



Springfield-Chicopee School Districts Striving Readers (SR) Program

222 Richmond Street, Suite 300
Providence, Rhode Island 02903-4226

Phone: 401.274.9548

Fax: 401.421.7650

E-mail: information@alliance.brown.edu

Web: www.alliance.brown.edu

Final Report Years 1–5: Evaluation of Implementation and Impact

Final March 2012

Prepared by:
The Education Alliance at Brown University

Springfield-Chicopee School Districts Striving Readers (SR) Program

Final Report Years 1-5: Evaluation of Implementation and Impact

Final March 2012

Prepared for:

Office of Elementary and Secondary Education, U.S. Department of Education
Institute of Education Sciences, U.S. Department of Education

Prepared by:

Research & Evaluation Division
The Education Alliance at Brown University



THE EDUCATION ALLIANCE at Brown University

EQUITY AND EXCELLENCE FOR ALL SCHOOL

Since 1975, The Education Alliance, a department at Brown University, has helped the education community improve America's schools. We provide applied research, technical assistance, and informational resources to connect research and practice, build knowledge and skills, and meet critical needs in the field.

With offices located in Providence, Rhode Island, adjacent to the Brown University campus, and a dedicated team of skilled professionals, collaborators, and partners, we provide services and resources to K–12 schools and districts across the country and beyond. As we work with educators, we customize our programs to the specific needs of our clients.

Our Web site (www.alliance.brown.edu) describes our work and provides extensive information and resources about education reform. Information about all Alliance programs and services is available by contacting:

**The Education Alliance
at Brown University
4 Richmond Square
Providence, RI 02906**

**Phone: 800.521.9550
Fax: 401.421.7650
E-mail: information@alliance.brown.edu
Web: www.alliance.brown.edu**

Report Authors: Kimberley Sprague, Colleen Zaller, Anita Kite, Karen Hussar

ACKNOWLEDGEMENTS

The authors recognize the “above-and-beyond” contributions and commitment of our partners in the Springfield and Chicopee Public School Districts (administrators, teachers, and program staff). A special thanks, in particular, to the phenomenal Striving Readers District Implementation Team who worked tirelessly to ensure this study would contribute to the research base in the field of education. This study has benefited from the commitment and energy of this team of Matt Rigney and Justin Hurst, as well as former team members Ann Ferriter and Sheila Hoffman, who facilitated the implementation of the three interventions in this project with the research always in mind, facilitated access to classrooms and teachers as well as school and district staff for interviews, and responded willingly to requests for data documenting their work.

The authors acknowledge the significant contributions of the Project Officer Marcia Kingman at the Office of Elementary and Secondary Education; Stefanie Schmidt at the Institute for Education Sciences; Barbara Goodson, Cris Price, and Beth Boulay at Abt Associates, Inc. (all Abt technical assistance team members); and Julie Meltzer at the Public Consulting Group, Inc. In addition, the authors are grateful for the substantial contributions made in the past by Jennifer Borman, Sarah Cussler, Joan Ford, Chandra Haislet, Leslie Nevola, Bob St. Pierre, Hardeek Shah, Cynthia Way, Ryoko Yamaguchi, and Ivana Zuliani. Thanks also go to our former colleagues for contributions this year: Deborah Collins and Laurie Phillips. Any omission of acknowledgement is solely the responsibility of the authors.

TABLE OF CONTENTS

Executive Summary	i
<i>Implementation</i>	i
Targeted Interventions: Inputs, Classroom Model, and Context	ii
READ 180: Implementation Ratings	iii
Xtreme Reading: Implementation Ratings	v
Whole-School Intervention: Inputs, Classroom Model, and Context	vii
<i>Impact</i>	ix
Targeted Interventions Impacts	x
Targeted Interventions Impact and Classroom Implementation	xi
READ 180 Classroom Implementation and Impact	xi
Xtreme Reading Classroom Implementation and Impact	xii
Implementation Patterns as Predictor	xiv
Whole-School Intervention Impact	xiv
Whole-School Impact and Implementation	xvi
<i>Overall Summary</i>	xvii
I. Introduction and Study Background	1
II. District Context	2
<i>Characteristics of Districts and Student Population</i>	2
Adequate Yearly Progress (AYP) Status	3
III. Theoretical Rationale and Description of Interventions	5
<i>READ 180 Targeted Intervention</i>	5
READ 180: Instructional Approach and Curriculum	6
READ 180: Over Time	7
<i>Xtreme Reading Targeted Intervention</i>	10
Xtreme Reading: Instructional Approach and Curriculum	11
Xtreme Reading: Over Time	12
<i>Whole-School Intervention</i>	15
SIM-CERT: Instructional Approach and Learning Strategies	15
SIM-CERT Inclusion Criteria	19
SIM-CERT: Over Time	20
IV. Evaluation of the Implementation of the Targeted Interventions	24
<i>Targeted Implementation Research Questions and Methods</i>	24
<i>Targeted Implementation Teachers</i>	25
Characteristics of Teachers: Prior Study Participation	26
Characteristics of Teachers: Over Time and Across Groups	27
<i>Business as Usual</i>	29
Contamination of Control Condition	30
V. Targeted Interventions: Results and Implications	32
<i>Targeted Implementation Components</i>	32
Targeted Implementation Component Ratings	33
<i>Targeted Implementation Overall Ratings</i>	34
READ 180: Implementation Ratings	35

