly, these collaborative leadership practices built school capacity to sustain the changes they were implementing. Teachers now share a collaborative culture with their principal, team, and the overall faculty, and value an environment where their opinions are welcomed and where they feel comfortable discussing problems because the principal invited dialogue and provided them with a safety net to try something new. Teachers are now sharing their knowledge and experience with each other and this adoption of a “team concept” helped teachers to focus on the “whole” student. Time provided for collaboration, (e.g. PLCs, grade-level teams, interdisciplinary teams, etc.) often focused on the individual student, such as monitoring early indicator data, goals, and challenges to identify ways to provide a support system for each student and to celebrate their successes. Project schools had begun “teaching smarter” as a result of their collaboration since they were sharing ideas, coordinating lessons, and looking more critically at their teaching practices.

### ***The Guiding Vision of the STW Rubric***

The STW Rubric provided a vision to project schools for what they wanted their schools to become, a common language and framework to focus discussions, and a powerful tool for cultivating individual teacher and school-wide reflective practices since the rubric utilized an evidence-based assessment for decision making, as well as setting and monitoring targeted goals for school improvement. Using this tool, schools delved deeper through articulating the meaning of each criterion, how they could implement it successfully by identifying their school's strengths and weaknesses, and validating their existing best practices. Schools carefully monitored and evaluated progress on targeted school improvement goals and made adjustments to their implementation practices. Teachers also employed the rubric to evaluate their individual practices and make adjustments. The STW Rubric became the basis for continuous school improvement efforts and project schools identified other data sources to facilitate a better understanding for assessing and monitoring targeted goals. In some project schools the STW Rubric had a visible presence in the building as it was posted on the walls of the teacher's lounge or other common area. Finally, the STW Rubric provided a guide for teachers to optimize their observations, questions, and resource gathering during STW visits.

### ***The Importance of a Continuous School Improvement Model***

Through the i3 STW Project, schools adopted and then facilitated their own continuous school improvement model where data was used at every stage of a never-ending cycle to: inform planning, set targeted goals, reflect on and evaluate progress, and refine implementation. Improvement itself was now the end goal and this transformed school culture as teachers collaborated to discuss and reflect on data in order to improve practices. The adoption of a continuous improvement structure began when coaches introduced the STW Rubric as a vision and tool to inform school decision making with a data-driven approach. Coaches guided faculty through structured conversations utilizing STW criteria to



identify school-level strengths and weaknesses, and to validate existing best practices. These early coach-led discussions provided schools with the foundation to later formalize the process by employing existing groups (e.g., leadership team, PLCs, interdisciplinary teams), scheduling time on a regular basis, and utilizing a process or a protocol for focusing the way a group was examining data. This structured approach to ongoing school improvement was critical to keep the focus of project schools on their goals and its existence will assist these schools in sustaining their work after the project ends.

### ***The Powerful Impact of STW Visits***

STW visits allowed groups of teachers from i3 STW Project schools to share a common experience by observing how a STW school had translated the STW Rubric's "best" practices into "everyday" practices. Aside from observations, teachers were able to ask questions, receive resources to facilitate the implementation of these practices, and network with other teachers. These interactive visits were powerful as they allowed for a better understanding not only of the best practices described in the rubric's criteria, but the difference between implementing with "Good Quality" and "High Quality." Schools used a "train the trainer" approach as teachers returning from a school visit provided professional development on what they had learned and disseminated resources, or piloted a new practice that once implemented successfully (perhaps with some adaptations) could be replicated throughout the school with the pilot teachers providing technical assistance as the in-house experts.

Initially value was added by visiting STW schools with similar demographics as it dispelled the myth that student behavior precluded the i3 STW Project school's ability to be successful. This evolved as project schools acknowledged the value of visiting any STW school, regardless of their socioeconomic status, since the purpose of the visit was gaining knowledge on how the STW school had been successful, and discerning what practices the project school could adopt or modify that would enable their school to be successful. The visits encouraged schools that they were already doing some best practices and once they began consistently implementing additional best practices and seeing positive results, their confidence increased.

### ***The Value of Participating in a Network of Schools***

i3 STW Project schools valued the knowledge, resources, and support gained through networking opportunities and professional development provided at the state and national levels through the Forum's STW Network. For teachers, this meant increased opportunities to connect and learn from other teachers, to share ideas and resources, and to be inspired to try new ideas in the classroom based on shared lessons learned. Principals emphasized the benefits afforded to them from getting together with other i3 STW Project principals to collaborate and share experiences and challenges. Project schools now attend state and national middle-level conferences, as well as the STW Conference to receive additional training. Rural schools especially valued these learning opportunities and viewed them as a lifeline from their rural isolation. Finally, schools began advancing their leadership skills by presenting on experiences and lessons learned from their participation in the i3 STW Project at state and national conferences.

