

**I ILLINOIS**

CPRD | Center for Prevention  
Research & Development

SCHOOL OF SOCIAL WORK

# Power of Two Evaluation Report

March 2020



This work was supported by the U.S. Department of Education's Supporting Effective Educator Development program through a grant to the National Forum to Accelerate Middle-Grades Reform. The opinions expressed are those of the authors and do not represent the views of the U.S. Department of Education.

| 510 Devonshire Dr., Champaign, IL 61820 | (217) 333-3231 | website: [publish.illinois.edu/cprd](http://publish.illinois.edu/cprd) |

# CONTENTS

**Acknowledgements ..... 2**

**The Power of Two ..... 3**

**Evaluation Design..... 9**

**Coach Findings ..... 15**

**Student Findings ..... 21**

**Teacher Findings..... 33**

**Summary ..... 39**

**Appendices..... 42**

# Acknowledgements

**Power of Two Evaluation Report**

**March 2020**

**Produced for:**

**National Forum to Accelerate Middle-Grades Reform**

**P.O. Box 224**

**Savoy, IL 61874**

[middlegradesforum.org](http://middlegradesforum.org)

**Power of Two Project Leaders:**

**California - Aneta Murphy, Peter Murphy, Theresa Hancock**

**Illinois - Vicki Mogil**

**Michigan - Blake Pratt, Molly Sholten, Sherry Schock**

**North Carolina - Ran Barnes, Betty Terrell, James Davis**

**National Forum - Ericka Uskali, Deborah Kasak**

**Report Prepared by:**

**Center for Prevention Research and Development**

**School of Social Work**

**University of Illinois**

**510 Devonshire Drive**

**Champaign, IL 61820**

[publish.illinois.edu/cprd](http://publish.illinois.edu/cprd)

**Nancy Flowers, Director of Research Programs**

**Dawn M. H. Carpenter, Senior Research Program Coordinator**

**Shahana Begum, Research Data Coordinator**

## Introduction

The Power of Two is an initiative of the National Forum to Accelerate Middle-Grades Reform. Funded by a U.S. Department of Education Supporting Effective Educator Development (SEED) grant in 2015, the Power of Two provided middle-grades English/language arts and mathematics teachers with professional development and coaching to implement targeted daily warm-ups with students during a year-long paired intervention. The three year project was implemented in California by the California League of Schools, in Illinois by the Association of Illinois Middle-Grade Schools, in Michigan by the Institute for Excellence in Education, and in North Carolina by the North Carolina Association for Middle Level Education. The Center for Prevention Research and Development (CPRD) in the School of Social Work at the University of Illinois served as the evaluation partner for the Power of Two Project.

## Intervention Description

Students engaged in the Power of Two participated in the **Academic Language Development (ALD) intervention in their English/language arts classroom** and in the Focused Instructional Model (FIM) in their mathematics classroom.

The instructional anchor of ALD is the **Academic Vocabulary Toolkit** by Dr. Kate Kinsella with Theresa Hancock. The *Academic Vocabulary Toolkit* is a supplemental curriculum for English/language arts classrooms focusing upon high-yield words (e.g., analysis, characteristic, demonstrate, generalize, impact, precede, subsequently) students must readily comprehend and adeptly utilize across the subject areas. The toolkit is divided into eight units that address a crucial competency required for advanced literacy tasks and skillful communication. The eight consistent units emerged from a detailed analysis of the shifts in standards for literacy and language instruction and performance-based assessments. Students need a practical toolkit of words at their disposal to describe, sequence, interpret, create, analyze, make inferences, argue, and discuss cause and effect. At each grade level, the *Academic Vocabulary Toolkit* provides students with a new set of words to expand

ALD has provided a way for me to target specific skills that the majority of my students lack. The greatest impact it has made has been the ability of the routine to provide structure for our class. ALD has also encouraged me to look deeper into what is required to build students' knowledge and willingness to learn. I believe that student success in the ALD class required a growth mindset "buy-in" from all of the students in order to get the most impact. –ELA Teacher

# The Power of Two

their academic vocabulary. Each toolkit word is introduced and practiced in tandem with a few priority word partners (i.e., words and phrases commonly used) with the focus word in the academic speech and writing. Students learn pronunciation and meaning and how the word is applied both verbally and in written form. Through the five-day instructional sequence (20-25 minutes per day), students deepen their word knowledge by analyzing and applying the high-utility academic words in multiple contexts.

The FIM intervention for mathematics classrooms was developed by the Institute for Excellence in Education. FIM is an integrative, holistic process of daily practice that allows teachers to monitor their

I have enjoyed implementing FIM in my classroom and I believe that it has been of great benefit to my students. I firmly believe that the repetition provided by FIM is responsible for their greater retention of key math standards. FIM has made me more aware of the need to continually stress that students must show effort. Students who have been stopped by past failures are now learning that it's safe to "try." –Math Teacher

students' mastery of content and provides regular interventions based on student needs and deficiencies. Daily formative assessments, administered during the first 10-15 minutes of mathematics class, quickly inform teacher decision-making. Students use metacognition to own their learning and find their own opportunities for growth and sustained success. Teachers provide comprehension strategies for students by solving two daily math problems using a think-aloud strategy with both verbal and written explanations. Teachers also review the

data collected through the daily warm-ups after class and use it to make instructional decisions for future classroom instruction.

ALD and FIM, implemented together comprised the Power of Two Project.

ALD and FIM teachers received professional development in their content areas and ongoing one-to-one coaching to apply the key ALD and FIM strategies and instructional routines in their classrooms. ALD teachers received a teaching guide that included detailed planning and pacing guides, clearly explicated instructional routines, explicit grammar lessons targeting high-frequency errors, reproducible templates, formative and summative assessments, and an instructional DVD with every lesson phase modeled by the ALD authors. FIM teachers engaged in professional learning experiences that deepened their content and pedagogical expertise specific to the standards taught at their grade level. With FIM, mathematics teachers engaged in a process to examine grade-level content standards and collaborated around putting forth solid instructional approaches for each standard. For

some teachers, the interaction among grade level mathematics teacher peers provided clarity about the content to be taught and how to effectively relay the information through FIM to students using the best teaching practices.

Through the Power of Two, both ALD and FIM teachers received individual support from a specially trained coach. Coaching supported teachers to ensure that the ALD and FIM interventions were implemented with fidelity and that teachers were mentored in their individual growth. The coaching model used a combination of observing teacher practice followed by supportive coaching conversations with the teacher about implementation, practice, routines, and student engagement.

## Expected Outcomes

The Power of Two delivered a coordinated two-subject intervention in order to improve the instructional practices of middle-grades English/language arts teachers to dramatically increase verbal and written interactions by using the ALD pedagogy, and to mathematics teachers to directly impact student mathematical learning by using the FIM warm-up. As a result, it is expected that high-need, middle-grades students will perform at higher levels and will be better prepared for success in high school, resulting in readiness for college and career.

The expected long term outcomes of the Power of Two include: 1) improved teacher practice through the use of FIM and ALD; and 2) increased student performance on English/language arts and mathematics assessments. The expected intermediate outcomes include: a) increased teacher use of effective English/language arts and mathematics strategies and routines; b) increased self-efficacy as teachers of English/language arts and mathematics; c) improved student confidence in their academic dispositions and academic behaviors; and d) increased student use of FIM and ALD strategies and routines.

## Schools

The year-long Power of Two initiative was implemented with two cohorts of middle-grades students and teachers – the first during the 2016/17 school year and the second in 2017/18. A total of 106 6<sup>th</sup> grade and 7<sup>th</sup> grade classrooms in 28 schools (Table 1) from California, Illinois, Michigan, and North Carolina participated. The Power of Two served diverse populations of students in rural, suburban, and urban locales. The initiative specifically supported English language learners (i.e., students whose primary language is not English) as well as students performing below grade level expectations.

# The Power of Two

Table 1. Power of Two Participating Schools

School Name	District	State	City
Badger Springs Middle School Landmark Middle School Mountain View Middle School Palm Middle School Sunnymead Middle School Vista Heights Middle School	Moreno Valley USD	California	Moreno Valley
Chopin Elementary School Jose de Diego Community Academy Lowell Elementary School Nobel Elementary School Reilly Elementary School Sabin Dual Language Magnet School Stowe Elementary School West Park Academy Yates Elementary School	Chicago Public Schools	Illinois	Chicago
Anna M Thurston Middle School Delton Kellogg Middle School Handy Middle School Manton Middle School Marshall Middle School Sheridan Road STEM Springport Middle School Western Middle School	St Charles Community Schools Delton Kellogg Schools Bay City Public Schools Manton Consolidated Schools Marshall Public Schools Lansing School District Springport Public Schools Bay City Public Schools	Michigan	St Charles Delton Bay City Manton Marshall Lansing Springport Auburn
Albemarle Road Middle School JM Alexander Middle School Gamewell Middle School Kennedy Middle School Northeast Middle School	Charlotte-Mecklenburg Schools Charlotte-Mecklenburg Schools Caldwell County Schools Charlotte-Mecklenburg Schools Charlotte-Mecklenburg Schools	North Carolina	Charlotte Huntersville Lenoir Charlotte Charlotte

## Service Delivery

The Power of Two initiative was designed, developed, and led by the National Forum to Accelerate Middle-Grades Reform. The Forum is a well-respected and experienced leader in transforming educational practices and strategies for teachers of students in grades five to eight. The Forum serves as a unifying voice for middle-grades improvement through more than forty key stakeholders and organizational members and its signature initiative, Schools to Watch (STW). STW provides a process

# The Power of Two

and set of criteria for identifying and recognizing schools on a sustained, upward trajectory of growth and improvement. The Forum has a national reach with STW affiliates in eighteen states whose middle-grades population represents nearly 70 percent of all young adolescents in the country. There are currently 600 Schools to Watch sites representing over 250,000 students and 16,000 teachers.

Over the years as the STW initiative expanded across the country, state leaders received countless inquiries from schools seeking turnaround assistance. The Power of Two Initiative represents the Forum's third U.S. Department of Education funded project focused on improving the middle-grades. It builds upon the experiences and lessons learned from the Forum's 2010 Investing in Innovation (i3) development grant, the *Schools to Watch: Schools Transformation Network*, and the 2013 i3 development grant, the *Middle-Grades Leadership Development* project. Collectively these grants reached 13,000 students and 900 teachers annually with on-the-ground technical assistance, coaching, networking, and intervention implementation. The grants focused on building organizational capacity using the STW criteria (coupled with analysis of student data) as a comprehensive framework incorporating self-assessment, instructional goal-setting, action planning, and evaluation.