Xtreme Reading: Implementation Ratings	37
<i>Targeted Intervention Implications: What Ratings Do Not Illuminate</i>	39
READ 180 Inputs	39
READ 180 Classroom Model	44
Xtreme Reading Inputs	49
Xtreme Reading Classroom Model	53
Cross-Targeted Intervention Barriers	57
VI. Evaluation of the Impacts of the Targeted Interventions	61
<i>Measures, Screening, and Random Assignment</i>	62
Screening as Planned	62
Randomization Process as Planned	64
<i>Final Sample</i>	66
Student Screening and Random Assignment	66
<i>Intent-to-Treat</i>	68
<i>Power to Detect Effects</i>	68
<i>Statistical Analyses</i>	70
Analytic Model and Specifications	71
<i>Analytic Sample</i>	72
<i>Impacts on Students</i>	74
VII. Targeted Intervention Impacts and Implementation	78
READ 180 Classroom Implementation and Impact	79
Xtreme Reading Classroom Implementation and Impact	80
Implementation Patterns as Predictor	81
VIII. Evaluation of the Implementation of the Whole-School Intervention	83
<i>Whole-School Research Questions and Methods</i>	83
<i>Whole-School Implementation Teachers</i>	84
Selection of SIM-CERT Teachers	84
Characteristics of SIM-CERT Teachers: Over time	85
<i>Whole-School Implementation Coaches</i>	87
Characteristics of SIM-CERT Coaches: Over Time	87
IX. Whole-School Intervention Implementation: Results and Implications	88
<i>Whole-School Implementation Components</i>	88
Professional Development	88
Professional Development Context	89
<i>Professional Development Ratings</i>	90
Number of Days in Attendance	90
Professional Development Training Ratings Context	93
Receipt of Training in Specific SIM-CERT Routines	96
Professional Development Ratings Context	99
<i>Classroom Implementation Ratings</i>	100
Classroom Implementation Rating Context	103
<i>Whole-School Intervention Implications: What Ratings Do Not Illuminate</i>	108
Intervention and Implementation Specifications	109
Professional Development Scheduling and Recruitment	111
Support and Accountability	113
Satisfaction with Professional Development	118

X. Whole-School Intervention Impacts	123
<i>Analytic Sample</i>	123
<i>Statistical Analyses</i>	125
Analytic Model and Specifications	125
<i>Whole-School Impact</i>	127
<i>Impact Results Summary</i>	130
XI. Whole-school Intervention Impact and Implementation	132
<i>Levels of Implementation</i>	132
<i>Analytic Sample</i>	133
<i>Statistical Analyses</i>	134
Analytic Model and Specifications	135
<i>Impact and Implementation Results Summary</i>	136
Between School Results	136
Within School Results	138
<i>Whole-School Impact and Implementation Summary</i>	140
XII. Evaluation Summary	142
References	144