### ***Cultivating School District Involvement and Support***

State hub leaders, coaches, principal mentors, and principals experienced a variety of challenges throughout the project when collaborating with school districts. Most notably were the district's lack of

understanding about the i3 STW Project and the disruption of relationships and support that occurred with unexpected turnover of district personnel. The lesson learned from these challenges is the importance of cultivating school district involvement and support of the project from the beginning and maintaining it throughout the project. Project leaders agreed that in retrospect, they would have done a careful evaluation of potential school district participation in the project which involved a review of their capacity to support the grant and the commitment of their superintendent. They also felt they should have explained and documented the i3 STW Project, the services to be provided to the school(s) and the district, as well as the expectations and role of the district in supporting schools.

It was also important to have ongoing collaborative relationships with key district personnel throughout the project. By encouraging their active attendance and participation project meetings, as well as the state and STW conferences, it assists in supporting the schools. In particular, it was important that the state hubs cultivated district understanding of the i3 STW Project by facilitating training on the STW rubric and participation in STW visits. Principals believed that through a better understanding of STW, the school district would recognize that the rubric captured all of the district's goals and vision but utilized different language. Finally, alignment of the current school district requirements and reporting with the STW Rubric criteria should be undertaken to illustrate the common goals.

### ***Project Oversight and Management***

Oversight and management of a multi-faceted project such as the i3 STW Project required the development and adoption of communication strategies, collaboration protocols, and methods to oversee progress, identify challenges, and make mid-course corrections as needed. The first key lesson learned was that the state hub teams started working immediately and independently after the grant was awarded without first meeting as a national leadership team (i.e., the three state teams, the National Forum, and CPRD) to discuss, clarify, and define each programmatic component and how to best to implement them. Despite the shared goals and the vision of the logic model, each state team initiated the project with an approach that best fit their state's needs resulting in variance in implementation. In hindsight, the national team should have taken time to strategize on how to best "rollout" the project such as: defining roles and expectations for coaches and principal mentors; providing training; introducing the i3 STW Project to the school district and securing their support formally through MOUs; scheduling time to build relationships with schools and address issues of school culture; and establishing protocols for relationships with mentor schools, etc.

The second lesson about oversight and project management was the importance of recruiting and hiring experienced and knowledgeable coaches and principal mentors to work with project schools. Through trial and error, the state hub leaders came to recognize that coaches and mentors needed to be experienced in STW, knowledgeable about middle-grades best practices, and have the dispositions and mindsets for the coaching/facilitating/mentoring role. It was also important to consider the assignments of coaches and mentors to project schools in terms of the right fit for the environment and needs of each school. Finally, the state teams learned that if the fit between coach and school is not working, it is best to reassess and reassign sooner rather than later, so that the work is not stalled.

The last key lesson about project oversight and management is the critical importance of the national team engaging in regular communication through bi-monthly phone calls, face-to-face quarterly

meetings, and frequent emails in order to both monitor progress, address challenges, and to minimize variance in project implementation. This formalized communication became the real strength of the i3 STW Project, allowing the national team to collaborate and support each other through discussions of project-level data, as well as state-by-state sharing of implementation experiences, challenges, and successful strategies, which were then communicated to schools and everyone working with schools, making it possible for the project to work across multiple states.

## List of Project Reports, Papers, Presentations, and Products

The following is a list of selected reports, papers, presentations, and products on the i3 STW Project that have been generated throughout the project period. Many are available for download from the National Forum to Accelerate Middle-Grades Reform's website at [www.middlegradesforum.org](http://www.middlegradesforum.org).

### *i3 STW Project Reports*

2011 i3 STW Coaching Focus Groups: Summary of Questions Asked  
February 2012

Year 2 Focus Group Findings on Change: i3 Project STW School Transformation Network  
September 2012

Year 2 Focus Groups: Principal Mentor Findings, i3 Project STW School Transformation Network  
October 2012

Year 2 Focus Groups: STW Coach Findings, i3 Project STW School Transformation Network  
December 2012

Year 2 Focus Groups: Principal Findings, i3 Project STW School Transformation Network  
January 2013

i3 Project: Schools to Watch Transformation Network Fact Sheet  
June 2013

Lessons Learned for Sustainability: 2013 Focus Group Findings from the i3 Project STW School Transformation Network  
September 2013

2014 i3 STW Project: National Aggregate Longitudinal STW Rubric Data Report  
June 2013