With the Power of Two initiative, the Forum focused on the need to ensure that middle-grades students have a foundation of skills in the two critical content areas of English/language arts and mathematics that are necessary for future success in education and career. The Power of Two, therefore, was designed to produce strong evidence about effective teacher practices, instructional routines, and strategies that can be learned and replicated in other middle-grades schools through the Forum's STW network. By employing deeper interventions in the instructional areas of English/language arts and mathematics together, every student had two different "doses" of intervention for a full school year in order to more rapidly bolster performance.

The on-the-ground work with teachers at schools was provided by the Forum's affiliate organizations in each state: the California League of Schools, the Association of Illinois Middle-Grade Schools, the Institute for Excellence in Education in Michigan, and the North Carolina Association for Middle Level Education. Each state affiliate hired experienced coaches, worked with the ALD and FIM model developers to provide training for the coaches, and monitored the implementation of the Power of Two initiative at the schools in their state. Each state affiliate utilized their existing organizational structures and staffing to recruit schools and implement the Power of Two within the requirements of districts in their states. As a result, there was some variability in service delivery, including:

- [School Locations](#) – In two states, all Power of Two schools were located in a single district. Whereas in the other two states, there were multiple districts.
- [Coaches](#) – In one state, the Power of Two coaches were employees of the district, whereas in all other states, the coaches were independently hired. Coaching assignments also varied by state in terms of the number of teachers served. While half the states hired one coach per intervention, one state hired two coaches per intervention, while another state hired one coach for both interventions. Overall, between 2 to 10 teachers were assigned to each coach.

A model of meta-coaching by the ALD and FIM model developers was employed as a mechanism to assist them with facilitating consistent coaching services across states. Both the ALD and FIM model leaders engaged coaches in monthly meetings with facilitated conversations, provided guided observation opportunities, offered email and phone support, and offered follow-up training opportunities.

## Introduction

The evaluation of the Power of Two initiative employed a cluster randomized controlled trial (RCT) design that meets the What Works Clearinghouse standards of Evidence (What Works Clearinghouse, 2014). Two cohorts of middle-grades students and teachers – the first during the 2016/17 school year and the second in 2017/18 – participated in the evaluation. The formal impact study (Cohort 1), where data were collected from both schools implementing the Power of Two and from schools serving as a control group, occurred in 2016/17. The following year, 2017/18, control schools received a delayed intervention (Cohort 2) and these data were used for exploratory analyses.

The objectives of the Power of Two evaluation were to: 1) collect data to inform and guide the Power of Two implementation; 2) analyze the intermediate impact of the Power of Two on teacher and student attitudes, practices, and behavior; and 3) measure the impact of the Power of Two on student achievement growth.

## Randomized Controlled Trial (RCT)

The Forum's state affiliates in California, Illinois, Michigan, and North Carolina recruited 28 schools that met the following criteria to participate in the Power of Two: 1) public school; 2) contain the middle grades (6<sup>th</sup> or 7<sup>th</sup> grades); and 3) serve high-need student populations (e.g., students at risk of educational failure, such as students who are living in poverty, English learners, or students who are far below grade level). Once the schools were on board, the state affiliates then worked with principals to select 53 pairs of teachers (53 English/language arts teachers and 53 mathematics teachers) to participate in the Power of Two initiative. Each mathematics/English language arts pair taught the same group of students, so students in treatment classrooms received both the English/language arts and mathematics intervention from the same pair of teachers.

Through simple cluster random assignment, the pairs of teachers within each state were randomly assigned to treatment or control. During the impact study period (2016/17 school year), treatment classrooms implemented the Power of Two and collected data for the evaluation; control classrooms collected the same data as treatment classrooms, but they did not receive the Power of Two. To incentivize the control schools to participate in the impact study, they received a delayed intervention of the Power of Two initiative during the subsequent 2017/18 school year.

# Evaluation Design

## Data Collection Tools

Data collection methods for the evaluation of the Power of Two initiative included both quantitative and qualitative components:

- [Coach's Log](#) – Coaching activities were tracked at treatment schools using CPRD's *Online Coach's Log*, which documented the coaching work (e.g., quantity of time, techniques, reflection of needs, resources used, successful outcomes, and assessments of barriers). Each Power of Two coach completed the log throughout the Power of Two initiative.
- [Classroom Observations](#) – The implementation of the English/language arts and math teaching practices were assessed with classroom observational ratings of instruction and student engagement. The ALD authors gave permission to the evaluation team for their observation tool to be used. For FIM, the evaluation team worked with FIM authors to develop an observation tool. Observations were conducted at treatment and control schools twice each year (beginning and end).
- [Teacher and Student Surveys](#) - Intermediate outcomes for students (academic efficacy, academic expectations, and engagement) and teachers (teacher efficacy, buy-in, confidence, and teaching practices) were measured prior to treatment and at the end of treatment in both treatment and control classrooms by CPRD's *School Improvement Self-Study Student and Teacher Surveys*. The constructs on the surveys have been demonstrated in prior research to be correlated with increases in student achievement (Flowers, Begum, Carpenter, & Mulhall, 2015; Flowers, Begum, Carpenter, Mulhall, & Poes, 2014; Flowers, Mertens, & Mulhall, 2003; Mertens & Flowers, 2004, 2006).
- [NWEA Measures of Academic Progress \(MAP\)](#) – The tool for assessing the impact on student achievement at treatment versus control classrooms, was measured with NWEA's MAP reading and mathematics assessments. The MAP assessments are rigorous and comparable across schools. MAP is used throughout the country and is designed to measure growth, project proficiency on high-stakes tests, and inform how educators differentiate instruction, evaluate programs, and structure curriculum. The assessments have been tested to ensure test reliability, validity, and fairness across all populations tested (Wang, Jiao, & Zhang, 2013; Wang, McCall, Jiao, & Harris, 2013; Zhang, Lau, & Wang, 2013). Students in all treatment and

control classrooms took the NWEA MAP assessments three times each year (fall, winter, and spring).

## Sample and Attrition Rates

At the start of the Power of Two Initiative, a total of 53 teacher pairs (i.e., 53 ELA teachers and 53 math teachers) were randomized within each state into either the treatment group or the control group. Several teachers dropped out of the initiative either before the project began or during the implementation period. The final sample included 40 English/language arts teachers and 43 math teachers who taught 1,095 students (Table 2).

Table 2. Power of Two Randomization and Attrition

	Randomization	Teachers Dropped	Final Sample
<b>California</b>	28 teachers (14 ELA and 14 math)	3 teachers (1 ELA and 2 math)	25 teachers (13 ELA and 12 math)
<b>Illinois</b>	26 teachers (13 ELA and 13 math)	8 teachers (5 ELA and 3 math)	18 teachers (8 ELA and 10 math)
<b>Michigan</b>	24 teachers (12 ELA and 12 math)	2 teachers (1 ELA and 1 math)	22 teachers (11 ELA and 11 math)
<b>North Carolina</b>	28 teachers (14 ELA and 14 math)	10 teachers (6 ELA and 4 math)	18 teachers (8 ELA and 10 math)
<b>Power of Two</b>	106 teachers (53 ELA and 53 math)	23 teachers (13 ELA and 10 math)	83 teachers (40 ELA and 43 math)

Since an RCT evaluation design was employed, attrition was carefully tracked (i.e., teachers and students who dropped out of the initiative) in order to report the attrition rates. Since the evaluation examines two separate achievement outcomes – reading and math - overall rates of attrition as well as the differential rate of attrition (i.e., between the treatment group and the control group) were calculated for ELA teachers and math teachers separately (Table 3). Among ELA teachers, the overall attrition rate was 24.5% and the differential rate was 4.7%. For math teachers, the overall attrition rate was somewhat lower than ELA teachers, at 18.9%, but the differential attrition rate for math teachers was higher at 8.3%.

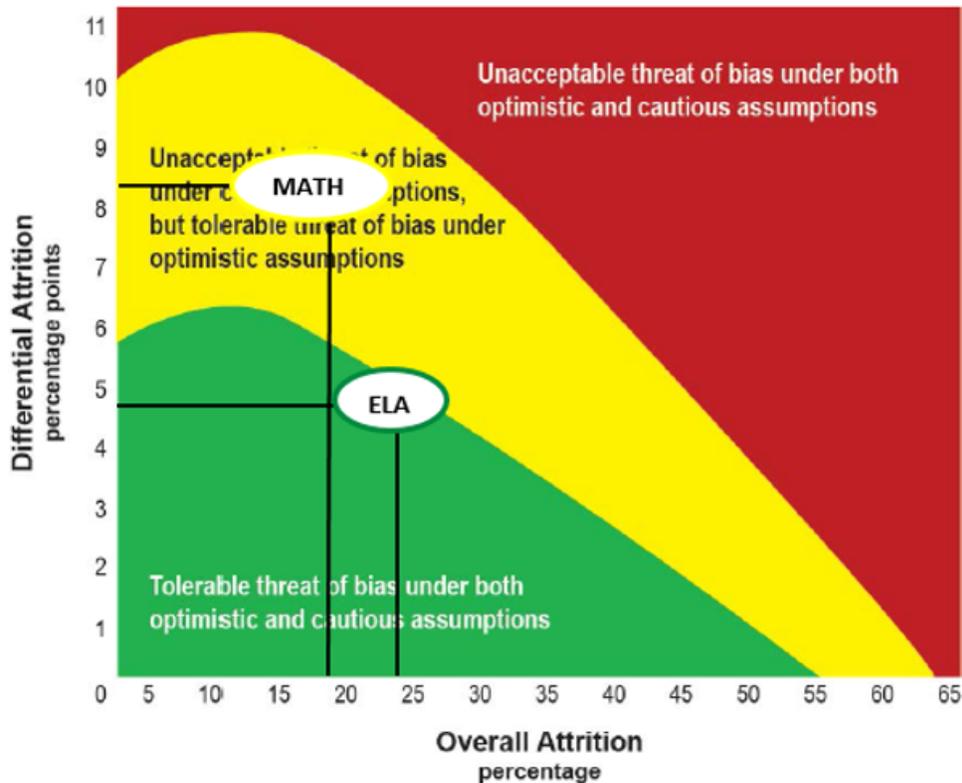
Table 3. Power of Two Attrition Rates

ELA	Treatment	Control	Total
Number teachers randomized	27	26	53
Number teachers lost	6	7	13
<b>Overall attrition rate</b>	22.2%	26.9%	<b>24.5%</b>
<b>Differential attrition rate</b>			<b>4.7%</b>
Math	Treatment	Control	Total
Number teachers randomized	27	26	53
Number teachers lost	4	6	10
<b>Overall attrition rate</b>	14.8%	23.1%	<b>18.9%</b>
<b>Differential attrition rate</b>			<b>8.3%</b>