LIST OF EXHIBITS

EXHIBIT 1. SUMMARY READ 180 INPUT RATINGS YEARS 1-5 (N = 14)	iii
EXHIBIT 2. SUMMARY READ 180 CLASSROOM MODEL RATINGS YEARS 1-5 (N = 14)	iv
EXHIBIT 3. SUMMARY XTREME READING INPUT RATINGS YEARS 1-4 (N = 11)	v
EXHIBIT 4. SUMMARY XTREME READING CLASSROOM RATINGS YEARS 1-4 (N = 11)	vi
EXHIBIT 5. PROFESSIONAL DEVELOPMENT DAYS REQUIRED: PERCENT OF TEACHERS RECEIVING ADEQUATE RATINGS BY DISTRICT AND COHORT	vii
EXHIBIT 6. PERCENTAGE OF TEACHERS WHO RECEIVED ADEQUATE LEVELS OF TRAINING IN THE REQUIRED ROUTINES FOR THE FIRST YEAR OF IMPLEMENTATION	viii
EXHIBIT 7. PERCENTAGE OF TEACHERS WHO MET AND EXCEEDED MINIMUM REQUIREMENTS FOR CLASSROOM MODEL IMPLEMENTATION	ix
EXHIBIT 8. IMPACT OF READ 180 BY LEVEL OF CLASSROOM IMPLEMENTATION (YEARS 1-5)	xii
EXHIBIT 9. IMPACT OF XTREME READING BY LEVEL OF CLASSROOM IMPLEMENTATION (YEARS 1-5)	xiii
EXHIBIT 10. CHARACTERISTICS OF PARTICIPATING SCHOOLS (2010-11)	3
EXHIBIT 11. AYP DETERMINATION FOR ELA BY DISTRICT (2006-10)	4
EXHIBIT 12. READ 180 LOGIC MODEL	9
EXHIBIT 13. SIM CONTENT LITERACY CONTINUUM (CLC)	10
EXHIBIT 14. XTREME READING LOGIC MODEL	14
EXHIBIT 15. SIM CONTENT ENHANCEMENT ROUTINES FOR TEACHING (SIM-CERT)	16
EXHIBIT 16. SIM-CERT LOGIC MODEL	18
EXHIBIT 17. SIM-CERT DELIVERY OF PROFESSIONAL DEVELOPMENT (AS PLANNED, YEARS 1-4)	21
EXHIBIT 18. INTERVENTION TEACHING EXPERIENCE BY YEAR (AS PLANNED, YEARS 1-5)	26
EXHIBIT 19. AVERAGE YEARS OF TEACHING EXPERIENCE ACROSS STUDY YEARS, BY GROUP	27
EXHIBIT 20. PERCENTAGE OF TEACHERS WITH HIGHEST DEGREE AND CERTIFICATION BY GROUP	28
EXHIBIT 21. SUMMARY READ 180 INPUT RATINGS YEARS 1-5 (N = 14)	35
EXHIBIT 22. SUMMARY READ 180 CLASSROOM MODEL RATINGS YEARS 1-5 (N = 14)	36
EXHIBIT 23. SUMMARY XTREME READING INPUT RATINGS YEARS 1-5 (N = 11)	37
EXHIBIT 24. SUMMARY XTREME READING CLASSROOM RATINGS YEARS 1-5 (N = 11)	38
EXHIBIT 25. SRI RANGES FROM NORMS FILE: UNPUBLISHED DATA PROVIDED BY SCHOLASTIC	63
EXHIBIT 26. PROCESSES FOR THE FINAL RANDOMIZATION (NINTH-GRADE SCREENING TEST)	65
EXHIBIT 27. SCREENING AND ASSIGNMENT AND SAMPLE	67
EXHIBIT 28. FINAL NUMBERS OF THE INTENT-TO-TREAT RANDOMLY ASSIGNED STUDENTS BY SCHOOL	68
EXHIBIT 29. MDES FOR PAIR-WISE COMPARISONS: BY N OF STUDENTS AND COVARIATE	69
EXHIBIT 30. STUDENT SAMPLE CHARACTERISTICS BY DISTRICT: PRE-AND POST-TEST SAMPLE	72
EXHIBIT 31. STUDENT SAMPLE CHARACTERISTICS BY TREATMENT: PRE- AND POST-TEST SAMPLE	73
EXHIBIT 32. MEAN STUDENT READING ACHIEVEMENT SCORES BY GROUP (SDRT-4 SCALED SCORES)	75
EXHIBIT 33. IMPACT OF INTERVENTION ON STUDENT READING ACHIEVEMENT BY GROUP (SDRT-4 NCE SCORES)	76
EXHIBIT 34. IMPACT OF READ 180 BY LEVEL OF CLASSROOM IMPLEMENTATION (YEARS 1-5)	79
EXHIBIT 35. IMPACT OF XTREME READING BY LEVEL OF CLASSROOM IMPLEMENTATION (YEARS 1-5)	80
EXHIBIT 36. SIM-CERT TEACHER RATES OF CERTIFICATION AT THE PROFESSIONAL LEVEL	86
EXHIBIT 37. SIM-CERT TEACHER AVERAGE NUMBER OF YEARS OF TEACHING EXPERIENCE	86
EXHIBIT 38. SIM-CERT TRAINING: NUMBERS OF TEACHERS ATTENDING ANY TRAINING THAT OCCURRED	89
EXHIBIT 39. PROFESSIONAL DEVELOPMENT DAYS REQUIRED: PERCENT OF TEACHERS RECEIVING ADEQUATE RATINGS BY DISTRICT AND COHORT	91
EXHIBIT 40. SPRINGFIELD SIM-CERT TRAINING: DELIVERY OF PROFESSIONAL DEVELOPMENT	94
EXHIBIT 41. CHICOPEE SIM-CERT TRAINING: DELIVERY OF PROFESSIONAL DEVELOPMENT	95