2014 i3 STW Project: National Aggregate Longitudinal Self-Study Survey Data Notebook  
June 2013

The Impact of the i3 Project: School Visit Focus Group Findings  
September 2014

California Coach's Log Report – Year 4  
September 2014

Illinois Coach's Log Report – Year 4  
September 2014

North Carolina Coach's Log Report – Year 4  
September 2014

Reflections on the i3 Project: Year 4 Focus Group Findings  
May 2015

### ***i3 STW Project Papers***

Turning Around Low-Performing Middle-Grades Schools: Emerging Research from a Nationally Funded Investing in Innovation (i3) Project  
Paper presented at the Annual Meeting of the American Educational Research Association, April 2014.

Measuring the Impact of a Nationally Funded i3 Project on Middle-Grades Student Academic Performance  
Paper presented at the Annual Meeting of the American Educational Research Association, April 2015.

### ***i3 STW Project Presentations***

Schools to Watch: School Transformation Network  
Presentation at the GradNation Summit, 2011

Investing in Innovation (i3): Middle Grades Lessons from the Field  
Presentation at the annual conference of the Association for Middle Level Education, 2011

Emerging Lessons in Turning Around Low-Performing Schools: i3 STW School Transformation Network  
Presentation at the annual conference of the Association for Middle Level Education, 2012

Schools to Watch: School Transformation Network, 2010 i3 Development Grant  
Presentation at the U.S. Department of Education Briefing, May 2013

Turning Around Struggling Schools: The Proof is in the Research: i3 STW School Transformation Network  
Presentation at the annual conference of the Association for Middle Level Education, 2013

Culture Change: Building Capacity for School Improvement  
Presentation at the annual conference of the Association for Middle Level Education, 2014

Professional Learning in i3 Middle Grades Schools: Then, Now, and Next  
Presentation at the annual conference of the Learning Forward Professional Learning Association, 2014

Research on the Key Building Blocks for Struggling Schools: A Study of an i3 (Investing in Innovation) School Turnaround Project  
Presentation at the annual conference of the Association for Middle Level Education, 2015

Turning Around High-Poverty Middle Grades Schools  
Presentation at the annual conference of the Learning Forward Professional Learning Association, 2015

### ***i3 STW Project Products***

Online Coach's Log data collection, tracking, and reporting system

Online Schools to Watch (STW) Self-Study and Rating Rubric data collection and reporting system

Coaching Middle Grades Schools to Success: A Field Guide for Schools to Watch Coaches

Coaching Middle Grades Schools to Success: A Guide for Schools to Watch State Service Hubs

Coalinga Middle School: Building Bridges to Success! Through Vision, Values, Character and Achievement, Case Study

Hamlet Middle School: A Journey to Excellence, Case Study

Middle School Philosophy and the K-8 Schedule: Academic Excellence, Developmental Responsiveness, Social Equity, and Organizational Structures and Processes Within the Successful K-8 School, Case Study

From Data to Success: Using Early Warning Indicators to Shape Interventions for Students in the Middle Grades, Case Study

## References

- Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, 333(6045), 1034–1037.
- Brownstein, A. (2012). What studies say about school turnarounds. *The National Education Writers Association (EWA) Research Brief*. Retrieved from [http://www.ewa.org/sites/main/files/file-attachments/ewabrief.schoolturnarounds.final3\\_.pdf](http://www.ewa.org/sites/main/files/file-attachments/ewabrief.schoolturnarounds.final3_.pdf)
- Flowers, N., Begum, S., Carpenter, D. M. H., & Mulhall, P. F. (2015, April) *Measuring the impact of a nationally funded i3 project on middle-grades student academic performance*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Flowers, N., Begum, S., Carpenter, D. M. H., Mulhall, P. F., & Poes, M. J. (2014, April). *Turning around low performing middle-grades schools: Emerging research from a nationally funded investing in innovation (i3) project*. Paper presented at the annual meeting of the American Educational Research Association, Philadelphia, PA.
- Flowers, N., & Mertens, S. B. (2003). Professional development for middle-grades teachers: Does one size fit all? In P. G. Andrews & V. Anfara, Jr. (Eds.), *Leaders for a movement: professional preparation and development of middle level teachers and administrators* (pp. 145-160). Greenwich, CT: Information Age Publishing.
- Flowers, N., Mertens, S. B., & Mulhall, P. F. (1999). The impact of teaming: Five research-based outcomes of teaming. *Middle School Journal*, 31(2), 57-60.
- Flowers, N., Mertens, S. B., & Mulhall, P. F. (2000a). What makes interdisciplinary teams effective? *Middle School Journal*, 31(4), 53-56.
- Flowers, N., Mertens, S. B., & Mulhall, P. F. (2000b). How teaming influences classroom practices. *Middle School Journal*, 32(2), 52-59.
- Flowers, N., Mertens, S. B., & Mulhall, P. F. (2002). Four important lessons about teacher professional development. *Middle School Journal*, 33(5), 57-61.
- Flowers, N., Mertens, S. B., & Mulhall, P. F. (2003). Lessons learned from more than a decade of middle grades research. *Middle School Journal*, 35(2), 55-59.
- Flowers, N., Mertens, S.B., Mulhall, P.F. (2007). *Applying current middle grades research to improve classrooms and schools*. Westerville, OH: National Middle School Association.
- Henry, D. L., Nistor, N., & Baltés, B. (2014). Examining the relationship between math scores and English language proficiency. *Journal of Educational Research and Practice*, 4(1), 11-29.

