COMPASS

Collaborative Organizational Model to Promote Aligned Support Structures

Final Evaluation Report

US Department of Education
Investing in Innovation (i3) Fund

Submitted by Lisa Holliday and Joel Philp

to

The Iredell-Statesville School District

Iredell, NC

October, 2015

Iredell - Statesville Schools
Igniting A Passion for Learning

TEG THE EVALUATION GROUP

COMPASS

Collaborative Organizational Model to Promote Aligned Support Structures

Final Evaluation Report

Lead Agency: Iredell-Statesville School District
410 Garfield St,
Statesville NC 28677

Project Directors: Sherrard Martin and Brie Beane

Third-Party Evaluators: Lisa Holliday, Ph.D. (Implementation Evaluation)
Joel Philp, Ph.D. (Impact Evaluation)
The Evaluation Group

Contact information:

- Martin:704-832-2549; sherrard_martin@iss.k12.nc.us
- Beane: 704-832-2524; brianne_beane@iss.k12.nc.us
- Holliday: 704-996-5291; lisa@evaluationgroup.com
- Philp: 803-719-5102; joel@evaluationgroup.com

Grant Period: 01/01/2010 – 01/01/2015

Funder: US Department of Education, Office of Innovation and Improvement (OII), Investing in Innovation (i3) Fund

PR Award # U396C100105

i3 Cohort Year: 2010

i3 ID DEV19

Funding for this report came from the U.S. Department of Education under its Investing in Innovation (i3) initiative within the Office of Innovation and Improvement through Grant U396C100105 to the Iredell-Statesville School District. The i3 grant award required an independent evaluation of the implementation and impacts of the program. The opinions, findings and conclusions in this report do not necessarily represent the official positions or policies of the funders. Correspondence regarding this document should be addressed to Lisa Holliday (Implementation Evaluation) or Joel Philp (Impact Evaluation), The Evaluation Group, 169 Laurelhurst Ave., Columbia SC 29206. Phone: (803) 719-5102; email Joel@evaluationgroup.com. For information about The Evaluation Group, see our website: www.evaluationgroup.com.

Table of Contents

List of Tables and Figures

Executive Summary

In 2010, Iredell-Statesville Schools was awarded an Investing in Innovation grant (i3) from the Office of Innovation and Improvement within the Federal Department of Education. **C**ollaborative **O**rganizational **M**odel to **P**romote **A**ligned **S**upport **S**tructures (COMPASS) is a development grant that seeks to meet the needs of students with disabilities, academically struggling high-needs students, and students with limited English proficiency by providing timely and targeted professional development to teachers through the alignment of support structures.

The Evaluation Group conducted a fidelity of implementation (FOI) study to determine the extent to which the program was delivered as intended. We also conducted an impact study to determine the effect of the program on reading achievement in grades 3-8.

The implementation evaluation assessed both educative and procedural components. Educative components included delivery and attendance at early release professional development sessions, Response to Intervention (RtI) sessions, and specialized COMPASS trainings to targeted support staff. Assessing the fidelity of the procedural/pedagogical components included evaluating the alignment of the program with district leadership teams and professional learning communities (PLCs). COMPASS was well- implemented. The district met its fidelity targets for all years of the evaluation. Fidelity scores ranged from 86%-100%.

COMPASS was found to have a positive impact overall and at some but not all grade levels. The impact study used a short interrupted time-series with comparison group design (C-SITS) to examine the effect of COMPASS on school-level standardized test scores. Reading outcomes were measured at the school level before and after implementation of the program in COMPASS schools, and at the same time points in comparison schools. Effects were assessed for grades 3-8 combined, and for each of grades 3 through 8 separately using two-level and three-level hierarchical linear modelling. School-level ELA scaled scores were converted to Z-scores with a mean of 0 and a *sd* of 1 to ensure comparability across grades and years. The effect of COMPASS, combined across grades 3-8, produced a significant impact estimate of 0.39, which translates into a gain of almost 4 scale score points over schools in the comparison group. We also found positive and statistically significant impacts within grade 4, grade 5, and grade 7, with estimates ranging from 0.42 (grade 4) to 0.64 (grade 5). Impact estimates for grade 3 and grade 6 were in the intended direction but did not reach statistical significance. The impact estimate for the eighth grade was in the negative direction (that is, COMPASS achievement was lower than comparison schools) but the estimate of -.17, was not statistically different from zero. It should be noted that this is school-level data and the impact at the individual student level is likely to be somewhat smaller.

Because all COMPASS schools reside in the same district, this study suffers from an N=1 confound. This occurs when a program can only be implemented in one classroom (or one school or one district) and the effects cannot be disentangled from other factors that may be operating within that classroom (or school or district). This opens up the possibility that the change in test scores may be due to other influences within Iredell-Statesville Schools, including other interventions being implemented, and not solely a result of the effects of COMPASS.

However, due to the high degree of fidelity (>80%) maintained by the district in all years of implementation, it is reasonable (but certainly not definitive) to conclude that COMPASS is a valid explanation for the improvement in test scores. COMPASS had an overall effect (grades 3-8) and an effect in grades 4, 5, and 7.

Acknowledgments

A number of people have made substantial contributions to this report. At the Iredell-Statesville School District, we wish to thank Sherrard Martin and Brie Beane, COMPASS Project Directors who enthusiastically offered their cooperation and assistance throughout the process; Laura Elliot, Director of Testing, for her assistance with identifying data sources and providing timely information; and Melanie Taylor, Deputy Superintendent of Curriculum and Instruction, for providing leadership and oversight to the entire project. At Abt, the Nei3 Technical Assistance providers, we would also like to thank our TA liaison, Linda Caswell, for her unwavering support, insights, and acumen throughout the evaluation, particularly as it relates to fidelity of implementation. We also wish to express our deep gratitude to Chris Price for his patience and gentle guidance in helping us to navigate the complexities of the comparative interrupted times series design and the associated impact analysis.

1. Implementation Evaluation

1.1 Program Description

In 2010, Iredell-Statesville Schools was awarded an Investing in Innovation grant (i3). The **C**ollaborative **O**rganizational **M**odel to **P**romote **A**ligned **S**upport **S**tructures (COMPASS) is a development grant that seeks to meet the needs of students with disabilities, academically struggling high-needs students, and students with limited English proficiency by providing timely and targeted professional development to the teachers through the alignment of support structures.

The long-term goal of COMPASS is to increase the academic achievement of all students with a focus on students with high-needs, disabilities, and limited English proficiency. To this end, COMPASS provides training to school-based support staff, their Executive Directors, and teachers. By increasing the expertise of support structure staff, COMPASS aims to provide higher quality support to teachers. In turn, this will increase the knowledge, skills, and performance of teachers, which will improve student performance.

In the COMPASS model, presented in Figure 1, alignment is the process of bringing together support structure staff through training and the creation of an online request system through which support staff assistance can be requested. Support structures provide appropriate professional development to teachers as they work to improve the academic achievement of all students, and specifically those students who are struggling or are at-risk for academic failure.

In order to align the support structures, the COMPASS management team developed a series of professional development sessions that target the school-level support structures, their Executive Directors, and faculty. These sessions focused on areas identified by district leadership, including the Common Core, Positive Behavioral Support, SIOP, RtI, AIMSweb, and Progress Monitoring.