EXHIBIT 42. REQUIRED AND RECOMMENDED CONTENT FOR SIM-CERT TRAININGS	97
EXHIBIT 43. PERCENTAGE OF TEACHERS WHO RECEIVED ADEQUATE LEVELS OF TRAINING IN THE REQUIRED ROUTINES FOR THE FIRST YEAR OF IMPLEMENTATION	98
EXHIBIT 44. CLASSROOM MODEL RATINGS BY DISTRICT ACROSS YEARS 2, 3, 4, AND 5	102
EXHIBIT 45. YEAR 5 CLASSROOM MODEL RATINGS BY DISTRICT AND COHORT	103
EXHIBIT 46. CLASSROOM USAGE OF SIM-CERT ROUTINES: YEAR 3	105
EXHIBIT 47. CLASSROOM USAGE OF SIM-CERT ROUTINES: YEAR 4	105
EXHIBIT 48. CLASSROOM USAGE OF SIM-CERT ROUTINES: YEAR 5	106
EXHIBIT 49. FREQUENCY OF CLASSROOM IMPLEMENTATION: UNIT ORGANIZER	107
EXHIBIT 50. TEACHER SATISFACTION LEVELS WITH SIM-CERT TRAINING WORKSHOPS	119
EXHIBIT 51. TEACHER PERCEPTIONS OF SIM-CERT COACH SUPPORTIVENESS	121
EXHIBIT 52. SAMPLE CHARACTERISTICS FOR TREATMENT AND COMPARISON GROUPS BY DISTRICT	124
EXHIBIT 53. MULTILEVEL MODELS ESTIMATING SLOPE OF MCAS ELA SCORES IN PRE-TREATMENT YEARS FOR TREATMENT, COMPARISON, AND COMBINED SCHOOLS	126
EXHIBIT 54. MULTILEVEL MODEL DESCRIBING THE RELATIONSHIP BETWEEN MCAS ELA SCORES AND SIM-CERT, ACROSS FIVE PRE-TREATMENT AND FOUR STUDY YEARS	128
EXHIBIT 55. MODEL PREDICTED MEANS OVER TIME FOR TREATMENT AND COMPARISON SCHOOLS	129
EXHIBIT 56. SAMPLE CHARACTERISTICS FOR TREATMENT GROUP BY DISTRICT AND SCHOOL	134
EXHIBIT 57. MULTILEVEL MODELS DESCRIBING THE RELATIONSHIP <i>BETWEEN</i> PARTICIPATING STRIVING READERS SCHOOLS' MCAS ELA SCORES AND SIM-CERT	137
EXHIBIT 58. MULTILEVEL MODELS DESCRIBING THE RELATIONSHIP WITHIN PARTICIPATING STRIVING READERS SCHOOLS' MCAS ELA SCORES AND SIM-CERT ACROSS STUDY YEARS	139

Executive Summary

This evaluation report presents implementation and impact findings to date regarding the Striving Readers grant as implemented by the Springfield and Chicopee Public School Districts. Any questions regarding this final report should be directed to the Office of Elementary and Secondary Education (OESE) at the U.S. Department of Education.

There were 25,213 students enrolled in Springfield and 7,845 in Chicopee in the 2010–11 school year. The districts differed in terms of student demographics as well as in size. In Springfield, 88% to 92% of the students were designated as minority in the participating schools as compared to 25% to 35% in Chicopee. Over three-quarters of the students in Springfield were also eligible for free or reduced lunch (80% to 84%) as compared to approximately one half in Chicopee (44% to 51%). District accountability data trends demonstrate the need for student literacy support. The Striving Readers grant requires the implementation of both targeted and whole-school literacy interventions. In collaboration with developers, five high schools within Springfield and Chicopee—three in Springfield and two in Chicopee—are implementing two targeted interventions to promote the reading skills of struggling readers as well as a whole-school intervention designed to promote content literacy throughout the student population.