- Larwin, K. H. (2010). Reading is fundamental in predicting math achievement in 10th graders. *International Electronic Journal of Mathematics Education*, 5(3), 131-145. Retrieved from [http://www.mathedujournal.com/dosyalar/IJEM\\_v5n3\\_2.pdf](http://www.mathedujournal.com/dosyalar/IJEM_v5n3_2.pdf)
- Mertens, S. B., & Flowers, N. (2003). Middle school practices improve student achievement in high poverty schools. *Middle School Journal*, 35(1), 33-43.
- Mertens, S. B., & Flowers, N. (2006). Middle Start's impact on comprehensive middle school reform. *Middle Grades Research Journal*, 1(1), 1-26.
- Mertens, S. B., Flowers, N., Hesson-McInnis, M., & Bishop, C. (2006, April). *Middle grades teaching practices: A validation study*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Mertens, S. B., Flowers, N., Hesson-McInnis, M., & Bishop, C. (2007, April). *The impact of middle grade structures, organizations, and practices on student outcomes: Moving toward a multivariate/multi-level model*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Roderick, M., (2006). *Closing the Aspirations-Attainment Gap: Implications for High School Reform. A Commentary from Chicago*. New York, NY: MDRC. Retrieved from <http://files.eric.ed.gov/fulltext/ED491184.pdf>
- What Works Clearinghouse. (2014). Procedures and standards handbook version 3.0. Retrieved from: [http://ies.ed.gov/ncee/wwc/pdf/reference\\_resources/wwc\\_procedures\\_v3\\_0\\_standards\\_handbook.pdf](http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf)

## Appendix A. Psychometric Properties of the Self-Study Survey Constructs

### School Improvement Self-Study Teacher Survey Constructs

Teacher Survey Constructs	Definition	Response Metric	Cronbach's Alpha
Team Practices	Four composite scores using 22 survey items to measure the activities of core academic subject teachers working together on teams, including: planning and coordination; curriculum integration; coordination of student assignments/assessments; and contact with parents.	Seven-point metric (1=never; 2=once a year; 3=several times a year; 4=quarterly; 5=monthly; 6=weekly; and 7=daily).	$\alpha = .89$
Quality of Team Interactions	Four composite scores using 19 survey items to measure effective strategies for core academic teachers working together in a team setting to improve student success, including: cohesion and harmony, readiness for teaming, addressing student needs, and having consistently high expectations for all students.	Five-point metric (1=strongly disagree; 2=disagree; 3=neither; 4=agree; and 5=strongly agree).	$\alpha = .89$
Team Decision Making	Three composite scores using 14 survey items measuring how much decision-making authority interdisciplinary teams (core academic teachers) have regarding team practices, school-wide practices, and student performance and assessment.	Five-point metric (1=very little; 2=little; 3=moderate amount; 4=much; and 5=very much).	$\alpha = .94$
Work Climate	Two composite scores using 8 survey items to measure work climate, including teachers' commitment to their school and teachers' feelings that they are recognized for their contributions by the school administration.	Five-point metric (1= Never; 2= hardly ever; 3= sometimes; 4= most of the time; and 5= always).	$\alpha = .76$
Collective Teacher Efficacy	Single composite score of 5 survey items that measures the collective commitment teachers have to the success of their students.	Five-point metric (1=strongly disagree; 2=disagree; 3=neither; 4=agree; and 5=strongly agree).	$\alpha = .93$
Teacher Decision Making	Two composite scores using 6 survey items measuring how much teachers participate in decisions and how often teachers experience autonomy in making decisions.	Five-point metric (1=strongly disagree; 2=disagree; 3=neither; 4=agree; and 5=strongly agree).	$\alpha = .79$
Administrative Leadership	Single composite score of 6 survey items measuring teachers' perceptions of the school administrator's leadership practices (e.g., communication, problem-solving strategies).	Five-point metric (1=strongly disagree; 2=disagree; 3=neither; 4=agree; and 5=strongly agree).	$\alpha = .96$
Classroom Instructional Practices	Nine composite scores using a total of 49 survey items to measure the frequency that core academic teachers engage in "best" instructional practices in their classrooms, including: small group active instruction; integration and interdisciplinary practices; authentic instruction and assessment; citizenship and social competence instruction; critical thinking practices; mathematical skill practices; reading skill practices; writing skill practices; and listening and verbal skill practices.	Seven-point metric (1=never; 2=several times a year; 3=monthly; 4=several times a month; 5=weekly; 6=several times a week; and 7=daily).	$\alpha = .89$