Training was offered through several media. North Carolina requires that schools hold six Early Release Professional Development (ERPD) days per year. During this time, the COMPASS management provided professional development. During Year Two, there were a series of COMPASS training sessions offered. These were targeted at support staff and designed to provide them with detailed information on the following content areas: Positive Behavioral Support Overview, SIOP/ESL, RtI Overview, AIMSweb, Curriculum Based Measures, Interventions, and Progress Monitoring. Lastly, in Years Two and Three, the state of North Carolina offered five eight-hour training sessions on Responsiveness to Intervention (RtI). The COMPASS management team added district-specific RtI training beginning in Years 4-5, i3+C3=RtI Success! Workshops. These trainings provided a refresher on information covered by state training.

After the support staff had been trained, they then began to provide support to teachers. This included leading COMPASS training sessions at schools, facilitating ERPD training, and providing individual training as requested. Additionally, support staff participated with the COMPASS management team to offer district level training to faculty and principles on topics such as RtI.

In order to streamline the support request process, the COMPASS management team created an on-line support request system, which was utilized beginning in Year Three. This system allowed school level support staff and principals to place requests for support. These requests could be for instructional support, content area support, or behavioral support. School level support staff could request that members of the COMPASS management team come to their school and provide assistance with data analysis, model teaching, or other support as needed. Principals could schedule whole level school professional development for members of the COMPASS management team or individual faculty support.

Each school utilized six types of data in order to determine if support was needed:

1. End of Grade and End of Course testing. Test results are reviewed at the beginning of the school year by teachers, PLCs, and principals. Scores below a three indicate that the student is eligible for RtI intervention, and that instructional support may be needed.
2. Pre- and post- benchmark assessments. Results are reviewed at least twice per year by PLCs, principals, and IFs. Scores below the 75th percentile on AIMSweb testing indicate a need for RtI intervention and show that instructional support may be needed.
3. Formative assessments. These are reviewed at least every 4.5 weeks by PLCs and IFs. Adequate performance scores will be established by the PLCs. Failure to make adequate performance indicates a need for support staff help.
4. RtI assessments. These are reviewed at least every 15 days by the RtI team. Adequate performance standards will vary by students and treatment plan. Failure to make adequate performance indicates a need for additional RtI intervention services and that instructional support may be needed.
5. Classroom Walk-Throughs. These are conducted periodically by principals.
6. Teacher observations of students while teaching.

A need for content area or behavior support was indicated after a review by PLCs of items 1, 2, 3, 5, and 6 above.[1] If the data indicated that students were struggling, PLCs contacted their school's Instructional Facilitator (IF) to request additional support. If within the IF's scope of expertise, the IF provided support. If not, the IF contacted the appropriate COMPASS support personnel and arranged for him/her to meet with the PLC.

At least twice per month, each school's Leadership Team met to review items 1, 2, 5, and 6. They also determined if a school or individual faculty member was in need of instructional support. After reviewing the data, the Leadership Team completed a Strategic Curriculum and Instruction (CI) form, which outlined a request for support for teachers who were working with struggling students. In order to complete the form, the Leadership Team did the following:

- Identify which content area will be addressed and which support structures are needed to support the content area;
- Determine what type of support is needed (coaching, training, etc.) and will be provided;
- Determine the support delivery method (coaching, courses, training, etc.);
- Determine what, if any, additional professional development needs the support staff have;
- Determine the duration and frequency of the support;
- Establish measurable outcomes and a timetable for data collection;
- Determine, based on analyzed data, whether to continue, adjust or discontinue the support.

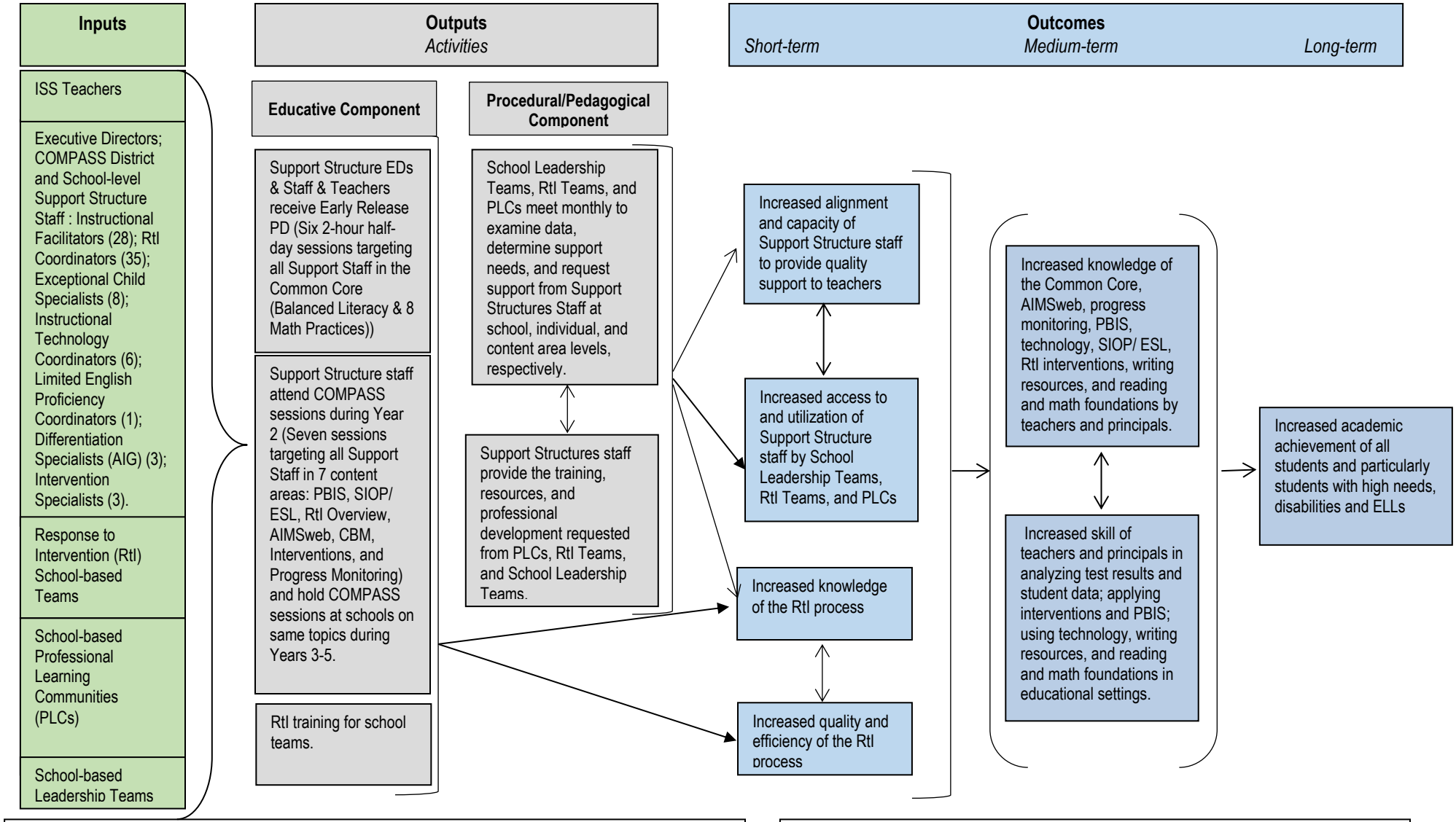
The identified support structures then provided targeted support to teachers in the form of:

- District Level Courses
- Specialist-Delivered Training
- ERPD Content/Experts
- Principal/AP Meetings
- Department Meetings
- COMPASS Meetings
- Online module/resources

Figure 1 presents the COMPASS logic model.