The targeted interventions are: (1) READ 180 Enterprise Edition (Scholastic, Inc.) and (2) Strategic Instruction Model (SIM) Xtreme Reading (University of Kansas, Center for Research on Learning). Both targeted interventions were to be provided as a supplement to the regular English Language Arts curriculum in the participating schools. The whole-school intervention is the Strategic Instruction Model Content Enhancement Routines for Teachers (SIM-CERT), which is a part of the University of Kansas’s Content Literacy Continuum (University of Kansas, Center for Research on Learning).

Implementation

The evaluation of the Springfield-Chicopee’s Striving Readers Program implementation focused on *the extent to which the intensive targeted and school-wide interventions were implemented*

on-model and also sought to describe the general context of implementation for the interpretation of outcomes. For this study, the extent to which an intervention was “on-model” was the extent to which the intervention was implemented according to the developers’ and districts’ specifications and plans. Each intervention encompassed both specifications related to classroom model implementation (e.g., use of instructional practices) and specifications related to the necessary inputs for achieving an appropriate level of classroom implementation (e.g., professional development training for teachers). Implementation levels characterize the complexity of the context in a meaningful and understandable way. In addition, defining levels of implementation provides a way to gauge the magnitude of an identified influence on study outcomes. Implementation of all interventions was evaluated within and across years. The implementation study entailed assigning ratings for adequacy based on the presence of observed and reported model components. Additional data sources (e.g., documents, interviews, surveys) provided a broad picture of the context of study implementation. Additional data sources (e.g., documents, interviews, surveys) provided a broad picture of the context of study implementation.

Targeted Interventions: Inputs, Classroom Model, and Context

In Year 5, a total of 15 teachers implemented the program: five READ 180 teachers, five Xtreme Reading teachers, and five Control classroom teachers. The same numbers of teachers implemented the program in Years 1–4, with the exception of an additional co-teacher in one READ 180 classroom in Year 1. Random assignment was employed to help ensure that teacher quality would be as equally distributed among the conditions as possible. In the final years, the district replaced ninth-grade intervention teachers with those teaching the intervention in the upper grades (non-RCT grades). Across the five years of implementation, a total of 14 teachers have taught READ 180, 11 have taught Xtreme Reading, and 9 have been designated as control classroom teachers. Of the 34 total in the study, 6 taught for all grant years, while 17 taught for only one year of the grant implementation. The majority of the 17 teachers leaving the study after one year did so in the first and second year of grant implementation, 8 and 6, respectively.

Overall, teacher turnover among READ 180 teachers was higher than those for Xtreme Reading (9 and 7 teachers, respectively). Rates of teacher attrition were higher in the three Springfield schools for both interventions. It is important to note that the interventions were not equivalent, and therefore their ratings should not be compared.

READ 180: Implementation Ratings

The summary of input ratings for READ 180 model implementation is presented by teacher, over time, in the Exhibit 1. For the inputs, all READ 180 teachers received aggregate ratings of adequate or high in Year 5, indicating that the professional development, materials, and classroom structure required for implementation had been provided for the majority of teachers.

Exhibit 1. Summary READ 180 input ratings Years 1–5 (n = 14)

Teacher	Year 1	Year 2	Year 3	Year 4	Year 5
1	Adequate	--	--	--	--
2	Moderate	--	--	--	--
3	Moderate	--	--	--	--
4	Adequate	Adequate	--	--	--
5	Adequate	--	--	--	--
6	Adequate	--	Adequate	Adequate	--
7	--	Adequate	Adequate	Adequate	Adequate
8	--	Adequate	--	--	--
9	--	Moderate	--	--	--
10	--	Adequate	--	--	--
11	--	--	Moderate	Adequate	Adequate
12	--	--	Moderate	Moderate	Adequate
13	--	--	Moderate	Adequate	Adequate
14	--	--	--	--	Adequate

Note. Implementation levels were defined as: 1 = No evidence (0–24%); 2 = Low (25–49%); 3 = Moderate (50–74%); and 4 = Adequate or High (75–100%).