## School Improvement Self-Study Student Survey Constructs

Student Survey Constructs	Definition	Response Metric	Cronbach's Alpha
Academic Efficacy	Single composite score of 5 survey items that measures students' belief in their ability to be academically successful if they are willing to work hard.	Four-point metric (1=strongly disagree; 2=disagree; 3=agree and 4=strongly agree).	$\alpha = .82$
Academic Expectations	Single composite score of 9 survey items that measures students' short and long-term academic expectations for themselves, as well as their perceptions of the academic expectations that their parents and teachers have for them.	Five-point metric (1=definitely won't; 2=probably won't; 3=might; 4=probably will; and 5=definitely will).	$\alpha = .87$
Family Involvement	Single composite score of 9 survey items to measure students' reports of how often their family discusses the importance of school with them, checks or helps them with homework, or motivates them to do well at school.	Four-point metric (1=never; 2=hardly ever; 3=sometimes; and 4=often)	$\alpha = .85$
Belonging	Single composite score of 7 survey items that measures students' feelings of belonging and connectedness to their school environment.	Five-point metric (1=never; 2=hardly ever; 3=sometimes; 4=most of the time; and 5=always).	$\alpha = .78$
School Climate	Five composite scores using 19 survey items to measure school climate, including students' ratings of whether teachers are supportive, expectations and rules are clear, teachers engage students in learning, student interactions are negative, and whether discipline is harsh.	Five-point frequency metric (1=never; 2=hardly ever; 3=sometimes; 4=most of the time; 5=always).	$\alpha = .77$

## Appendix B. Fidelity Matrix for National Forum’s STW School Transformation Network Project

	Key Elements of Component	Operational Definition for Indicator	Data Source(s) for Measuring Indicator	Data Collection Schedule for Obtaining Data on Indicator	Raw Scoring and Scoring Criteria for Fidelity				
					Each indicator will be scored and given an adequate threshold at only one level				
					Indicator Score (Teacher)	Implementation Score	Indicator Score Range (School)	Implementation Score	Fidelity Score
<b>Component 1: Create a powerful vision for high performance using the STW criteria</b>									
1	Planning meeting for schools	Planning meeting each year for school and district representatives from each school, as well as coaches to: 1) Provide a project overview for the year, 2) Go over the STW criteria, 3) Go over responsibilities of each partner, 4) Introduce the support team, and 5) Engage in networking.	Coach’s log	Ongoing entry by coaches into coach’s log			0-2	0 (low)=not attended 1 (med)=school reps attended 2 (high)=school reps and district reps attended	
2	Teachers complete the self-rating STW Rubric	The self-rating STW Rubric is a tool for continuing improvement. The rubric is divided into 4 sections depicting the STW criteria. Each section contains detailed examples of excellence. Teachers rate the implementation annually.	Online STW Rubric data collection system	Annual administration of the online STW Rubric			0-2	0 (low)=<50% staff completed 1 (med)=50-75% staff completed 2 (high)=76%+ staff completed	
3	Review the data results from the STW Rubric with school leadership team	A report on the results of the self-rating STW Rubric is generated for each school. The coaches/state team reviews the results with the leadership team for future use in assessing needs and developing improvement goals.	Coach’s log	Ongoing entry by coaches into coach’s log			0-6	0 (no)=not completed 6 (yes)=completed	
4	Faculty supports the vision for the school	The degree to which the faculty supports the vision of the school via their response to a question on the Self-Study Staff Survey (I support the vision/goals/practices of this project).	Self-Study Staff Survey	Annual administration of the Self-Study Staff Survey	1=Strongly disagree 2=Disagree 3=Neither 4=Agree 5=Strongly agree	0=does not support vision (raw scores of 1, 2, 3) 1=support vision (raw scores of 4, 5)	0-6	0 (low)=<50% of staff support vision 3 (med)=50-75% of staff support vision 6 (high)=76%+ of staff support vision	
<b>Component Level Score</b>							0-16	0 (low)=0-8 1 (med)=9-12 2 (high)=13-16	Implementation of component with fidelity= At least 70% of schools: Years 1-3 – with high implementation (a score of 2), Year 4 – with med or high implementation (a score of 1 or 2)