1 PLCs have been in place in ISS schools since 2004. They are based on the DuFour (2007) model, and all teachers participate in at least one PLC that meets once per week for one hour. PLCs consist of at least two or more people, preferably teaching at the same school in the same content area. Each school has common course PLCs in English, Social Studies, Math, Science, Fine Arts, Foreign Language, Health/PE, and Exceptional Children, and, at high schools, ROTC Leadership. Depending on the course offerings at the school, there may be additional PLCs. In PLCs, teachers use data from assessments to evaluate student performance, target areas for improvement, implement strategies for improving instruction, and assess the effectiveness of those strategies.

Figure 1. COMPASS Logic Model



Assumptions

- Support structure request system will be finalized by the end of the Spring 2012 semester.
- Improving the expertise of support structure staff will result in higher quality support for teachers.
- Increased training of support structure staff will lead to higher collaboration between staff and teachers.
- Increasing knowledge, skills, and performance of teachers will result in improved student performance.

External Factors

- Project design hinders the use of a rigorous evaluation design.
- There are few data points available for the SITS design.
- School culture and staff and student receptiveness to change will likely affect implementation.

1.2 Deployment

COMPASS was phased into all district schools during a five-year period.

- Year 1 (2010-2011) Pre implementation phase: Staffing, training, and deployment plans were drafted throughout the year.
- Year 2 (2011-2012) Phase 1: Four schools implemented the program; twelve schools received training throughout the year in preparation for implementing the program in Year 3.
- Year 3 (2012-2013) Phase 2: Sixteen schools implemented the program; fifteen schools received training to implement to program beginning in Year 4.
- Years 4-5 (2013-15) Phase 3: COMPASS was deployed in all district schools.

1.3 Fidelity of Implementation

Fidelity of implementation (FOI) was measured annually from Years Two through Four using schools included in the impact study. The FOI method utilized was an adaptation of the approach described by Century, Rudnick, and Freeman (2010)[2]. Using a critical components framework, they classify program elements into two categories: Structural-Critical components and Instructional-Critical components. Structural Critical components are further subdivided into Procedural and Educative components. Procedural components focus on "the basic steps of the procedures and the ways the intervention are physically organized" (p. 205). As a subset of procedural components, pedagogical components "represent the actions, behaviors and interactions that the user is expected to engage in when enacting the intervention" (p.205). Educative components include "what the user needs to know" and "are analogous to built-in professional development or training" (p. 205). Within this model, FOI is "the extent to which the critical components of an intended program are present when that program is enacted" (Century, Rudnick, & Freeman, 2010, p. 202).

For purposes of this study, the above was modified slightly to focus on two critical components. The Educative and Procedural/Pedagogical components were most applicable to the COMPASS model. The Procedural and Pedagogical components are closely intertwined: each procedural activity has a corresponding pedagogical activity that occurs simultaneously. Therefore, for the COMPASS model, there are two main critical components: Educative and Procedural/Pedagogical.

1.3.1 Critical Components

Through interviews conducted with program staff during Year One of the grant, the critical components of COMPASS' intervention and nine indicators were identified. Because the innovation of COMPASS is its unique alignment model, the FOI evaluation measured alignment of support structures within school-based Leadership Teams and PLCs.

To reduce data collection burden, this study utilized existing documentation, such as meeting minutes and attendance records, where possible. Leadership Teams and PLCs maintained detailed meeting minutes, which included requests for support, the application of COMPASS related training and resources, and when training was provided by COMPASS management staff. Additionally, the district maintained attendance records on training. Yearly targets were established in consultation with district staff.

[2] Century, J., M. Rudnick, and C. Freeman. (2010). A framework for measuring fidelity of implementation: a foundation for shared language and accumulation of knowledge. *American Journal of Evaluation 31*(2), 199-218.

Table 1 shows each critical component and the indicators used to measure fidelity.

Table 1: COMPASS Critical Components and Indicators

Critical Component	Indicators	Indicators	Indicators
Educative	**Early Release PD:** Six two-hour half-day sessions targeting all Support Staff and teachers.	**COMPASS Sessions**: These sessions cover seven content areas (PBIS, SIOP/ESL, RtI, AIMSweb, CBM, Interventions, and Progress Monitoring) and can be held at the district office or at schools. The district sets the number of required sessions yearly.	**RtI training**: Sessions focusing on RtI; can be held by the state or district. State RtI training includes five sessions; district training will vary yearly.
Procedural/Pedagogical	**Alignment within School Leadership Teams:** Make requests for support for faculty from COMPASS support staff. **Documentation of support:** For example, noting in Leadership Team minutes that faculty are applying the training they received from support staff.	**Alignment within PLCs:** Make requests for support in areas such as the Common Core, AIMSweb, etc. from COMPASS support staff. **Documentation of support:** For example, noting in PLC minutes that faculty are applying the training they received from support staff.	

1.4 Method

The Fidelity of Implementation (FOI) study addressed the following research questions:

1. What was the overall level of fidelity of implementation?
2. How much variation in implementation fidelity was there across schools?

1.4.1 Selection Criteria

The FOI evaluation focused on the 21 schools that were included in the impact evaluation.

1.4.2 Fidelity Metric

The evaluation team developed a comprehensive fidelity index that utilized differential weighting. The differential weighting took into account the fact that while COMPASS staff had a great deal of control over their ability to offer trainings (weighted .3), it is not enough to offer the training. Targeted participants must also attend the trainings (weighted. 7).

Table 2: Fidelity Metric.

Component	Indicators of Fidelity		District Target[3]	Weight In School FOI Score
Educative	Early Release PDs	% of sessions held (.3) + % of targeted participants attending (.7)	.8	.20
	Compass Sessions	% of workshops held (.3) + % of targeted participants attending (.7)	.8	.20
	RtI Training	% of sessions held (.3) + % of targeted participants attending (.7)	.9	.10
Procedural/ Pedagogical	Alignment with School Leadership Teams	% of completed requests (.7) + % leadership team agendas noting applied support (.3)	.7	.30
	Alignment with PLCs	% completed requests (.6) + % of PLC agendas noting applied support (.4)	.7	.20

The assessment of FOI indicators varied by phase and by year. During initial implementation, all schools participated in the Educative component (ERPD, COMPASS workshops, and RtI training). However, while they were undergoing training, most schools did not implement the Procedural/ Pedagogical component. An exception to this occurred during Year Two, when Phase I schools participated in training and program implementation. Table 3 shows the critical components that were assessed in each intervention year, by phase.

Table 3: Critical Components Measured, Year by Phase.

	Year 2, 2011-2012		Year 3, 2012-2013		Year 4, 2013-2014	
Phase	Educative	Procedural	Educative	Procedural	Educative	Procedural
1	✓		✓	✓	✓	✓
2	✓		✓	✓	✓	✓
3					✓	✓

1.4.3 Critical Component Scores

The fidelity index allowed for the calculation of separate scores for the Educative and Procedural/Pedagogical critical components. This is useful because schools can meet overall expectations for fidelity, but this does not mean that all components were implemented with equal fidelity.