All RCT studies, even well-designed RCTs, experience rates and patterns of attrition. Teachers and students move, teachers change jobs, principals re-assign students, teachers go on leave, etc. Our goal was to ensure that the sample attrition in our study did not compromise the initial comparability of the treatment and control groups that could potentially lead to biased estimates of the Power of Two’s outcomes. Using the What Works Clearinghouse guidance on sample attrition, we examined both *overall attrition* (i.e., the rate of attrition for the entire sample, measured as the percentage of the randomized sample that has been lost) and *differential attrition* (i.e., the percentage point difference in the rates of attrition for the treatment and control groups), because both types of attrition contribute to the potential bias of the estimated effect (What Works Clearinghouse, 2014).

The WWC provides an attrition standard, based on a theoretical model for attrition bias and empirically based assumptions. The standard offers researchers a range from more cautious assumptions (i.e., more strict) to more optimistic assumptions (i.e., more lenient) to measure the levels of expected bias. The WWC offers three levels: 1) tolerable under both optimistic and cautious assumptions; 2) unacceptable under cautious assumptions but tolerable under optimistic assumptions; or 2) unacceptable under both cautious and optimistic assumptions. An examination of the attrition levels for the Power of Two indicate that the attrition rates among both ELA teachers and math teachers are acceptable using the optimistic assumptions (Figure 1). Therefore, the WWC standards for teacher attrition and the threshold for a tolerable level of bias were met.

Figure 1. Power of Two Attrition Rates and Potential Bias Estimates



In addition to examining teacher attrition, we also tracked and calculated attrition rates at the student level to ensure that the Power of Two student achievement data was consistent and would not lead to biased estimates of effectiveness. For reading achievement, the overall student attrition rate was 12.6% and the differential rate was 6.7%. For math, the overall student attrition rate was 14.2% and the differential attrition rate was 1.8%. When comparing these rates to the WWC guidance, both are acceptable using the optimistic assumptions. Therefore, the WWC standards for student attrition and the threshold for a tolerable level of bias were met.

# Evaluation Design

## Sample Sizes

Table 4 below depicts the sample sizes for each data source that was part of the Power of Two evaluation.

Table 4. Power of Two Sample Sizes

Data Component	Cohort 1 (2016/17)	Cohort 2 (2017/18)
<b>Coach's Log</b>	513 contacts with ELA teachers; 505 contacts with math teachers	544 contacts with ELA teachers; 527 contacts with math teachers
<b>Classroom Observations</b>	48 pre & post observations of ELA teachers; 47 pre & post observations of math teachers	23 post observations of ELA teachers; 24 post observations of math teachers
<b>Teacher Survey</b>	46 ELA & 45 math teachers with pre & post survey data	23 ELA & 22 math teachers with pre & post survey data
<b>Student Survey</b>	911 students with pre & post survey data	472 students with pre & post survey data
<b>NWEA MAP Reading Assessment</b>	957 students with Fall & Spring data	424 students with Fall & Spring data
<b>NWEA MAP Math Assessment</b>	940 students with Fall & Spring data	501 students with Fall & Spring data

# Coach Findings

## Introduction

Each of the Forum's state affiliates in California, Illinois, Michigan, and North Carolina hired their own experienced coaches, who were trained by the ALD and FIM model developers. For Cohort 1, ten coaches (5 ALD and 5 FIM) were employed to support teachers by providing on-the-job mentoring and assisting the model developers with a series of trainings throughout the year. There was very little turnover of coaches between Cohort 1 and Cohort 2. Nine of the 10 coaches were retained to work with Cohort 2. The tenth coach was replaced by a Cohort 1 ALD teacher, who retired from teaching and was hired as an ALD coach. In addition, one state affiliate hired a second FIM coach, who was already trained and working as a FIM coach in the state, bringing the total to 11 coaches (5 ALD, 6 FIM) for cohort 2.

After an initial 2-day training for coaches (and teachers), the Power of Two Initiative utilized a model of meta-coaching hoping to ensure that coaching services across the states were consistent. Both ALD and FIM model developers engaged coaches in monthly meetings with facilitated conversations, provided guided observation opportunities, offered email and phone support, and presented follow-up training opportunities. The model developers provided three follow-up state-level trainings annually. The first day of each training consisted of the coach accompanying the model developer to make observations of all of the Power of Two teachers. These classroom observations then informed the second day of professional development led by the model developers for the coaches and their Power of Two teachers. Coaches also benefited from networking with their fellow coaches to share their experiences and challenges as well as collaborating on strategies to inform their practices. For cohort 2, coaches assisted and sometimes led the trainings for teachers.

On-the-job mentoring of Power of Two teachers was facilitated through coaching contacts, which

“To help plan the review unit for end of year testing based off our FIM data, we charted our last Progress Monitoring Test scores. This helped us to determine what power standards did not need attention, which power standards needed small group attention, and what power standards needed re-teaching. From there, we discussed re-teaching strategies based off the FIM 15 strategies. We talked about review strategies such as Flip teaching, targeted homework, small group, differentiated digital learning on Canvas and planned differentiated review stations.” —FIM Coach

included the following core components for both the ALD and FIM interventions:

- Co-planning between teacher and coach;
- Coach modeling of instructional routines;
- Co-facilitation of lessons by teacher and coach;
- Coaching observations;
- Coach-led conversations with teachers;
- Data analysis for formative assessment; and
- Utilization of relevant resources.

“The teacher benefitted from last week’s professional development session by implementing the ideas which were discussed for pushing students to use higher level vocabulary words in their partnering session and using the collaboration cards more strategically during the lesson by directing students to use a specific phrase while partnering.”

—ALD Coach

Naturally, the content of the coaching contacts varied by intervention. In addition to the core coaching components, ALD coaches also helped teachers to prepare PowerPoint presentations to display response frames for each lesson, develop word banks to increase student academic register and augment Toolkit words, identify grammar targets, etc. FIM coaches provided additional assistance that helped teachers to identify different strategies for solving problems other than using algorithms, incorporate non-cognitive learning into lessons, develop a system for administering cognitive and non-cognitive brag tags to encourage student engagement, etc.

The Coach’s Log is an electronic data collection system for ALD and FIM coaches to complete after each coaching contact with a Power of Two teacher. For the purpose of the Coach’s Log, a coaching contact was typically comprised of a classroom observation and a reflective conversation with the teacher that documented the following:

- [Number of contacts](#) for each individual teacher;
- [Coach’s preparation](#) for the visit (i.e., length of time to prepare, preparation of any data);
- [Classroom observation](#) (i.e., date, length of observed activity, coach’s role);
- [Reflective conversation](#) (i.e., date, length of conversation, focus of conversation, resources utilized); and
- [Overall summary](#) of the visit (i.e., description of the visit, challenges, and successful outcomes).

FIM coaches also recorded the non-cognitive factors (e.g., persistence, flexibility, reflection) discussed with FIM teachers.

The Coach’s Log data were examined, analyzed, and summarized in reports at 2 or 3 intervals per year to identify the change process and quality of implementation. The qualitative data provided in the overall summary (i.e., description of the visit, challenges, and successful outcomes) section of the log were examined using an inductive approach. Recurrent themes were identified with a focus on whether the content represented group-shared ideas. Feedback reports, aggregated by intervention, were disseminated semiannually to the project team for use in monitoring the improvement progress and refining the interventions.

Coaches were required to provide three coaching contacts a month from September through May for each teacher that participated in the yearlong intervention. Overall, coaches provided more than 500 contacts a year to intervention teachers (Table 5). English/language arts teachers in Cohort 1 received an average of 21 contacts. Cohort 2 received an average of 23 contacts for the ALD model. Mathematics teachers in both Cohort 1 and 2 received an average of 22 contacts for the FIM model. Although the average teacher received 20 to 23 contacts (depending on the cohort and the intervention), annual coaching contacts ranged from 8 to 35 contacts per teacher.

Table 5. Power of Two Coaching Contacts by Intervention and Project Year

	Cohort 1 (2016/17)		Cohort 2 (2017/18)	
	ALD	FIM	ALD	FIM
<b>Number of Coaches</b>	5	5	5	6
<b>Number of Teachers</b>	23	23	24	24
<b>Number of Contacts</b>	513	505	544	527
<b>Average Contact/Teacher</b>	21	22	23	22
<b>Coach Preparation Time</b>	15 minutes	17 minutes	16 minutes	17 minutes
<b>Average Contact Time</b>	47 minutes	51 minutes	48 minutes	48 minutes

## Successful Outcomes from Coaching Teachers

As coaches observed positive changes and successful outcomes occurring during coaching contacts, they carefully described both the changes in the teacher’s practices (i.e., planning, instruction, data analysis) and the impact on student engagement and learning. **Coaches identified many common successes, of which the most impactful are highlighted below.**

# Coach Findings

“This teacher continues to thoroughly plan and write out his lessons. We had co-planned this lesson and it went just as planned. He added some students to the reporting that I had not heard report before and they did well. He guided students to incorporate precise academic words. He continues to do a great job.” –ALD Coach

First, coaches reported **conscientious teachers who were receptive to the coach’s constructive feedback** resulting in successful outcomes from the coach/teacher partnership.