	Key Elements of Component	Operational Definition for Indicator	Data Source(s) for Measuring Indicator	Data Collection Schedule for Obtaining Data on Indicator	Raw Scoring and Scoring Criteria for Fidelity		
					Each indicator will be scored and given an adequate threshold at only one level		
					Indicator Score Range (School)	Implementation Score	Fidelity Score
<b>Component 2: Engage in an in-depth assessment and planning process using the STW criteria</b>							
1	Identify needs with school leadership team	The STW coach works with the leadership team to examine data (student outcome data, early indicators, self-rating STW Rubric), and reflect on the STW criteria in order to identify the areas for improvement.	Coach's log	Ongoing entry by coaches into coach's log	0-1	0 (no)=not completed 1 (yes)=completed	
2	Develop clear and measurable goals with school leadership team	The STW coach works with the leadership team to develop SMART (Strategic, Measurable, Achievable, Realistic, and Time-bound) goals for school improvement.	Coach's log	Ongoing entry by coaches into coach's log	0-1	0 (no)=not completed 1 (yes)=completed	
3	Implement a coherent and integrated action plan with performance benchmarks with school leadership team and faculty	The STW coach works with the leadership team and faculty to implement an integrated action plan that outlines concrete steps/strategies to accomplish school improvement goals. Performance benchmarks are included in the action plan in order to assess progress, monitor results, and make any mid-course corrections as needed.	Coach's log	Ongoing entry by coaches into coach's log	0-4	0 (no)=not completed 4 (yes)=completed	
<b>Component Level Score</b>					0-6	0 (low)=0 1 (med)=1-2 2 (high)=4-6	Implementation of component with fidelity = At least 80% of schools with high implementation (a score of 2)

	Key Elements of Component	Operational Definition for Indicator	Data Source(s) for Measuring Indicator	Data Collection Schedule for Obtaining Data on Indicator	Raw Scoring and Scoring Criteria for Fidelity		
					Each indicator will be scored and given an adequate threshold at only one level		
					Indicator Score Range (School)	Implementation Score	Fidelity Score
<b>Component 3: STW coach</b>							
1	Planning meeting for STW coaches	Planning meeting occurs at the beginning of each year or when a new coach is hired. Meeting includes project overview, STW criteria, and coaching activities and expectations.	State hub activity reports	Quarterly submissions to project director	0-1	0 (no)=not completed 1 (yes)=completed	
2	STW coaches work with schools	Each year, STW coaches visit the school 2 days/month for 9 months. (Notes: 1) coaches spend additional time planning for visits. In Illinois, they are paid 7 hours per month to plan. In California and North Carolina, they are not paid for planning; 2) coaching in year 1 did not begin until Jan, so in year 1 coaches visit the school 2 days/month for 5 months.)	Coach's log	Ongoing entry by coaches into coach's log	0-6	<i>Year 1:</i> 0 (low)=0-4 visits 3 (med)=5-8 visits 6 (high)=9 or more visits  <i>Years 2-4:</i> 0 (low)=0-9 visits per year 3 (med)=10-15 visits per year 6 (high)=16 or more visits per year	
3	STW coaches participate in ongoing training	Ongoing training includes participation in coach's conference calls, state meetings, and attendance at STW conferences.	State hub activity reports	Quarterly submissions to project director	0-1	0 (no)=not completed 1 (yes)=completed	
<b>Component Level Score</b>					0-8	0 (low)=0-2 1 (med)=3-5 2 (high)=6-8	Implementation of component with fidelity = At least 70% of schools with high implementation (a score of 2)