The fidelity of Educative and Procedural/Pedagogical components were assessed by school**.** If a school earned a score of 0.50 or less, it was said to have met adequate implementation; if a school earned a score of less than 0.50, it was rated as having not met adequate implementation.
To determine the extent of implementation across the program, we assessed the percentage of schools that had met school-level fidelity targets and set a program-wide threshold of 80%. Table 4 shows the district fidelity targets by component.

Table 4: Program Fidelity Thresholds.

[3] Targets and final weight were established by the district.

Critical Component	Threshold for Sample for Implementation with Fidelity	Component Implemented with Fidelity?
Educative	≥80% of schools with a score of .50 or higher.	Yes
Procedural/Pedagogical	≥80% of schools with a score of .50 or higher.	Yes

1.5 Overall Findings

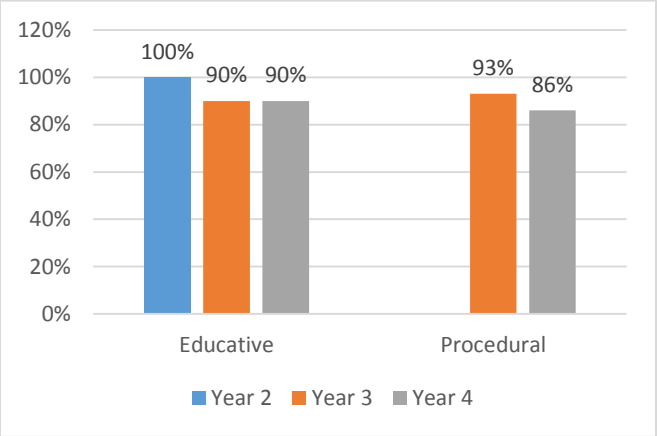


Figure 2: Percentage of COMPASS Schools Meeting Fidelity Targets, by Year.

As shown in Figure 2, Iredell-Statesville schools met the fidelity targets for both the Educative and Procedural/Pedagogical components for all years of the grant. Percentages range from 86% of schools having met the Procedural fidelity targets in Year 3, to having all schools (100%) meeting the Educative targets in Year 2. Table 5 below shows the school-level scores across the educative and procedural components for Years 2-4.

Table 5: School-Level Educational and Procedural/Pedagogical Scores, by Year.

School	Level	Year 2 , 2011-2012		Year 3, 2012- 2013		Year 4, 2013-2014	
		Educative	Procedural	Educative	Procedural	Educative	Procedural
Brawley	MS	x	x	0.58	x	0.84	0.66
Celeste Henkel	ES	0.54	x	1.24	0.70	0.81	0.71
Central	ES	x	x	0.58	x	0.55	0.38
Cool Spring	ES	x	x	0.35	x	0.33	0.71
East Iredell ES	ES	x	x	0.73	x	0.55	0.70
East Iredell MS	MS	x	x	0.83	0.59	0.62	0.47
Harmony	ES	x	x	0.40	x	0.80	0.54
Lake Norman	ES	x	x	0.82	0.66	0.58	0.60
Lakeshore ES	ES	x	x	0.85	0.71	0.52	0.71
Lakeshore MS	MS	x	x	0.70	0.59	0.77	0.61
NB Mills	ES	x	x	1.17	0.69	0.58	0.54
North Iredell	MS	x	x	0.72	0.66	0.66	0.51
Scotts	ES	x	x	0.73	0.71	0.62	0.54
Sharon	ES	0.56	x	0.53	0.68	0.58	0.60
Shepherd	ES	x	x	0.87	x	0.70	0.50
Statesville	MS	x	x	0.71	0.71	0.73	0.67
Third Creek	ES	x	x	0.68	x	0.70	0.52
Troutman ES	ES	0.54	x	0.82	0.71	0.77	0.58
Troutman MS	MS	x	x	0.78	0.64	0.33	0.53
Union Grove	ES	x	x	0.78	0.30	0.95	0.54
West Iredell	MS	0.56	x	0.72	0.68	0.84	0.45

1.5.1 Implementation of the Educative Component

Attendance at trainings was high throughout the program. Schools consistently exceeded the target enrollment; this was especially true of RtI training. The demand was so great for this training that some schools kept waiting lists for faculty. Table 6 shows the number of required attendees, the actual number of attendees, and the percentage attending each type of trainings (ERPD, COMPASS, and RtI) in Years 2-4 combined, by implementation phase.

Table 6: COMPASS, ERPD, and RtI Training Attendance in Years 2-4, by Phase.

Phase	COMPASS Training			ERPD Training			RTI Training		
	N Required	N Actual	Percent	N Required	N Actual	Percent	N Required	N Actual	Percent
1	60	74	123%	394	364	92%	33	105	318%
2	163	162	101%	853	812	95%	100	290	290%
3	159	231	145%	837	1238	148%	46	242	526%
Total	**382**	**467**	**122%**	**2084**	**2414**	**116%**	**179**	**637**	**356%**

1.5.2 Implementation of Procedural/Pedagogical Component

One goal of COMPASS was to streamline the manner in which support is solicited, coordinated, and provided throughout the district. This was assessed by reviewing the meeting minutes from the School Leadership Teams and the PLCs, as well as determining the percentage of on-line requests for support that were actually fulfilled. We found that School Leadership Team and PLC meeting minutes consistently indicated that faculty and principals were regularly analyzing data to determine students' needs and coordinating a plan for delivering or assessing support. Table 7 shows the number of minutes that were analyzed by type and the number (%) that showed the support was actually provided for Years 2-4.

Table 7: Leadership Team and PLC Meeting Minutes Noting Support, by Phase.

Phase	Number of Meeting Minutes Reviewed	Number of Meeting Minutes Noting *Support Provided*	Percent Noting *Support Provided*
Leadership Team Minutes			
1	105	101	96%
2	349	318	91%
3	208	169	81%
Total	**662**	**588**	**89%**
Professional Learning Communities Minutes			
1	949	822	87%
2	1487	1129	76%
3	889	679	76%

The online support request system was utilized extensively during Years 3-4. Table 8 shows the total number of support requests by implementation phase and, of these, the percent of fulfilled requests. COMPASS maintained a 96% fulfillment rate for all years tracked. Some factors that influenced reporting include 1) undocumented yet fulfilled support request and 2) undocumented cancelled support requests.

Table 8: Support Requests, Years 3-4, by Phase.

Phase	Total Support Requested	Total Support Provided	Percent of Fulfilled Requests
1	159	151	95%
2	371	363	97%
3	313	297	95%
Total	**843**	**811**	**96%**

2. Impact Evaluation

The impact evaluation used a short interrupted time-series with comparison group design (C-SITS) to examine the effect of COMPASS on school-level standardized test scores. To address the study's primary confirmatory question, the study produced a combined impact estimate for students in grades 3-8. To address exploratory questions, the study produced impact estimates for students in each grade: 3, 4, 5, 6, 7, and 8. Reading outcomes were measured at the school level before and after implementation of the program in COMPASS schools, and at the same time points in comparison schools.

2.1 Counterfactual

The impact study compared the reading achievement of students in grades 3- 8 attending COMPASS schools with the reading achievement of students from non-COMPASS schools in neighboring school districts sharing similar demographic characteristics. It examined changes in student reading achievement as measured through NC's End-of-Grade Reading test. At comparison schools, students received any instruction that was offered by the district, i.e., "business as usual." Schools served as the unit of analysis.