Second, there was the **thoughtful co-planning that led to better preparation and implementation**. While co-planning occurred with Cohort 1 teachers during the second half of the year, the more experienced Cohort 2 coaches started this process almost immediately with the Cohort 2 teachers. For FIM teachers, this co-planning focused on classroom management, development of student non-cognitive skills, systemization for awarding brag tags, and improvement of “think alouds,” the metacognitive demonstration of problem

solving. For ALD teachers, co-planning focused on improving instructional routines and strategies to increase student academic register such as grammatical targets, robust word banks for frames, etc., resulting in some highly implementing ALD teachers “thoroughly planning” lessons on their own before the end of the school year.

Third, coaches demonstrated to teachers how to **utilize formative student assessments to inform teacher practice**. For FIM teachers, coaches facilitated “quick sorts” of strategies utilized and common mistakes made on warm-ups and progress monitoring tests to indicate strategies to model, select power standards to reteach, create fluid intervention groups for differentiated instruction, and occasionally to inform regular instruction outside of the FIM warm up. Some highly implementing FIM teachers began utilizing “purposeful data analysis” to improve student learning outside of coaching sessions. For ALD teachers, coaches assisted with reflective analysis of the NWEA MAP Assessment data to inform small group instruction, goal setting, and student conferencing.

“The 6th and 7th grade teachers met together as a team to look at their data. They did a quick sort of the most recent Progress Monitoring Test and looked at content that had been taught in a previous unit. The group brainstormed ways to teach percentages because a teacher said he was stuck and didn’t know how to explain it another way. This led to a rich discussion and the teacher left the meeting with a plan for how he wants to model the problem.” – FIM Coach

Finally, coaches reported **facilitating collaborations between teachers to share challenges and successful strategies to plan for mutual benefit**. This occurred during statewide professional

development opportunities and within schools with multiple Power of Two teachers, with the latter occurring more frequently in year 2 (2017/18) between multiple Cohort 2 teachers or in combination with Cohort 1 teachers continuing to implement in year 2. For ALD teachers, collaborations focused on co-planning lessons, creating lesson PowerPoints, and developing word banks. FIM teachers collaborated to share different problem-solving strategies to augment algorithms.

## Challenges with Coaching Teachers

Coaches faced some common challenges in working with teachers. First, **classroom management and culture** were impeding the abilities of some teachers to implement the interventions. Student engagement and behavior was inhibiting focus and learning. Prioritizing teacher needs over the need to coach on the intervention, coaches responded by providing and modeling classroom-management strategies, as well as helping teachers seek additional supports from the school. Teachers used strategies provided by coaches and achieved some successes in the classroom by the end of the year.

Second, coaches struggled with **pacing concerns** as teachers were not always able to implement ALD and FIM daily given scheduling constraints (e.g., school alterations to calendar, holidays, testing, weather). While coaches encouraged teachers to “double up” lessons to maintain the scheduled pace, some teachers were unable to add a second lesson and fell behind the requirements for the yearlong intervention.

The biggest challenge for coaches was **helping teachers to develop fidelity of implementation of the Power of Two**. Coaches facilitated lessons to model best practices and provided feedback on classroom observations of teachers with reflective conversations. Coaches utilized data to encourage teacher reflection and growth. FIM coaches also used Cognitive Coaching to mediate teacher thinking and empower teachers to be self-directed. The majority of teachers improved their fidelity of implementation over the yearlong intervention. However, **there was a small sample of English/language arts teachers, who chose not to implement with fidelity**. These teachers, or their schools, believed the ALD lesson took too much time away from regular instruction so they modified or omitted steps to complete the lesson in less time, or did not implement the ALD lesson daily.

A final challenge identified from the coaching log was the **fidelity of implementation of coaching services**. The Power of Two Initiative expected coaches to provide three coaching contacts a month from September through May for each teacher, or 27 coaching contacts for the year. However, there was great variability in the actual number of coaching contacts, as teachers received a range of 0 to 5

# Coach Findings

coaching contacts per month, totaling 8 to 35 contacts per year. Furthermore, qualitative descriptions of the coaching contact showed state-level variability in the coach's focus and practices which occurred during the reflective conversations with teachers. This variability occurring with implementation may have had an impact on outcome growth.

## Lessons Learned

There were several coaching lessons learned during the implementation of the Power of Two with Cohort 1 that were addressed for Cohort 2 or have been identified for further refinement.

- [Importance of timing for initial intervention training](#) – The timing of the initial training on the intervention for Cohort 1 coaches and teachers afforded many lessons learned to inform Cohort 2 practices.
  
- **Train coaches beforehand so they can serve as a resource for training teachers.** For Cohort 1, coaches were trained at the same time as the teachers. As a result, Cohort 1 coaches could not provide support during the initial two-days of training for teachers as they were learning the intervention at the same time. For Cohort 2, however, coaches utilized their yearlong coaching experiences with Cohort 1 teachers to augment the Cohort 2 teacher training by answering teacher questions, modeling routines and practices, and advising teachers during their initial training, thereby establishing their credibility as an expert and a resource. In particular, Cohort 2 coaches contributed to the initial training by advising ELA teachers in the selection of ALD units of words to compliment ELA instruction and math teachers in the selection of FIM power standards to address grade-level content standards that guided the yearlong implementation.
  
- **Train teachers closer to the start of the school year.** Teachers received their initial training in late June during their attendance at the National Forum's annual STW Conference. While the conference provided teachers with networking and professional development opportunities, the elapsed time between learning and implementing the intervention proved to be very challenging for most Cohort 1 teachers, as teachers did not implement the intervention until 3 months later in late August or early September. As a result, coaches developed and provided "refresher" trainings for Cohort 2 teachers immediately preceding the start of school year to provide resources, review intervention routines, and co-plan the first week of lessons. A recommendation for

future work would be to structure teacher trainings to include more lesson modeling and opportunities for teachers to practice and receive feedback.

- **Develop a plan for training replacement teachers.** Teacher attrition occurred with both cohorts after the initial 2-day training in late June, which resulted in two different training models. In the first model, coaches oversaw the training of the new replacement teachers with an on-the-job training. This model was more successful with Cohort 2 teachers since coaches had a year of experience with the intervention and could both anticipate teacher challenges and be ready with solutions at the onset of the year. In the second model, the teacher was able to attend another training offered by the model developers. Since teacher attrition is inevitable, developing a plan for the consistent training of replacement teachers is an area that could still be refined.
  
- **Fidelity of Coaching** – As mentioned in the “Challenges with Coaching Teachers” section, there was great variability in the frequency and descriptions of coaching contacts leading to the following recommendations.
  
- **The number of coaching contacts needs to be consistent for all teachers.** Either the National Forum’s state affiliates, or the intervention’s meta-coaches, need to monitor the frequency of monthly coaching contacts. When a deficit in coaching contacts occurs, they need to intervene with coaches more quickly to ensure that teachers are receiving the prescribed dosage.
  
- **Interventions need to define coaching sessions to promote fidelity of implementation.** The intervention’s meta-coaches need to convene regular meetings to define what a coaching session should look like at various stages in the year and review each teacher’s implementation performance to ensure teacher progress and consistency of message. Meta-coaches should provide documentation and videos to reinforce fidelity of implementation. A recommendation for future work would be to make attendance for meta-coaching sessions mandatory for coaches as not all coaches valued and attended these sessions for the Power of Two Initiative.

## Introduction

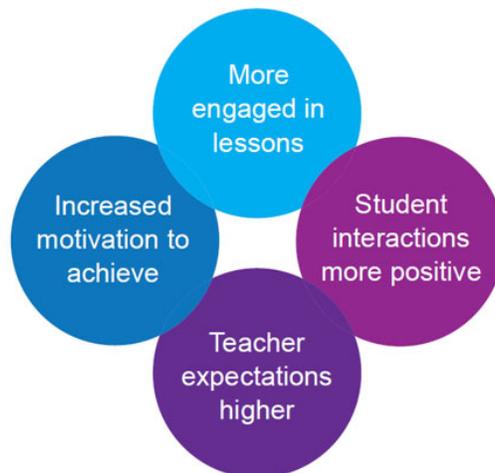
Students engaged in the Power of Two initiative received a paired intervention experience in ELA and math. The pairing of these subject areas addressed the need for students to make rapid academic improvement in a short period of time by supporting them in their academic effort, providing daily practice and repetition of key concepts, and monitoring mastery.

Student intermediate outcomes (academic efficacy, academic expectations, and engagement) were measured prior to treatment and at the end of treatment in both treatment and control classrooms by CPRD’s School Improvement Self-Study Student Survey. To assess the impact of the Power of Two on student achievement growth, the NWEA Measures of Academic Progress (MAP) Assessment was used. Achievement growth was examined at treatment versus control classrooms, using the NWEA’s MAP reading and math assessments. Students in all treatment and control classrooms took the NWEA MAP assessments three times each year (beginning, middle, and end).

## Intermediate Outcomes on Students

Data from teacher surveys, student surveys, and the coach’s logs documented improvements in student learning. Teachers and coaches reported that Power of Two students were more eager and ready to learn resulting in greater engagement, increased motivation to learn, and more positive student interactions (Figure 2). As a result, teachers had higher expectations for students.

Figure 2. Impact of Power of Two on Students



# Student Findings

Coaches observed students that were eager and ready to learn resulting in greater student engagement and confidence in adopting and understanding ALD/FIM instructional routines. In ALD classrooms, students demonstrated increased engagement and confidence when using a high academic register for verbal and written frames, as well as productive partnering. Some teachers shared with coaches that students continued to utilize this increased academic register during regular ELA instruction and other core classes. In FIM classrooms, students demonstrated increased engagement and confidence with positive, academic risk-taking, scaffolding their prior knowledge while completing the “effort” box during their warm-up as well as completing progress monitoring tests.

“Students continue to challenge each other to create more academic responses. The classroom teacher is also holding students to higher standards by incorporating more challenging word banks and encouraging students to do the same.” –ALD Coach

Coaches observed that students in FIM classes demonstrated increased motivation to achieve through goal setting and “brag tags.” Student actively took ownership of their data by setting goals (i.e., PMT, NWEA), reflecting on their progress, and showing excitement to see their growth, which further increased their motivation. The charting of PMT progress on a classroom “data wall” helped

“Students are actively participating with their effort on warm-ups and are showing growth on Progress Monitoring Tests. Also, students are receiving cognitive and non-cognitive brag tags as an incentive for showing growth. Students are engaged with learning math.”