	Key Elements of Component	Operational Definition for Indicator	Data Source(s) for Measuring Indicator	Data Collection Schedule for Obtaining Data on Indicator	Raw Scoring and Scoring Criteria for Fidelity		
					Each indicator will be scored and given an adequate threshold at only one level		
					Indicator Score Range (School)	Implementation Score	Fidelity Score
<b>Component 4: Principal mentor</b>							
1	Planning meeting for principal mentors	Planning meeting occurs at the beginning of each year or when a new mentor is hired. Meeting includes project overview, STW criteria, and mentoring activities and expectations.	State hub activity reports	Quarterly submissions to project director	0-1	0 (no)=not completed 1 (yes)=completed	
2	Principal mentors work with principals	Each year, principal mentors contact the school one time/month for 9 months. Contacts include visits, phone consultation, and/or email communication. (Notes: 1) In Illinois and North Carolina, contacts are visits. Due to large geographic distances in California, mentors work with schools via a combination of visits, phone, and email; 2) mentoring in year 1 did not begin until Jan, so in year 1 mentors contact the school one time/month for 5 months.)	Coach's log	Ongoing entry by coaches into coach's log	0-6	<i>Year 1:</i> 0 (low)=0-2 contacts 3 (med)=3-4 contacts 6 (high)=5 or more contacts  <i>Years 2-4:</i> 0 (low)=0-4 contacts per year 3 (med)=5-8 contacts per year 6 (high)=9 or more contacts per year	
3	Principal mentors participate in ongoing training	Ongoing training includes participation in coach's conference calls and meetings.	State hub activity reports	Quarterly submissions to project director	0-1	0 (no)=not completed 1 (yes)=completed	
<b>Component Level Score</b>					0-8	0 (low)=0-2 1 (med)=3-5 2 (high)=6-8	Implementation of component with fidelity = At least 70% of schools with high implementation (a score of 2)

	Key Elements of Component	Operational Definition for Indicator	Data Source(s) for Measuring Indicator	Data Collection Schedule for Obtaining Data on Indicator	Raw Scoring and Scoring Criteria for Fidelity		
					Each indicator will be scored and given an adequate threshold at only one level		
					Indicator Score Range (School)	Implementation Score	Fidelity Score
<b>Component 5: Mentor school</b>							
1	Mentors schools* work with school faculty and/or leadership team	Mentor schools serve as role models, hosting visits from faculty/leadership teams, providing observation opportunities (e.g., grade level meetings, team meetings, classrooms), discussion opportunities, and share resources. Visits occur at least 2 times per year. (Note: Due to the large size of schools in California and North Carolina, leadership teams visit mentor schools and present findings to faculty upon return – training of trainers model. In Illinois, smaller schools allow all teachers to visit mentor schools.)	State hub activity reports	Quarterly submissions to project director	0-2	0 (low)=0 visit 1 (med)=1 visit 2 (high)=2 or more visits	
Component Level Score					0-2	0 (low)=0 1 (med)=1 2 (high)=2	Implementation of component with fidelity = At least 70% of schools with med or high implementation (a score of 1 or 2)

\*Note: In Illinois, close geographic proximity allows one-to-one assignments of mentor schools to project schools. In California and North Carolina, a cadre of 3 to 8 schools serve as mentors, chosen based on the reform issue needed.)

	Key Elements of Component	Operational Definition for Indicator	Data Source(s) for Measuring Indicator	Data Collection Schedule for Obtaining Data on Indicator	Raw Scoring and Scoring Criteria for Fidelity		
					Each indicator will be scored and given an adequate threshold at only one level		
					Indicator Score Range (School)	Implementation Score	Fidelity Score
<b>Component 6: Implement early indicators program</b>							
1	Preliminary training on early indicators for schools	Early indicators is a program schools use to compile data about student progress in order to identify students “off track” & in need of assistance. The preliminary training is to understand the program and learn implementation approaches. At least 1 person from school attends in yr 1.	State hub activity reports	Quarterly submissions to project director	0-1	Year 1: 0 (no)=not attended 1 (yes)=attended  Years 2-4: NA	
2	Comprehensive tailored training on early indicators for schools	Schools participate in comprehensive training on early indicators that is tailored to their school. This may occur for each individual school or for a cadre of schools in a district. At least 1 person from school attends in yr 2.	State hub activity reports	Quarterly submissions to project director	0-1	Year 1, 3, and 4: NA  Year 2: 0 (no)=not attended 1 (yes)=attended	
3	Implement a process for regularly compiling and examining student data	STW coach assists schools to implement a process for compiling the data on each student and a structure for regularly examining the data.	State hub activity reports	Quarterly submissions to project director	0-1	0 (no)=not completed 1 (yes)=completed	
4	Implement a process for regularly identifying students who need additional support	STW coach assists schools to implement a process for regularly identifying students who are “off track” and assessing root causes for being off track.	State hub activity reports	Quarterly submission to project director	0-4	0 (no)=not completed 4 (yes)=completed	
5	Schools implement a 3-tiered, progressive intervention process to assist students who are “off track”	The 3-tiered levels of intervention include: Tier 1) Preventive strategies for an entire school; Tier 2) Targeted strategies for students who need extra support; and Tier 3) Intensive interventions for students who require one-to-one support.	State hub activity reports	Quarterly submissions to project director	0-4	0 (low)=no implementation 2 (med)=Implemented Tier 1 & 2 4 (high)=Implemented all 3 Tiers	