2.2 Research questions

The overarching research question asks, *"Did COMPASS schools make gains in reading achievement as reflected in their EOG Reading test scores compared to similar non-COMPASS schools?"*

The intervention wass expected to impact all grades 3–8. It would not have been surprising if it affected some grades more than others. If the intervention did affect some grades more than others, our planned approach to estimate a single, combined average impact estimate (specified below) may have masked important variations in impacts between grades. However, we balanced that concern against the concern that if we specified a large number of confirmatory contrasts, we might have faced the prospect of getting spurious results by chance (if we didn't correct for multiple comparisons), or severely reducing the study's power to detect effects (if we did correct for multiple comparisons). We therefore specified a single combined estimate as our single **confirmatory** contrast, C1, and specified a set of six **exploratory** contrasts, E1-E6, that allowed us to explore any variation in impacts among grade levels.

2.2.1 Confirmatory Research Question:

C1. What is the effect of COMPASS on EOG Reading scores on grades 3-8 combined compared to similar non-COMPASS Schools?

2.2.2 Exploratory Research Questions:

E1. Did COMPASS schools make gains in **3rd grade** EOG reading scores compared to similar non-COMPASS schools?

E2. Did COMPASS schools make gains in **4th grade** EOG reading scores compared to similar non-COMPASS schools?

E3. Did COMPASS schools make gains in **5th grade** EOG reading scores compared to similar non-COMPASS schools?

E4. Did COMPASS schools make gains in **6th grade** EOG reading scores compared to similar non-COMPASS schools?

E5. Did COMPASS schools make gains in **7th grade** EOG reading scores compared to similar non-COMPASS schools?

E6. Did COMPASS schools make gains in **8th grade** EOG reading scores compared to similar non-COMPASS schools?

2.3 Design

The impact study used a short interrupted time-series with comparison group design (C-SITS) to examine the effect of COMPASS on school-level standardized test scores. Reading outcomes were measured at the school level before and after implementation of the program in COMPASS schools, and at the same time points in comparison schools. For a full description of the C-SITS design, see Price (2013)[4].

2.3.1 Sample Eligibility

Schools in the impact study had to be a public elementary school or middle school in the state of North Carolina serving grades 3-8 with publicly available ELA state standardized test data for those same grades for at least three successive years prior to the intervention (2007, 2008, and 2009). Alternative schools, charter schools, and high schools were not eligible. Outcomes were measured at the school-level (i.e. there would be school-mean 3rd - 8th grade scores each year). Thus, all 3rd – 8th grade students that normally take the state tests were included.

2.3.2 Sample Selection

COMPASS Schools

Twenty-one of the district's 34 schools operating in 2010 were included in the impact evaluation. All district schools were required to provide the following information to the COMPASS management team:

- a list of current school initiatives,
- the teacher turnover rate,
- current demographics,
- whether or not the school met its AYP,
- the total EOG Math and Reading scores,
- the number of students referred for EC evaluation last year,
- the number of students who qualified for EC.

Schools were also required to provide documentation attesting that 100% of administrators and 75% of staff supported their school's participation in the program. Schools were then placed in Phase I-III depending upon when they successfully completed the above requirements i.e. the first four schools to do so were placed in Phase I and so on.

Research recommends that C-SITS studies have at least three years of pre-intervention data (Bloom, 2003). Four schools were excluded from the study because, having opened in 2008-2009, they had less than three years of baseline data available. Additionally, one school was excluded from the study because no testing information was available, as was the district alternative school. State standardized ELA testing does not go beyond grade 8. Therefore the seven district high schools were not part of the impact study. However, schools that were not part of the impact study still received the intervention.

[4] Price, C. (2013). Research on Educational Effectiveness (SREE) 2013 Spring Conference Workshop: Planning for a Short Interrupted Time Series Design. Workshop materials updated 03-30-2015.

Comparison Schools

Propensity score matching (PSM) was used to construct a well-matched comparison group. Using data available at http://www.ncreportcards.org/src/, the following information was gathered for treatment schools:

1. Enrollment of the school in the grades tested;
2. Economically disadvantaged students (0-20%, 21-40%, 41-60%, 61-80%, 81-100%)[5]; and
3. Number of students tested in grades 3-5 for elementary schools, 6-8 for middle schools.

Using the above information, each treatment school was initially matched with 2-4 non-treatment schools. The data was visually examined and a pool of potential comparison schools was selected based upon how closely they matched treatment schools in terms of size, percentage of economically disadvantaged students, and the number of students tested. A pool of potential comparison schools comprised of 78 schools (53 elementary and 25 middle) was created. Additionally, only schools that had not implemented RtI were chosen as possible comparison schools. RtI status was determined using information sent from the North Carolina Department of Instruction that listed all NC schools that had implemented RtI or were planning to implement RtI.

For PSM, additional information about the percentage of students in each grade (3, 4, and 5 for elementary schools; 6, 7, and 8 for middle schools) who scored proficient on the 2011 End-of-Grade Reading assessment and school's annual yearly progress (AYP) was also gathered for both treatment and the pool of comparison schools.

Logistic regression in SPSS was performed to generate propensity scores. Information on the following variables was used for both comparison and treatment schools:

1. Percentage of economically disadvantaged students (coded on a 1-5 scale with 1=0-20%; 2=21-40%; 3=41-60%; 4=61-80%; 5=81-100%)
2. AYP (0= did not make AYP; 1 did make AYP)
3. Number of students who took Reading EOG by grade
4. Percentage of students who scored proficient on Reading EOG by grade

Nearest neighbor matching with calipers was used to select the two best matches for each treatment school from the full pool of comparison schools. Upper and lower ranges were calculated using .5 (standard deviation of the logit) +/- the propensity score for treatment schools. Comparison schools whose propensity scores were close to treatment schools and fell within the calipers were then selected. The table below shows the propensity scores of the 21 COMPASS schools and the calipers used:

Table 9: Propensity Scores on the 21 COMPASS Schools.

Name of School	Propensity Score	Lower Caliper	Upper Caliper
Elementary Schools (n=14)			
Harmony	0.117530	-1.045930054	1.280990054
C. Henkel	0.157799	-1.005661054	1.321259054
Union Grove	0.233996	-0.929464054	1.397456054
Troutman	0.237346	-0.926114054	1.400806054
Third Creek	0.305289	-0.858171054	1.468749054
Shepherd	0.319451	-0.844009054	1.482911054
Cool Spring	0.341500	-0.821960054	1.504960054
Lakeshore	0.361645	-0.801815054	1.525105054

[5] North Carolina reports percentage of economically disadvantaged students within these ranges and does not provide the exact percentage for each school.

East Iredell	0.385274	-0.778186054	1.548734054
Sharon	0.413357	-0.750103054	1.576817054
Central	0.596236	-0.567224054	1.759696054
Lake Norman	0.618739	-0.544721054	1.782199054
NB Mills	0.660071	-0.503389054	1.823531054
Scotts	0.793858	-0.369602054	1.957318054
Middle Schools (n=7)			
N. Iredell	0.114583	1.313485817	1.542651817
E. Iredell	0.209265	1.042752469	1.461282469
W. Iredell	0.629646	0.622371469	1.881663469
Troutman	0.644522	0.607495469	1.896539469
Statesville	0.697534	0.554483469	1.949551469
Lakeshore	0.814067	0.437950469	2.066084469
Brawley	0.851779	0.400238469	2.103796469

The table below shows the treatment schools and the matched schools' propensity scores. All propensity scores of the two matched comparison schools were within the calipers of each COMPASS school presented above.