–FIM Coach

to facilitate these conversations and motivated student practice. Teachers also used “brag tags” as incentives to acknowledge and reward student academic and non-cognitive improvements resulting in increased student motivation.

Coaches observed that student interactions were more positive between students and teachers in both ALD and FIM classrooms. Positive rapport between teachers and students improved the overall classroom climate, with teachers encouraging student effort and creating a

positive learning environment and students responding with higher engagement. Teachers in some classrooms communicated higher expectations to students.

Finally, coaches (and teachers) observed a positive impact on student math and ELA outcomes as measured on the NWEA MAP Assessment (see “Achievement Growth” section below) in addition to classroom measurements of growth by intervention. ALD coaches noted improvement of overall student knowledge and usage of high academic vocabulary as measured by the End of Unit

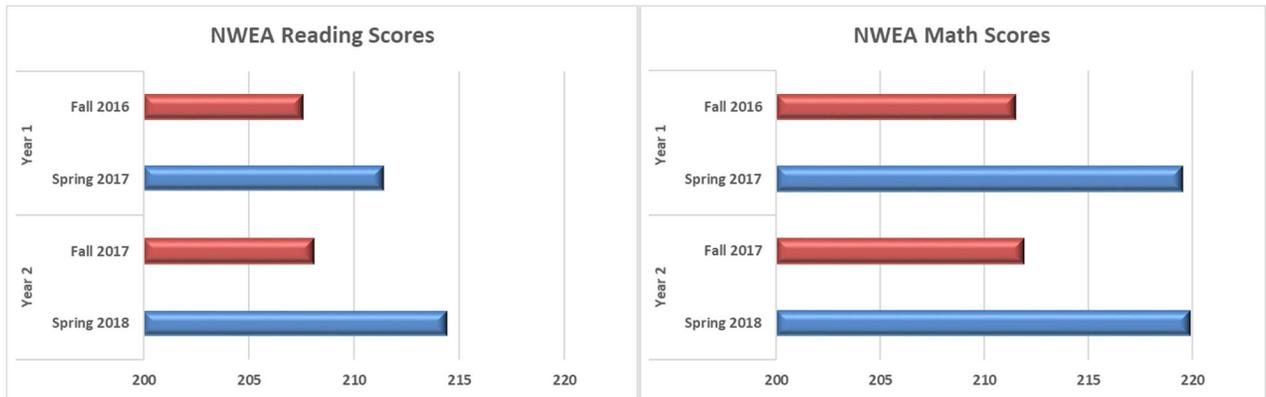
# Student Findings

Assessments. FIM coaches identified improvement in student knowledge of Power Standards (aligned to the Common Core) as measured on the Progress Monitoring Tests and through the earning of cognitive brag tags.

## Achievement Growth

Both cohorts of Power of Two students had gains in NWEA reading and math achievement scores from fall to spring. For Cohort 1, students had a 3.8 point gain in reading scores and an 8.0 point gain in math scores. For Cohort 2, students had a 6.3 point gain in reading and an 8.0 point gain in math (Figure 3).

Figure 3. Power of Two Student Achievement Gains



Interestingly, the achievement gain in Math was the same for Cohort 1 and Cohort 2. The gain in reading, however, was stronger among Cohort 2 students (6.3 points) than Cohort 1 students (3.8 points). Using a linear mixed effect models statistical approach, we ran exploratory analyses to compare the Cohort 1 reading growth to the Cohort 2 reading growth. We found that the Cohort 2 reading growth was significantly higher than Cohort 1. In other words the Cohort 2 Power of Two students showed higher growth than the Cohort 1 Power of Two students. We hypothesize that both coaching services were stronger (see Coach Findings section) and implementation was better (see Teacher Findings section) for ALD teachers in Year 2.

When analyzing the achievement gain among Cohort 1 Power of Two students against the control group, we note that the control group showed similar gains (Figure 4). Although the overall gain is similar, the Power of Two students demonstrated quicker gain in both reading and Math by the winter assessment than the control students.

Figure 4. Power of Two and Control Group Student Achievement Gains (2016-17)



## Impact Study Results

The Power of Two intervention effect was assessed through linear mixed effect models which incorporated the correlation between students in the same class, due to the nature of the randomization of the interventions by classroom (Appendix A). The data identified 14% intra-class correlation (ICC) in reading and 21% in math. A series of random intercept (2-level) models were executed to determine if the Power of Two Initiative resulted in any difference in student achievement scores among Cohort 1 treatment students compared to the control group.

The final analytic sample for reading included 957 students at level-1 (500 at treatment classrooms and 457 at control classrooms) and 47 classrooms at level-2. The final analytic sample for math included 940 students at level-1 (478 at treatment classrooms and 462 at control classrooms) and 46 classrooms at level-2. Any students with missing fall or spring achievement scores were excluded from the analytic samples. The models utilized the following variables to test the impact on reading and math achievement: fall scores as the baseline, spring scores as the outcome, gender and English Language Learner (ELL) status as level-1 covariates, state as a level-2 covariate and intervention condition as independent factor.

The final models, adjusting for all possible covariates, suggest no significant intervention effect on either reading scores or math scores (Table 6). In other words, the Cohort 1 students that received the Power of Two initiative performed the same on both the reading test (Appendix B) and the math test (Appendix C) as the control students. Neither the p-values nor effect sizes suggest a significant intervention effect on achievement.

# Student Findings

Table 6. Power of Two Impact Study Results

Outcome Domain	Treatment Group				Control Group				Impact Estimate (Model Adjusted Mean Difference)	Effect Size	Impact Standard Error	P-value
	Cluster Sample Size	Student Sample Size	Unadjusted		Cluster Sample Size	Student Sample Size	Unadjusted					
			Mean	Standard Deviation			Mean	Standard Deviation				
Reading	24	500	211.4	12.54	23	457	212.5	14.11	-.872	-.065	1.07	.420
Math	23	478	219.5	12.81	23	462	218.6	15.46	.723	.05	1.02	.482

Even though the impact analyses did not find an overall intervention effect on either reading scores or math scores, it should not be interpreted to mean that the Power of Two initiative was not effective, but that classrooms may need additional time to fully implement and refine the initiative with middle grades students to advance achievement scores. Coaches and teachers also wondered if the NWEA MAP was sensitive enough to detect changes in outcomes for the Power of Two. We can further hypothesize that the positive changes noted in other sections of this report on student attitudes, behavior, and engagement, as well as on teacher practices and efficacy are a positive precursor to improvements in achievement since achievement does not improve without corresponding improvements in the teaching and learning environment.

## Exploratory Analysis of Language Arts Achievement Test

Since coaches and teachers observed that the NWEA MAP language arts test may have been a better measure of achievement growth in the constructs taught by the ALD intervention than the NWEA MAP reading test, an exploratory analysis of the language arts test was conducted. Half of the Cohort 1 classrooms (i.e., 19 of the classrooms) administered the NWEA MAP language test in addition to the reading test (9 were treatment and 10 were control). There was slightly higher growth in scores for treatment students (4.4 points) than control students (3.6). Although there was no significant treatment effect on language arts achievement for Cohort 1, the intervention group had a slightly upward (positive) change compared to the control group. Using the language arts test rather than reading test in the future implementation of the Power of Two is worth consideration.

## Lower Achieving Quartile of Students

Since the Power of Two specifically supported second language students as well as students performing below grade level expectations, we examined achievement growth among Cohort 1 and Cohort 2 students by quartile. Among students who began the school year (i.e., fall) performing in the lowest quartile of achievement (below 25<sup>th</sup> percentile), Power of Two students showed more gains by spring than the overall average gains among all students in both reading and math (Table 7). Additionally, for Cohort 1, Power of two students in the lowest quartile had higher gains (10.0 gain) than the control students (7.0 gain).

Table 7. Power of Two Student Achievement Gains – Lowest Quartile

### AVERAGE NWEA TEST GROWTH FROM FALL TO SPRING

		Overall	Lowest Quartile*
Year 1	Reading	3.84	8.0
	Math	8.06	10.0
Year 2	Reading	6.31	12.08
	Math	8.0	9.61

\*Students with baseline NWEA scores below the 25th percentile.

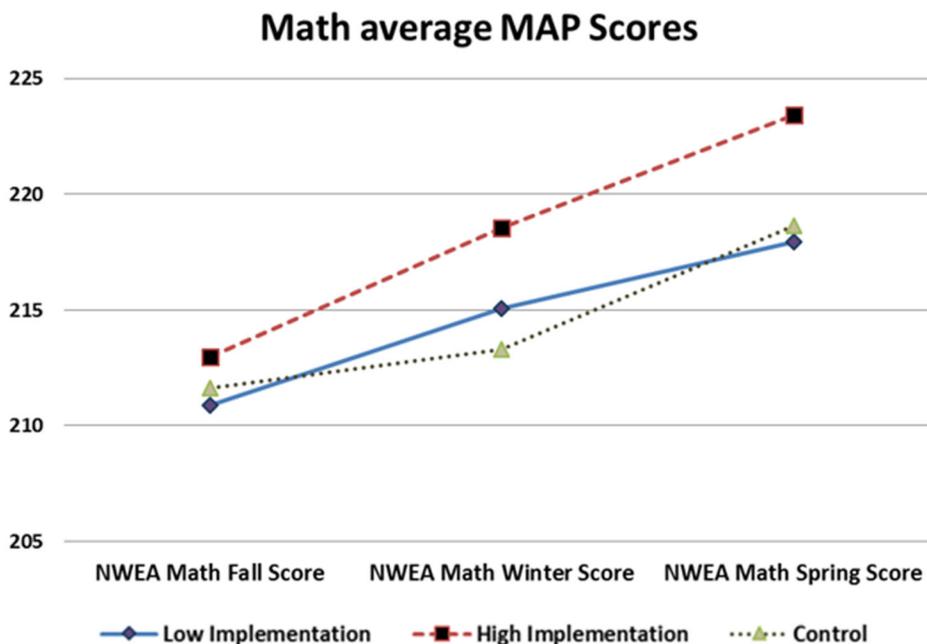
## Impact of Level of Implementation on Achievement

An examination of achievement gains among Power of Two students by state and by classroom revealed significant variability. Several states had greater gains in math, others had greater gains in reading. The same was true among individual classrooms with some classrooms showing larger gains while others did not. We hypothesized that these varying results may have been due to several factors, including amount of coaching received, level of consistent and quality implementation of the ALD and FIM interventions, and school-specific or district factors influencing the ability of teachers to spend time on Power of Two practices.