Component Level Score					Years 1-2: 0-10	Years 1-2: 0 (low)=0-2 1 (med)=3-8	Years 3-4: 0 (low)=0-1 1 (med)=2-8	Implementation of component with fidelity = At least 60% of schools with high implementation (a score of 2)
Key Elements of Component	Operational Definition for Indicator	Data Source(s) for Measuring Indicator	Data Collection Schedule for Obtaining Data on Indicator	Raw Scoring and Scoring Criteria for Fidelity				
				Each indicator will be scored and given an adequate threshold at only one level				
				Indicator Score Range (School)	Implementation Score	Fidelity Score		
<b>Component 7: Participate in national and state STW Network</b>								
1	School accesses peer support from the broader national and state-level STW network	Peer support from the STW network includes opportunities such as: professional development on mathematics, literacy, or other middle-grades topics; annual STW national conference, state-level STW events; and a variety of online discussion groups and other experiences sponsored by the Forum and its partners. At least 1 person from the school attends at least two events.	State hub activity reports	Quarterly submissions to project director	0-2	0 (low)=0 events attended 1 (med)=1 event attended 2 (high)=2 or more events attended		
Component Level Score					0-2	0 (low)=0 1 (med)=1 2 (high)=2	Implementation of component with fidelity = At least 70% of schools with high implementation (a score of 2)	



	Key Elements of Component	Operational Definition for Indicator	Data Source(s) for Measuring Indicator	Data Collection Schedule for Obtaining Data on Indicator	Raw Scoring and Scoring Criteria for Fidelity		
					Each indicator will be scored and given an adequate threshold at only one level		
					Indicator Score Range (School)	Implementation Score	Fidelity Score
<b>Component 8: Implement professional learning communities (PLCs)</b>							
1	Schools create professional learning community (PLC) structures for teachers	PLCs are collegial groups of teachers and/or other personnel (i.e., administrator, curriculum director) who work together regularly to analyze student needs, identify evidence-based practices, and reflect on results.	State hub activity reports	Quarterly submissions to project director	0-1	0 (no)=not completed 1 (yes)=completed	
Component Level Score					0-1	0 (low)=0 1 (high)=1	Implementation of component with fidelity = At least 60% of schools with high implementation ( a score of 1)

	Key Elements of Component	Operational Definition for Indicator	Data Source(s) for Measuring Indicator	Data Collection Schedule for Obtaining Data on Indicator	Raw Scoring and Scoring Criteria for Fidelity			
					Each indicator will be scored and given an adequate threshold at only one level			
					Indicator Score Range (School)	Implementation Score	Fidelity Score	
<b>Component 9: Focused professional development designed to build a learning community and addresses the needs of students who need additional support</b>								
1	Schools participate in professional development on PLC structures and functioning	Each school participates in at least 1 professional development session per year on PLC structures or functioning (i.e., analyzing achievement data, examining teacher work). <sup>*</sup> At least 1 person from the school attends.	State hub activity reports	Quarterly submissions to project director	0-2	0 (no)=not completed 2 (yes)=completed		
2	Schools participate in professional development on meeting the needs of those students who are most at risk	Each school participates in at least 1 professional development session per year on meeting the needs of at-risk students. <sup>*</sup> At least 1 person from the school attends.	State hub activity reports	Quarterly submissions to project director	0-2	0 (no)=not completed 2 (yes)=completed		
3	Schools participate in other professional development focused on their identified needs	STW coaches and state teams provide/arrange for professional development as needs arise, such as using universal design principles, differentiated instruction, literacy, teacher high level math to ELLs, technology, and creating a personalized learning environment. A minimum of 2 sessions per year. <sup>*</sup> At least 1 person from the school attends.	State hub activity reports	Quarterly submission to project director	0-2	0 (low)=0 sessions 1 (med)=1 session 2 (high)=2 or more sessions		
Component Level Score					0-6	Years 1-3: 0 (low)=0-2 1 (med)=3-5 2 (high)=6	Year 4: 0 (low)=0-2 1 (med)=3 2 (high)=4-6	Implementation of component with fidelity = At least 80% of schools with high implementation (a score of 2)

<sup>\*</sup>Note: Due to the large size of the schools in California and North Carolina, for some professional development activities, leadership teams participate and then present the information to the full faculty upon return (i.e., training of trainers model). In Illinois, smaller schools allow all teachers to participate in professional development activities.