Table 10: PSM Scores of COMPASS and Matched Schools.

COMPASS School	COMPASS Propensity Score	Match 1 School	Match 1 Propensity Score	Match 2 School	Match 2 Propensity Score
Elementary Schools (n=14)					
Harmony	0.117530	Wittenburg	0.098588	Banks	0.080922
C. Henkel	0.157799	C.T. Onerton	0.112821	Bell Fork	0.110550
Union Grove	0.233996	Shady Grove	0.154789	Pinewood	0.143278
Troutman	0.237346	Hurley	0.167587	P.Union	0.160529
Third Creek	0.305289	B.J Martin	0.195935	Z.Vance	0.190887
Shepherd	0.319451	Weddington	0.209711	PW Moore	0.201825
Cool Spring	0.341500	Baton	0.233612	Griffth	0.217769
Lakeshore	0.361645	Banoak	0.279315	Benhaven	0.274345
East Iredell	0.385274	Poe	0.286560	Holly Ridge	0.280508
Sharon	0.413357	Vienna	0.314833	Beverlyhill	0.309886
Central	0.596236	Bostian	0.408107	Whitaker	0.353141
Lake Norman	0.618739	O.Richmond	0.425859	Lewisville	0.411558
NB Mills	0.660071	C. Tuttle	0.595231	A. Springs	0.430991
Scotts	0.793858	Forest Park	0.641486	North Rowan	0.633625
Middle Schools (N=7)					
N. Iredell	0.114583	G.C. Hawley	0.021568	Woodington	0.004249
E. Iredell	0.209265	Wiley	0.040147	Hobbton	0.003253
W. Iredell	0.629646	West Rowan	0.091543	W.Alexander	0.075094
Troutman	0.644522	Hopes Mill	0.110113	C.Stanford	0.101179
Statesville	0.697534	Williamston	0.350695	C. Campus	0.118863
Lakeshore	0.814067	G. Culbreth	0.397155	E.B. Aycock	0.366775
Brawley	0.851779	Zebulon	0.682245	Warren Co.	0.536932

The final pool of comparison 42 schools was comprised of 28 elementary schools and 14 middle schools, matched with each of the 21 COMPASS Schools at a 2:1 ratio. There are two comparison schools for each treatment school.

COMPASS was phased in according to the following schedule:

- Phase 1 schools (n=4) were trained in 2011-12 and began implementation in that same year;
- Phase 2 schools (n=10) were also trained in 2011-12 but began implementation in 2012-13;
- Phase 3 schools (n=7) were trained in 2012-13 and began implementation in 2013-14.

It should be noted that for Phase II and Phase III schools, training years are pre-implementation years. For Phase I schools, training and implementation occurred simultaneously during the 2011-2012 year.

Table 11 below shows the matched schools with the treatment schools according to the phase of implementation.

Table 11: Matching by Implementation Phase.

COMPASS	Comparison	
Phase I		
Celeste Henkel Elementary	Carolle T. Onerton Elementary	Bell Fork
Sharon Elementary	Vienna Elementary	Beverlyhill Elementary
West Iredell Middle School	West Rowan Middle	West Alexander Middle
Troutman Elementary	Hurley Elementary	Pleasant Union Elementary
Phase II		
East Iredell Middle	Wiley Middle	Hobbton Middle
Lake Norman Elementary	Old Richmond Elementary	Lewisville Elementary
Lakeshore Elementary	Banoak Elementary	Benhaven Elementary
Lakeshore Middle	Grey Culbreth Middle	E.B. Aycock Middle
NB Mills Elementary	Charles Tuttle Elementary	Aurelian Springs
North Iredell Middle	G C Hawley Middle	Woodington Middle
Scotts Elementary	Forest Park Elementary	North Rowan Elementary
Statesville Middle	Williamston Middle	Centennial Campus Middle
Troutman Middle	Hopes Mill Middle	Charles W. Stanford Middle
Union Grove Elementary	Shady Grove	Pinewood Elementary
Phase III		
Brawley Middle	Zebulon Middle	Warren County Middle
Central Elementary	Bostian Elementary	Whitaker Elementary
Cool Spring Elementary	Baton Elementary	Griffith Elementary
East Iredell Elementary	Poe Elementary	Holly Ridge Elementary
Harmony Elementary	Wittenburg	Banks Elementary
Shepherd Elementary	Weddington	PW Moore Elementary
Third Creek Elementary	Benjamin J Martin	Zebulon Vance Elementary

2.4 Measure

North Carolina's End-of-Grade (EOG) Reading test (ABCs Reading) uses a standardized scale for scoring that is consistent in grades 3-5 and 6-8. The test consists of eight sections for grades 3-5 and nine sections for grades 6-8. It measures the goals and objectives specified in the 2004 NC English Language Arts Standard Course of Study, and it assesses comprehension and vocabulary.

2.4.1 Conversion of School-level Standard Scores to Z-Scores

In considering how to combine data across grades, our overarching question is whether the data from different grades are measured in the same scale. Data from different grades may have different means and different standard deviations. In the current study, all data are from a single state, and the tests are

supposed to be equated across grades, so re-scaling might not have been necessary. However, preliminary results indicated that the standard deviations decrease with increasing grade levels. Moreover, the state of NC rescaled their EOG Reading test in 2012-13, the final two years of post-intervention data, making the conversion of standard scores to Z-scores clearly necessary. Furthermore, the re-scaling approach ensures that the combined impact estimate, which represents a weighted mean of the impact estimates across each grade, will be a combination of impact estimates that each can be interpreted relative to the standard deviation of school-level test scores at each grade.

The re-scaling approach converts EOG school-level scaled scores to Z-scores at each grade level. This analysis uses only school-level data. We do not have student-level data and therefore cannot calculate student-level standard deviations. The Z-scores are standardized relative to school-level within-grade means and standard deviations.

An overview of the steps is as follows:

- Calculate a standard deviation of school-means for each year/grade level;
- Calculate the mean of school-means for each year/grade level,;
- Standardized the observed scores at each grade level by subtracting the grade-level mean and dividing by the standard deviation.

2.4.2 Data Collection

In the fall of each year, the evaluators compiled EOG test scores (for COMPASS and comparison schools) as they become publicly available on the North Carolina Department of Instruction's website. Data from the previous school year was typically released to the public between September and December of the following school year.

As indicated by the heavy blue line in the table below, for Phase 1 schools, data was collected across four pretreatment years, 2008-11, and three treatment years, 2012-2014; Phase II schools had five pretreatment years (2008-12) and two treatment years (2013-14); and Phase III schools had six pretreatment (2008-13) years and one treatment year (2004).

Table 12: Treatment Years and Pre-treatment Years, COMPASS and Comparison Schools.