# Student Findings

Three factors were shown to have the greatest impact on Power of Two student learning: 1) level of consistent implementation (i.e., daily implementation of the ALD and FIM instructional practices and strategies); 2) level of quality implementation (i.e., high fidelity of implementation of the ALD and FIM intervention); and 3) effectiveness of classroom management. Cohort 1 Power of Two students in math classrooms with teachers who were the most highly implemented, confident, and effective with their classroom management skills, showed *significantly* greater math achievement growth than students in lower implemented classrooms as well as control classrooms (Figure 5). A series of random intercept (2-level) models were executed to determine if the highly implemented Cohort 1 ALD and FIM classrooms resulted in higher student achievement. The final model that examined implementation, adjusting for all possible covariates, showed that for FIM classrooms, highly implemented teachers had significantly higher math achievement growth than control students (Appendix D).

Figure 5. Math Achievement Gains for High FIM Implementation vs. Control



Additionally, highly implemented FIM teachers reported that the students in their Power of Two classrooms were more likely to initiate positive interactions with each other, displayed lower disruptive behavior during class, had stronger focus on achievement, and were more motivated to do well in school when compared to students in the control classrooms (Table 8).

# Student Findings

Table 8. Classroom Climate of Power of Two Highly Implemented FIM Teachers

Teacher Survey Construct	Highly Implemented FIM Classrooms Mean (SD)	Other FIM Classrooms Mean (SD)	Control Classrooms Mean (SD)
Students initiate positive interactions	3.96 (0.89)	3.58 (0.76)	3.47 (0.60)
Disruptive student behavior	2.60 (1.01)	2.99 (0.85)	3.11 (0.76)
Students have a focus on achievement	3.71 (1.11)	3.38 (0.81)	3.57 (0.50)
Student Motivation	3.86 (1.14)	3.22 (1.23)	3.51 (0.99)

## Lessons Learned - Student Outcomes

- Several participating districts required signed parental consent in order to release the NWEA MAP assessment data to the evaluation team. This resulted in less data from two districts from two different states and thus smaller samples. Future implementation of the Power of Two should investigate the parent consent requirements and research application requirements in every district prior to recruiting schools so that an informed decision can be made about that district's participation.
- The administration of the NWEA MAP assessments were new for some teachers and the training they received for appropriately creating a test environment and adherence to administration procedures was limited. The quality of the data would have been improved if all Power of Two teachers (treatment and control) would have followed an established process of ensuring that the environment for testing was consistent, test duration was consistent, the same protocol was used for test administration, and the same decision making was used for any re-testing decisions. Power of Two partners reflected after the project that it would have been beneficial to do a simulation of the NWEA test environment to assist teachers with proper administration of the test.
- The window for Power of Two teachers (treatment and control) to administer the NWEA MAP test varied by district. According to NWEA, a critical issue with using the data is the number of

weeks of instruction prior to the test. If this varies, we can expect to have different outcomes among the students. It would have been beneficial to establish a tighter window of when the NWEA MAP is administered across all states.

- There was variance in how each district viewed and used the NWEA MAP assessment. It was used by some districts for accountability. It was used in other districts for regular formative assessment of students. For several districts, they had not used NWEA MAP before this project and only implemented in the project classrooms. These issues may have impacted growth on the NWEA MAP over time. Additionally, coaches and teachers wondered if the NWEA MAP was sensitive enough to detect changes in outcomes for the Power of Two, particularly for ALD where the reading test was used as opposed to the language arts test. Assessments are often not sensitive enough to changes in instruction.

# Teacher Findings

## Introduction

ELA and math teachers that implemented the Power of Two initiative received professional development and ongoing one-to-one coaching to apply the key ALD and FIM strategies and instructional routines in their classrooms. The professional development and coaching supported teachers to both ensure that the ALD and FIM interventions were implemented with fidelity and to mentor teachers in their individual growth.

Teacher intermediate outcomes (improved teaching practices, efficacy, engagement of students, using data to inform practice, and confidence) were measured prior to treatment and at the end of treatment in both treatment and control classrooms by CPRD's School Improvement Self-Study Teacher Survey. Additionally, the Coach's Log was also used to assess level of implementation, quality of practices, successes, challenges, and barriers to implementation.

## Intermediate Outcomes on Teachers

Data from teacher surveys and coach's logs documented the positive impact that the Power of Two initiative had on teacher instruction. Teachers had mastered the ALD and FIM instructional practices and engaged more frequently in lesson planning, which led to increased confidence in teaching practices (Figure 6). In addition, teachers were using data more frequently for "purposeful data analysis" to improve individual student learning and to reflect on their practice.

"ALD has helped me be consistent and intensive with vocabulary instruction. It has also helped me make vocabulary more engaging and interactive for students. I also understand the importance of verbally practicing vocabulary."

—ALD Teacher

Through the coach's log, coaches captured the journeys of their teachers as they mastered the ALD and FIM instructional routines. Early teacher excitement, but reliance on written instructional materials, gradually yielded to teacher confidence. This occurred after daily repetition of the ALD and FIM instructional routines and coach's feedback helped teachers to refine and solidify the new teaching practices. Regularly scheduled time for lesson planning, "purposeful data analysis," and reflection strengthened teacher quality and fidelity of implementation of the instructional routines and reinforced teacher confidence in their practice, ultimately improving student learning.

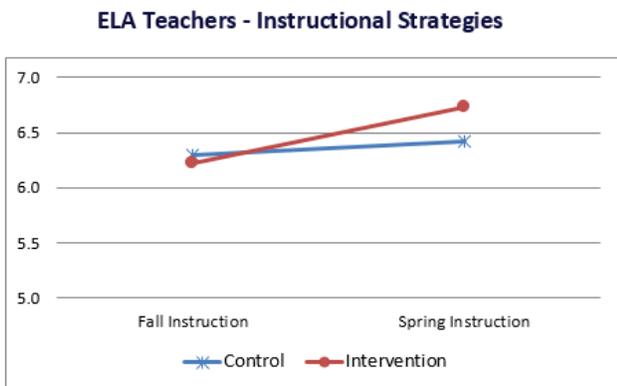
Figure 6. Impact of Power of Two on Teachers



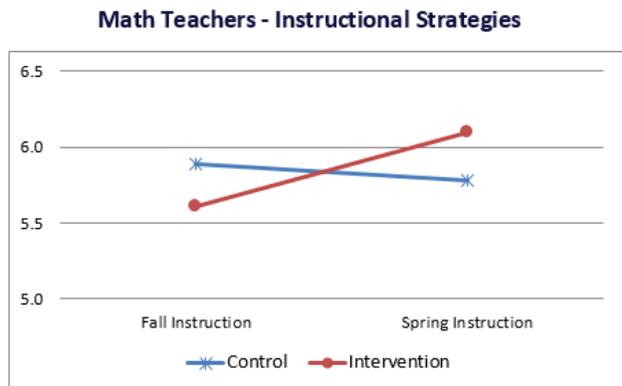
An examination of teacher survey data illustrates that ALD and FIM teachers significantly improved their instructional strategies as a result of the intervention (Figure 7), particularly in the areas of engaging students in lessons (e.g., clear expectations communicated, provide feedback to students, gather information about where students are struggling, and working with students to track their progress).

“With FIM, I review key concepts throughout the year instead of just hitting them once. I have a better sense of which concepts students are understanding and which they’re struggling with.”  
 –FIM Teacher

Figure 7. Improvements in Instructional Practices



1=Never; 2=Several time a year; 3=Monthly; 4=Several times a month; 5=Weekly; 6=Several times a week; 7=Daily



1=Never; 2=Several time a year; 3=Monthly; 4=Several times a month; 5=Weekly; 6=Several times a week; 7=Daily

“High levels of fidelity of implementation in this classroom. The teacher continues to hold students to high standards and lessons are taught thoroughly.”  
—ALD Coach

Essential to the successful implementation of the ALD and FIM instructional routines was lesson planning, which focused on improving quality and fidelity of implementation. During regularly scheduled reflective conversations, coaches reported that lesson planning was a primary focus of discussion with teachers. Initially modelled and led by coaches, lesson planning evolved into co-planning with the teacher taking ever-greater ownership

of their lessons. As teachers became more proficient in planning, they also experienced opportunities to plan with other teachers where they shared challenges and successful strategies to plan for mutual benefit. FIM teachers focused on developing both cognitive components (e.g., improving the meta-cognitive “think aloud” demonstration of problem solving, identifying multiple problem-solving strategies in addition to algorithms to model) and non-cognitive components (e.g., incorporating *Habits of Mind*, utilizing “brag tags” for motivation), as well as the classroom management strategies necessary to facilitate lessons. ALD teachers focused on implementing and improving each of the 20-steps of the ALD instructional routine, developing robust word banks and grammatical targets for each lesson’s response frames. Coaches reported that some ALD teachers were “thoroughly planning” lessons on their own by the end of the year.

Coaches reported that another frequent focus of reflective conversations with teachers was the interpretation of data results to inform teacher practice with the goal of improving student learning. Like lesson planning, “purposeful data analysis” began with the coach modeling how to utilize formative assessments to inform teacher practice and evolved into a collaborative partnership with teachers. ALD teachers utilized end of unit assessments and NWEA MAP data to inform small group instruction, goal setting, and student conferencing. FIM teachers focused on “quick sorts” of strategies utilized and common mistakes made on warm-ups and progress monitoring tests to indicate strategies to model, select power standard to reteach, create “fluid” intervention groups for differentiated instruction and occasionally to inform regular math instruction. Coaches reported that highly implementing FIM teacher continued utilizing their “purposeful data analysis” outside of coaching sessions. Coaches also observed FIM teachers engaging students in the data analysis and reflection process through student discussions of their progress monitoring tests, graphing of their longitudinal progress monitoring test results, and setting and adjusting growth goals, resulting in students taking more ownership of their learning. FIM teachers also created classroom data walls to graph and discuss class results on the progress monitoring tests.

# Teacher Findings

Coaches observed that improved understanding of the ALD and FIM instructional routines combined with better preparation resulted in greater teacher confidence and consistency in delivery of instructional routines. Another contributing factor was teacher recognition that their implementation of the ALD and FIM instructional routines was directly benefiting student engagement, motivation, and learning outcomes as measured by the NWEA MAP, ALD End of Unit Assessments, and FIM Progress Monitoring Tests. Furthermore, ALD teachers recognized an increase in student academic register that extended beyond the ALD lesson into regular ELA instruction and other core classes. As a result, teachers endorsed ALD and FIM to their principals and advocated for the continuation of these practices beyond the one-year intervention. The Power of Two Project supported Cohort 1 teachers who wanted to continue in Year 2 and had the support of their principals. Some schools continued the intervention beyond the grant, supporting the intervention teachers and training additional teachers in order to maintain the momentum of the Power of Two's impact.

“Teacher shared how she had students reflect on their test results in writing and set goals in their journals. We looked at student goals and compared them to PMT results to make a plan for how to help students measure and refine their goals based on their data. .”  
—FIM Coach

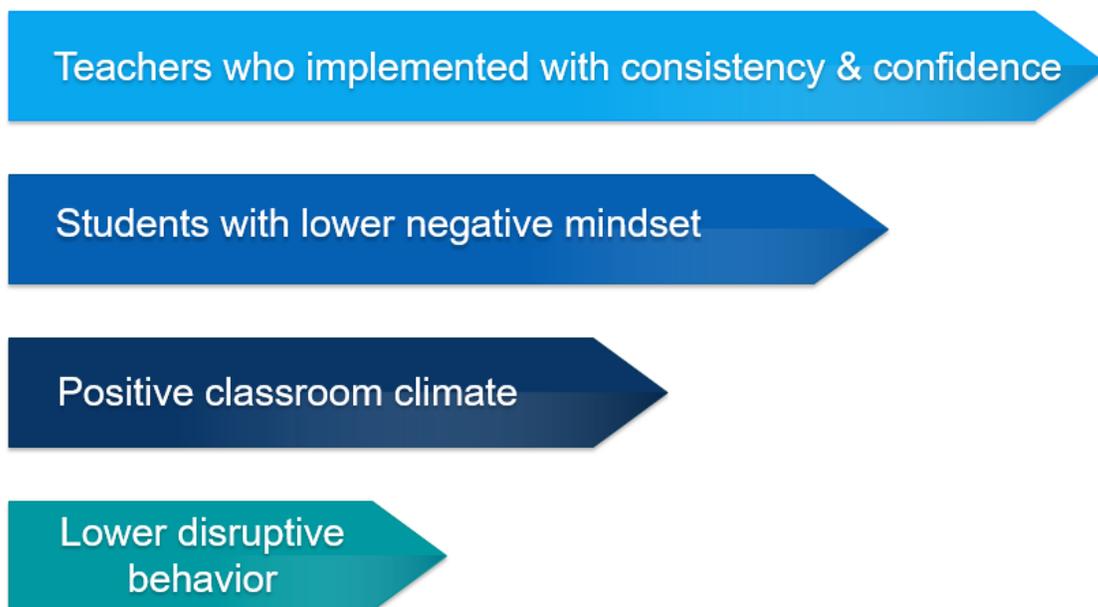
The basic structure of a Power of Two coaching contact was a classroom observation and a reflective conversation, so time was regularly scheduled to support teacher reflection on their practice three times a month. Some FIM coaches also guided teachers through mid-year and end-of-year reflections on the teacher's ongoing growth and progress. The reflective focus evolved over the year as coaches guided teachers to examine first the mechanics of implementing each of the instructional steps. Focusing next on the planning of lessons augmented the fidelity and quality of implementation. Finally, there was the emphasis on purposeful data analysis to identify growth areas for both teachers and students. Teachers adjusted their practices to respond to student needs, which informed the ALD and FIM lesson and sometimes informed regular ELA and math instruction. FIM coaches utilized Cognitive Coaching to mediate teacher thinking and empower teachers to be self-directed in order to sustain changes in teacher practices beyond the grant. Contributing to the success of reflective process were teachers who wanted to do well and were receptive to constructive feedback from their coach.

## Impact of Level Implementation on Teacher Findings

Data from teacher surveys, classroom observations, coach’s logs, and discussions with coach’s and state affiliates were used to track fidelity of implementation of the Power of Two. The major components of implementation included: teacher participation in training/professional development; teacher participation in ongoing coaching sessions; teacher daily implementation of ALD and FIM; teacher adherence to ALD and FIM protocols; and teacher attitude/buy-in to the Power of Two.

Figure 8. Characteristics of Highly Implemented Power of Two Teachers

### Highly Implemented Classrooms...



The majority of teachers attended training/professional development sessions. As described in the coach’s finding section, however, teachers were not always able to implement the Power of Two daily given scheduling constraints. Additionally, while coaching services were expected to be provided three times per month for each teacher, there was a great deal of variability in the actual number of sessions that teachers attended. While the average teacher received 20 to 23 coaching sessions, the range varied from eight to 35 coaching sessions per teacher, with additional variability in the focus and content of the sessions. While the coaches modeled lessons for teachers, observed teacher

# Teacher Findings

instruction and provided feedback, it took the majority of teachers several months of practice to improve their fidelity of implementation.

Through analyses of the available data sources, 13 teachers (6 ALD, 7 FIM) from Cohort 1 and 16 teachers (8 ALD, 8 FIM) from Cohort 2 were identified as high implementers. The characteristics of high implementers were powerful. 100% said they were “totally confident” in their ability to implement the Power of Two and 100% said they were “very satisfied” with their coach and the training they were receiving. Upon further examination, the highly implemented teachers also described the students in their classrooms as initiating positive interactions with each other more frequently. In the case of the highly implemented FIM teachers, they also reported that their students were more motivated to do well in school and displayed lower disruptive behavior when compared to students in the control classrooms. Additionally, the students in both highly implemented ALD and FIM classrooms reported a more positive overall climate and lower feelings of negative mindset (Figure 8).

Table 9. Teacher Survey Constructs by Highly Implemented, Other Intervention, & Control Classrooms

	Other Intervention Group			Highly Implemented Classrooms			Control		
	Mean	Valid N	SD	Mean	Valid N	SD	Mean	Valid N	SD
<b>Students Initiate Positive Interactions</b>									
Fall	3.47	15	0.75	3.64	7	0.78	3.39	22	0.63
Spring	3.58	15	0.76	3.96	7	0.89	3.47	22	0.60
<b>Disruptive Student Behavior</b>									
Fall	3.11	15	0.90	2.94	7	0.71	3.03	22	0.68
Spring	2.99	15	0.85	2.60	7	1.01	3.11	21	0.76
<b>Teacher Efficacy</b>									
Fall	4.32	15	0.43	4.18	7	0.59	4.20	22	0.55
Spring	4.25	15	0.67	4.46	7	0.39	3.99	22	0.54
<b>Students Are Motivated to Achieve</b>									
Fall	3.17	15	1.25	3.71	7	0.71	3.48	22	0.97
Spring	3.22	15	1.23	3.86	7	1.14	3.51	22	0.99

The constructs measured by the teacher survey depicting greater improvements among highly implemented classrooms include teacher ratings of self efficacy, positive student behavior, disruptive

# Teacher Findings

behavior, and student focus to achieve academically. Each of these constructs improved substantially in highly implemented classrooms as compared to other intervention classrooms as well as the control group (Table 9).

## Lessons Learned – Teacher Outcomes

- The Power of Two implementation experiences among Cohort 1 coaches and teachers provided many lessons learned to inform and improve upon the Cohort 2 practices.
  - **Coaches in Cohort 2 were experienced with ALD and FIM from coaching during the first year, so they were able to “hit the ground running” more quickly** with their Power of Two teachers. In Cohort 1, coaches were trained during the same sessions as teachers. Additionally, the coaches in Year 2 were aware of when and how teachers may struggle and were able to both anticipate this and be ready with solutions.
  - **Coaches in Year 2 had previously been trained**, so they served as a resource to the FIM and ALD trainers during the Cohort 2 training and were able to work with the teachers during the training in meaningful ways.
  - **FIM and ALD was implemented in Cohort 2 in all class sections** for treatment teachers. In Cohort 1, only a few teachers did this. Coaches observed that Cohort 2 teachers taught the lessons in more classes, and therefore received more practice in order to refine the protocols and adhere to fidelity of implementation.
  - **For Cohort 2, ALD coaches did two demonstration lessons per month** for Power of Two teachers for the first several months. This practice was added for Cohort 2 because it was noted by the model developer that teachers needed more exposure to complete and accurate adherence to the lesson implementation in order to build knowledge and skills.
- A stronger and more complete communication plan that was implemented across all schools for the Power of Two initiative was needed. This would ensure that all principals understood the requirements of scheduling students into treatment and control classrooms before scheduling occurred, that they understood more about the initiative in order to support their teachers’ change in practice, and that they were engaged in examining the impact of the initiative along the way. Additionally, the National Forum recognized that some Power of Two

## Teacher Findings

teachers did not understand the “why” of what they were implementing. More training and time was needed to communicate the big picture to both principals and teachers. Also, it would have helped if the principal was part of these trainings/conversations so that he/she could support the teacher. One idea would be to accomplish this with teachers and principals by doing PLCs on the project.