Phase & Type of School (COMPASS or Comparison)	Spring 2008	Spring 2009	Spring 2010	Spring 2011	Spring 2012	Spring 2013	Spring 2014	N of Ele. Schools	N of Middle Schools
Phase I COMPASS	x	x	x	x	T	T	T	3	1
Phase I Comparison	x	x	x	x	t	t	t	6	2
Phase II COMPASS	x	x	x	x	x	T	T	5	5
Phase II Comparison	x	x	x	x	x	t	t	10	10
Phase III COMPASS	x	x	x	x	x	x	T	6	1
Phase III Comparison	x	x	x	x	x	x	t	12	2

All reading scores come from assessments administered in the spring of the school year and reported in the fall of the following school year.
"x": indicates a pre-treatment year when a school-level 5th grade reading outcome score was obtained
"T": For COMPASS schools T indicates a treatment year
"t": For comparison schools, "t" indicates a year when the schools' treatment group counterparts have received treatment.

2.5 Analysis

2.5.1 Baseline Sample Size

There were 21 COMPASS schools and 42 comparison schools for a total sample of 63 schools. Forty-two schools were elementary schools (14 COMPASS, 28 Comparison) serving impact grades 3-5, and 21 were middle schools serving served grades 6-8 (7 COMPASS, 14 Comparison).

The characteristics of the sample at baseline are presented in Table 13 below.

Table 13: Characteristics of the Sample Schools at Baseline.

Characteristic	Comparison			COMPASS			Total		
	Mean	N	*sd*	Mean	N	*sd*	Mean	N	*sd*
Economically Disadvantaged	3.21	42	1.33	3.24	21	1.22	3.22	63	1.28
Met AYP (%)	21.0	42	.na	24.0	21	na	22.0	63	na
Enrollment in tested grades	368.07	42	174	363.90	21	182.96	366.68	63	175.94
% Proficient in tested grades	70.12	42	13.52	68.88	21	12.47	69.70	63	13.09

2.5.2 Analytic Sample Size

All data and analyses occurred at the school level. Because we used extant school-level data for the analyses, there was no way for a school to drop out of the study. We did not expect any attrition of schools from the study unless schools closed or merged with other schools. This did not happen. The analytic sample is identical to the baseline sample. It is not possible in this design to measure attrition at the student level.

2.5.3 Matching Blocks as Covariates

The propensity matching procedure used a nearest neighbor with caliper matching strategy to match two comparison schools to each treatment school. Rather than using an analysis model that conditions on the four variables that were used in the matching process, we included dummy variables for matching blocks.

2.5.4 Analytic Approach

For estimating the impact of COMPASS school-mean Reading achievement for a single grade (e.g. 5th grade) we fit two level hierarchical linear models with repeated observations over time (level-1) nested in schools (level-2). The model was a linear baseline trend model (Bloom 2003)[6] with random intercepts and slopes for schools, and allowed for different pre-treatment slopes for treatment and comparison schools. The model was essentially the same model as the "CITS model" described in Sommers et al. (2012)[7] except that our model also included dummy variables to represent the matching blocks. Additionally, our overall treatment effect estimate was calculated as the average of the treatment effect estimates from the three post-intervention years. Models were fit to the data using the mixed procedure SPSS with a repeated statement and an AR(1) option to account for potential autocorrelation assuming an autoregressive-1 structure which assumes that observations with a school that are closer together in time may be more correlated with one another than observations that are further apart in time.

[6] Bloom, H. S. (2003). Using "Short" Interrupted Time-Series Analysis To Measure The Impacts Of Whole-School Reforms: With Applications to a Study of Accelerated Schools. Evaluation Review, 27(3), 3-49.

[7] Somers, M. a., Zhu, P., Jacob, R., & Bloom, H. (2012) . The Validity and Precision of the Comparative Interrupted Time Series Design and the Difference-in-Difference Design in Educational Evaluation. New York, NY: MDRC..

Level 1 (school years within schools)

$$
\begin{aligned}
&Y_{jk} = \\
&\alpha_{0k} + \delta_{1k}(Time_{jk}) + \phi(TrtGrp_{jk}) + \gamma(TrtGrp_{jk} * Time_{jk}) \\
&+ \beta_1(I1_{jk}) + \beta_2(I2_{jk}) + \beta_3(I3_{jk}) \\
&+ \beta_4(TrtGrp_{jk} * I1_{jk}) + \beta_5(TrtGrp_{jk} * I2_{jk}) \\
&+ \beta_6(TrtGrp_{jk} * I3_{jk}) \\
&+ \sum_{m=1}^{M} \lambda_m(Block_m) \\
&+ \varepsilon_{jk}
\end{aligned}
$$

Level 2 (schools)

$$\alpha_{0k} = \alpha_{00} + r_{0k}$$

$$\delta_{1k} = \delta_{10} + r_{1k}$$

Where:

Y_{jk}	is the j^{th} observation on school *k*, (outcomes were z-scored prior to analysis as described elsewhere.)
$Time_{jk}$	= -4, -3, -2, -1 for four, three, two, or one year pre-treatment, = 0, 1, or 2 for one, two or three years post-treatment.
$TrtGrp_k$	= 1 if school is an intervention (treatment) school; = 0 if comparison school
$I1_{jk}$	=1 if first post-treatment year; = 0 else
$I2_{jk}$	=1 if second post-treatment year; = 0 else
$I3_{jk}$	=1 if third post-treatment year; = 0 else
$TrtGrp_k$	= 1 if school is an intervention (treatment) school; = 0 if comparison school
$Block_m$	=1 of school is in the m^{th} of M matching blocks; = 0 else
r_{0k}	= Between school variation in pre-treatment intercepts, assumed distributed $N(0, \tau_{00}^{\ 2})$
r_{1k}	= Between school variation in pre-treatment slopes, assumed distributed $N(0, \tau_{11}^{\ 2})$
ε_{jk}	= residual for j^{th} observation on school *k*, assumed distributed $N(0, \sigma^2)$

The overall impact estimate and its standard error were calculated using a /TEST option in the SPSS mixed procedure, which calculated the average of $\hat{\beta}_4, \hat{\beta}_5$ and $\hat{\beta}_6$.

To estimate impacts of the intervention on schools for all grades combined, a three level hierarchical linear model was used with repeated observations over time (level-1) nested in grades (level-2), grades nested in schools (level-3). Other than the three level hierarchical structure, the model was identical to that described for a single grade.

2.6 Findings

2.6.1 Baseline Equivalence Testing

We used two-level and three-level models to test for baseline equivalence, as follows:

Baseline Balance Testing for Confirmatory Contrast 1: Impact Estimated for Grades 3-8 Combined:

Level-1 Model: Time Level

$$Z_{iGj} = \beta_{0Gj} + \varepsilon_{iGj}$$

Level-2 Model: Grade Level

$$\beta_{0Gj} = \pi_{0j} + r_{Gj}$$

Level-3 Model: School Level

$$\pi_{0j} = \gamma_{00} + \gamma_{01}(T_j) + \sum_{m=1}^{M-1} \gamma_{0(m+1)}(MatchingBlock_m) + \mu_{0j}$$

Baseline Balance Testing for Exploratory Contrast E1-E6: Impact Estimates for a Single Grade Level:

Level-1 Model: Time Level

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

Level-2 Model: School Level

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(T_j) + \sum_{m=1}^{M-1} \gamma_{0(m+1)}(MatchingBlock_m) + \mu_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

Where:

Y_{ij}	= the reading score at time *i* at school *j*;
T_j	= 1 if school *j* is an intervention school, and 0 if comparison;
ε_{ij}	= the random effect representing the difference between score at year *i* for school *j* and the predicted mean score for school *j*. These residual effects are assumed normally distributed with mean 0 and variance σ^2, and independent from μ_{0j}.
μ_{0j}	= the deviation of school j's intercept from the grand mean intercept distributed with mean 0 and variance τ^2;
γ_{01}	= the difference between the baseline mean scores of treatment and comparison schools. The test of the null hypothesis that γ_{01} is equal to zero is a test of baseline equivalence of the two sets of schools.
$MatchingBlock_m$	=1 if school was in the m^{th} of *M* matching blocks, 0 otherwise

All other terms had been defined in the previous section.