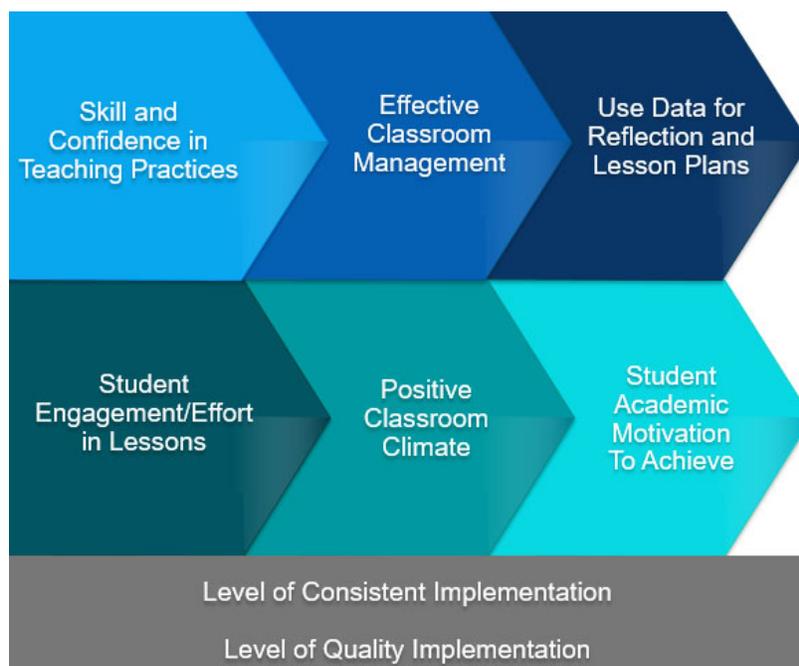
- The overall project goals may have benefited from intentional efforts/activities to facilitate knowledge and collaboration between the ELA and math pairs: introduce FIM to the ELA teachers; introduce ALD to the math teachers; provide time for ELA and math pairs to meet so they can discuss student progress and plan improvement efforts. We should have also introduced the FIM and ALD coaches to each other, and the FIM coach to the ELA teachers, and the ALD coach to the math teachers.

There are a myriad of lessons learned from the Power of Two Project that have been explicated in this report. They are important to document as part of the evaluation of the implementation and outcomes of the project. These lessons also provide valuable information for future implementation of this project and serve to inform other middle-grades schools embarking on subject-specific interventions. Although there were many differences between the project districts and schools, such as their location (rural/small town/suburban), enrollment, building grade configuration, and district and state requirements, they shared many of the same challenges and struggles to improve their instructional practices, classroom climate, and improve outcomes. The summary presented below and in Figure 9 reflect findings from the data collected as part of the evaluation from all stakeholders in the project including principals, teachers, coaches, and state project leaders.

## Summary of Findings

1. Power of Two Cohort 1 and 2 teachers mastered the FIM and ALD instructional strategies and routines, used data more frequently, and were more confident in their teaching.
2. Coaches observed proficiency and confidence in Power of Two Cohort 1 and 2 teachers with FIM and ALD instruction, higher engagement of students during lessons, greater use of data to improve student learning, and more reflection on practice.

Figure 9. Power of Two Intervention Outcomes



# Summary

3. Teacher survey data showed that Power of Two Cohort 1 and 2 ALD and FIM instructional strategies and routines increased.
4. Data from teacher surveys, student surveys, and the coach's logs documented improvements in student learning. Teachers and coaches reported that Power of Two students were more eager and ready to learn resulting in greater engagement, increased motivation to learn, and more positive student interactions. As a result, teachers had higher expectations for students.
5. Student NWEA MAP achievement test data showed that Cohort 1 and 2 Power of Two students experienced growth in math and reading achievement.
6. Students in Cohort 1 Power of Two classrooms with the most highly implemented and confident math teachers showed significantly ( $p < 0.01$ ) greater math achievement growth than control students. And:
  - Students had lower negative mindset.
  - Math teachers reported higher levels of teaching efficacy and that their students were more likely to initiate positive interactions with each other, were more motivated to do well in school, had higher academic expectations of them, and reported lower disruptive student behavior in the classroom.
  - Disruptive student behavior in the classroom is the main difference between math superstar classrooms and other treatment classrooms and control classrooms.
7. Power of Two students in Cohort 1 and 2 who began the school year in the lowest quartile of achievement tended to show the greatest growth in both reading and math achievement after participation in the Power of Two Project.
8. Students in Cohort 2 Power of Two classrooms did significantly ( $p < 0.01$ ) better on the NWEA MAP reading test than students in Cohort 1 Power of Two classrooms. We hypothesize that this could mean many things like implementation was better in Year 2 and/or coaching was refined/more effective in Year 2, and/or due to changes in how teachers were trained, etc.

## Summary of Process Factors that Influence Outcomes

The three factors that had the greatest impact on Power of Two student learning were: 1) level of consistent implementation (i.e., daily implementation of the ALD and FIM instructional practices and strategies); 2) level of quality implementation (i.e., high fidelity of implementation of the ALD and FIM intervention); and 3) effectiveness of classroom management. Having an intentional focus on implementation quality and defining quality for everyone as well as meeting regularly with teachers and coaches to do implementation checks are beneficial strategies for improving implementation. Another factor that was shown to improve implementation was in schools where the principal or

# Summary

leadership participated actively in the Power of Two (i.e., the coach met once a month with the principal) to support the teachers and encourage teacher buy-in and full implementation.

For all interventions, implementation levels and quality is always a challenge and requires thoughtful planning at the outset as well as regular check-ins throughout. For the Power of Two, which was implemented in many different districts in four states, local considerations such as district or state requirements, also play a role as implementation must be balanced with the realities of meeting the unique needs of teachers and schools in a variety of settings.

## Considerations for Future Implementation

The leadership team, state affiliate leaders, and coaches involved in the Power of Two expressed a belief in the positive impact of the intervention on the teachers and students who were involved. The data presented here have highlighted the many areas of growth as well as lessons learned throughout the project. During the concluding project meetings, the Power of Two team reflected on considerations for future implementation of the Power of Two project and these included:

- An intentional focus on tightening up the implementation so that it is more alike than different across states and school. For example, define what a coaching session should look like; have teachers implement in all classrooms; convene regular meetings to go over implementation and ensure that it remains consistent, etc.
- The overall project goals may have been augmented by including intentional efforts and activities to facilitate knowledge and collaboration between the ELA and math teacher pairs. For example: introduce FIM to the ELA teachers and introduce ALD to the math teachers; provide time for ELA and math pairs to meet so they can discuss student progress and plan improvement efforts; introduce the FIM and ALD coaches to each other, and the FIM coach to the ELA teachers, and the ALD coach to the math teachers.
- The Power of Two was implemented in a few classrooms within each school. The implementation may have benefited from implementation across entire grade levels or the entire schools to capitalize on the momentum of change that occurs when everyone is working on the project and can share lessons with each other and brainstorm when challenges arise.

# Appendices

## Appendix A. Statistical Model Structure

In the random intercept model, the intercepts  $\beta_{0j}$  are random variables representing random differences between classrooms:

Level 1:

$$Y_{ij} = \beta_{0j} + \beta_{1j} (Y_{pre.ij}) + \sum \beta_{k+1} (x_{ijk}) + \varepsilon_{ij}$$

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} (Trt) + \sum \gamma_{0j} (w_{ij}) + \mu_{0j}$$

Where

$Y_{ij}$  = Spring MAP scores (reading or math) of the  $i^{\text{th}}$  student in the  $j^{\text{th}}$  classroom.

$Y_{pre.ij}$  = Fall MAP scores (reading or math) of the  $i^{\text{th}}$  student in the  $j^{\text{th}}$  classroom.

$x_{ijk}$  represents a set of student level covariates included in the model; such as, gender, state dummies, LEP status.

$w_{ij}$  represents a set of school level covariates thought to include in the model; such as, intervention condition, state dummies

$Trt = 1$  if treatment classrooms,

$=0$  if control classrooms

$\varepsilon_{ij}$  = residual error,  $\varepsilon_{ij} \sim N(0, \sigma^2)$  and  $\mu_{0j} \sim N(0, \tau^2)$  and independent of  $\varepsilon_{ij}$ .

# Appendices

## Appendix B. Estimates of Statistical Models for Reading

Fixed Effect	Coefficient	S.E.	P value
$\gamma_{00}$ = Intercept	212.40	.772	.000
$\gamma_{10}$ = Coefficient of Fall MAP scores	.727	.024	.000
$\gamma_{20}$ = Coefficient of intervention	-.872	1.07	.420
IL	5.65	1.53	.001
NC	.121	1.66	.942
CA	1.37	1.43	.343
Gender	-.810	.571	.156
ELL status	-2.43	1.30	.062
Random Part	Variance Component	S.E.	P value
Level-one variance:			
$\sigma^2$ = Residual variance	73.40	3.49	.000
Level-two variance:			
$\tau_{02}$ = Classroom variance	9.13	2.85	.001
<b>-2Log Likelihood</b>	<b>6748.41</b>		

Number of student = 957, Effect size ES = 0.065

# Appendices

## Appendix C. Estimates of Statistical Models for Math

Fixed Effect	Coefficient	S.E.	P value
$\gamma_{00}$ = Intercept	218.63	.727	.000
$\gamma_{10}$ = Coefficient of Fall MAP scores	.887	.022	.000
$\gamma_{20}$ = Coefficient of intervention	.723	1.02	.482
IL	3.66	1.47	.016
NC	2.35	1.57	.142
CA	-.373	1.35	.784
Gender	-.412	.508	.418
ELL status	-2.03	1.15	.078
Random Part	Variance Component	S.E.	P value
Level-one variance:			
$\sigma^2$ = Residual variance	57.57	2.76	.000
Level-two variance:			
$\tau_{02}$ = Classroom variance	8.50	2.53	.001
<b>-2Log Likelihood</b>	<b>6386.88</b>		

Number of student = 940, Effect size, ES = 0.05

# Appendices

## Appendix D. Estimates of Statistical Models for Highly Implemented Math

Fixed Effect	Coefficient	S.E.	P value
$\gamma_{00}$ = Intercept	218.66	.679	.000
$\gamma_{10}$ = Coefficient of Fall MAP scores	.888	.022	.000
$\gamma_{20}$ = Coefficient of highly implemented classrooms	3.43	1.42	.020
$\gamma_{30}$ = Coefficient of low implemented classrooms	-.423	1.06	.691
IL	4.11	1.38	.005
NC	1.73	1.50	.253
CA	.100	1.27	.937
Gender	-.442	.508	.385
ELL status	-2.08	1.14	.070
Random Part	Variance Component	S.E.	P value
Level-one variance:			
$\sigma^2$ = Residual variance	57.59	2.76	.000
Level-two variance:			
$\tau_{02}$ = Classroom variance	6.99	2.20	.001
<b>-2Log Likelihood</b>	<b>6380.86</b>		

Control classrooms were the reference group, Number of student =  
 Effect size, ES for highly implemented classrooms = 0.24, ES for low implemented classrooms = - 0.03