2.6.2 Baseline Equivalence Results

The differences between the COMPASS and comparisons schools at baseline on Z-scores, computed from the school-level EOG ELA scaled scores, was minimal for each contrast and no difference was significantly different from zero. Differences ranged from -0.031 (E3, Grade 5) to 0.106 (E2, Grade 4). The treatment-comparison difference for at baseline C1, which measured impact across grades 3-8 combined, was -0.034. We conclude that COMPASS and comparison schools are equivalent at baseline, at least on this outcome measure.

Table 14: Baseline Equivalence of the COMPASS and Comparison Schools, by Contrast.

Contrast ID #	Contrast Grades	COMPASS N of Schools	Comparison N of Schools	Unadjusted Pooled *sd*[a]	Treatment – Comparison Difference	p-value
C1	3-8	21	42	1.0	-.034	.88
E1	3	14	28	1.0	-.056	.85
E2	4	14	28	1.0	.106	.58
E3	5	14	28	1.0	-.031	.90
E4	6	7	14	1.0	.071	.87
E5	7	7	14	1.0	.084	.85
E6	8	7	14	1.0	.088	.84

[a] **Pooled sd using Z-scores.** Outcome data were converted to Z-scores prior to analysis. Z-scores were standardized relative to within grade sample means and standard deviations of all treatment and comparison school data from all pre-treatment (baseline) years. Therefore, the standard deviation of the school-level Z-scores for each grade and for all grades combined (for pre-treatment years) is 1.0.

2.6.3 Impact Results

We tested for between-group differences at posttest using the model described in section 2.5.4, which tests for differences in Z-scores averaged across the three post intervention years and accounts for random slopes and random intercepts across the schools.

All impact estimates except for the single grade estimate for 8th graders was in the expected positive direction. The one confirmatory contrast tested the combined effect of COMPASS across grades 3-8, produced an impact estimate of 0.39, p<.001. That is, when averaged across three years, and combined across grades levels, COMPASS schools scored 0.39 of a standard deviation higher than similar comparison schools across those same grade levels. Because Z-scores have a *sd* of 1.0, the effect size is computed as .39/1.0 = 0.39, which according to Cohen (1988)[8] represents a medium effect. However, this is school-level data and the impact at the individual student level is likely to be somewhat smaller.

We also found positive and statistically significant impacts within grade 4, grade 5, and grade 7, with estimates ranging from 0.42 (grade 4) to 0.64 (grade 5). Impact estimates for grade 3 and grade 6 were in the intended direction but did not reach statistical significance. As mentioned, the impact estimate for the eighth grade was in the negative direction (that is, COMPASS achievement was lower than comparison schools) but the estimate of -.17, was not statistically different from zero. See Table 15.

Table 15: Summary of Impact Results, by Contrast.

[8] Cohen, J (1988) Statistical Power Analysis for the Behavioral Sciences (second ed.). Lawrence Erlbaum Associates. NY.

Contrast ID #	Contrast Grades	Treatment Group N of Schools	Comparison Group N of Schools	Pooled *sd*[a]	Impact Estimate	Impact Standard Error	p-value	Degrees of Freedom
C1	3-8	21	42	1.0	.39	.09	.001	808
E1	3	14	28	1.0	.28	.18	.126	138
E2	4	14	28	1.0	.42	.21	.043	117
E3	5	14	28	1.0	.64	.21	.003	135
E4	6	7	14	1.0	.40	.27	.155	53
E5	7	7	14	1.0	.57	.26	.031	55
E6	8	7	14	1.0	-.17	.25	.493	52

[a] **Pooled sd using Z-scores.** Outcome data were converted to z-scores prior to analysis. Z-scores were standardized relative to within grade sample means and standard deviations of all treatment and comparison school data. Therefore, the standard deviation of the school-level Z-scores for each grade and for all grades combined is 1.0.

As mentioned, the impact estimates can be interpreted relative to the standard deviation of school-level test scores at each grade. To do this, we computed the average *sd* of the ELA scaled scores across all post-intervention years, for grades 3-8 combined and separately for each grade, then multiplied the *sd* by the impact estimate to approximate the impact of COMPASS on the actual school-level ELA scaled scores. For grades 3-8, COMPASS schools improved by 3.65 scaled scores over comparison schools; in 4th grades, by 3.80 scaled scores; in 5th grades, by 5.52 scaled scores; and in 7th grades, by 5.69 scaled scores. Estimates for other grades are not reported computed because the actual impact estimates were not significantly different from zero. See Table 16.

Table 16: Estimated Difference in ELA Scaled Scores in COMPASS Schools vs. Comparison Schools, by Contrast.

Contrast ID #	Contrast Grades	Mean ELA Scaled Score across posttreatment years	Mean *sd* of ELA Scaled Scores across posttreatment years	Standardized Impact Estimate from Table 15	p-value	Estimated Difference in School-Level ELA Scores in COMPASS Schools v. Comparison Schools
C1	3-8	438.56	9.38	.39	.001	3.65
E1	3	428.45	9.53	.28	.126	NS
E2	4	433.60	9.07	.42	.043	3.80
E3	5	437.83	8.64	.64	.003	5.52
E4	6	444.95	9.80	.40	.155	NS
E5	7	448.03	9.99	.57	.031	5.69
E6	8	450.93	9.93	-.17	.493	NS

2.7 Discussion

The impact evaluation found 1) a positive overall effect, when effects were averaged across three years, and combined across grades levels, and 2) a positive effect within specific grades, namely grades 4, grade 5, and grade 7. Impact estimates for grade 3 and grade 6 were in the intended direction but did not reach statistical significance, while the impact estimate for the eighth grade was in the negative direction yet not statistically different from zero.

While these results are persuasive, the impact study does suffer from an N=1 confound. That is, because COMPASS was only implemented in one school district, we cannot disentangle to effect of the ISS school district from the effect of COMPASS. For example, there is the possibility that perhaps ISS implemented policies or dedicated additional resources to improving reading instruction in grades 3-8 at the same time

COMPASS was being implemented. Without having another school district delivering the program, it is impossible to rule out the possibility that something other than COMPASS, yet endogenous to ISS, was influencing the results.

One obvious question is to why COMPASS had differential effects across grade levels. If support structures and procedural processes were operating equally at the same level of efficiency across grade levels, then we would expect to see little difference in the impact at each grade. However, although all impact estimates for each grade were in the intended direction, only impacts in three of the six grades reached statistical significance. Of note, two of these were in the middle school, grades 6 and grade 8. It is possible that there was differential levels of implementation between grades, particularly at the middle school, that could account for these results but since implementation by grade level was not tracked in the FOI evaluation there is no way to confirm or disconfirm this possibility. Administrators at ISS were equally at a loss to offer any plausible explanation.

Nevertheless, these results demonstrate that with the proper training and support to critical support staff, procedural processes properly aligned to streamline the identification referral, and delivery of student services, real positive changes can occur.