Only because the ratings were aggregated for professional development, materials, and classroom structure did one of the teachers, new to READ 180, receive an adequate rating as this teacher had not received all of the professional development. All teachers indicated they had enough teacher materials and were provided with the required 90-minute daily class period. Input scores increased from prior years when fewer teachers received moderate scores.

The summary of classroom ratings for READ 180 model implementation is presented by teacher, over time, in Exhibit 2. For the classroom model, four of the five READ 180 teachers received aggregate ratings of adequate or high in Year 5, indicating fidelity of implementation as defined was achieved. The remaining READ 180 teacher (one of the five) was implementing with a low level of fidelity. Overall, ratings for classroom fidelity increased in Year 5.¹

Exhibit 2. Summary READ 180 classroom model ratings Years 1–5 (n = 14)

Teacher	Year 1	Year 2	Year 3	Year 4	Year 5
1	Adequate	--	--	--	--
2	Adequate	--	--	--	--
3	No evidence	--	--	--	--
4	Adequate	Low	--	--	--
5	No evidence	--	--	--	--
6	Low	--	Low	Adequate	--
7	--	Moderate	Adequate	Adequate	Adequate
8	--	Adequate	--	--	--
9	--	Moderate	--	--	--
10	--	Adequate	--	--	--
11	--	--	Moderate	Moderate	Adequate
12	--	--	Moderate	Moderate	Adequate
13	--	--	Moderate	Moderate	Adequate
14	--	--	--	--	Low

Note. Implementation levels were defined as: 1 = No evidence (0–24%); 2 = Low (25–49%); 3 = Moderate (50–74%); and 4 = Adequate or High (75–100%).

Patterns over time were difficult to discern because, with the exception of one teacher, different teachers implemented in Years 1 and 2 as compared to Years 3 and 4. However, ratings remained consistent over time despite teacher turnover in Years 1 and 2, likely due to the district decision to replace these teachers with those experienced in teaching the intervention in the upper grades when new hires and random assignment were not possible. Teachers who continued teaching READ 180 over time had higher classroom implementation ratings over time. Four of the five READ 180 teachers had implemented the intervention in the prior year; one of the four teachers with the highest ratings had taught READ 180 longest (four years as compared to three years for the remaining three teachers).

¹ Overall, ratings for classroom fidelity remained the same in Year 4 as compared to Year 3 with the exception of one teacher (a rating of low changed to a rating of high). In both Years 3 and 4, teachers received moderate scores rather than adequate because they were observed to be behind schedule as per the pacing calendar and did not devote the full 90 minute class period to READ 180 instruction.

Xtreme Reading: Implementation Ratings

The summary of input ratings for the Xtreme Reading model implementation is presented by teacher, over time, in Exhibit 3. For the inputs, all Xtreme Reading teachers received aggregate ratings of adequate or high in Year 5, with the exception of one teacher.² The teacher with a rating of moderate for implementation was new to teaching Xtreme Reading for this grade level.

Exhibit 3. Summary Xtreme Reading input ratings Years 1–4 (n = 11)

Teacher	Year 1	Year 2	Year 3	Year 4	Year 5
1	Adequate	--	--	--	--
2	Adequate	--	--	Adequate	--
3	Adequate	Moderate	Adequate	Adequate	Adequate
4	Moderate	Moderate	Adequate	Adequate	Adequate
5	Adequate	--	--	--	--
6	--	Low	Adequate	Adequate	Adequate
7	--	Adequate	Adequate	--	--
8	--	Adequate	--	--	--
9	--	--	Adequate	Adequate	--
10	--	--	--	--	Adequate
11	--	--	--	--	Moderate

Note. Implementation levels were defined as: 1 = No evidence (0–24%); 2 = Low (25–49 %); 3 = Moderate (50–74%); and 4 = Adequate or High (75–100%).

No professional development was required because the two replaced teachers in Year 5 had been teaching in the upper grades. The lower rating for one of the teachers was due to lower ratings for materials received. The summary of classroom ratings of Xtreme Reading model implementation is presented by teacher, over time, in Exhibit 4.

² For the inputs, all Xtreme Reading teachers received ratings of adequate or high in Year 4, as in Year 3. Ratings were lower in Year 2 (two teachers with moderate ratings and one teacher with a low rating), primarily due to the teacher-reported lack of receipt of all instructional materials and, for one teacher, insufficient provision of professional development.

Exhibit 4. Summary Xtreme Reading classroom ratings Years 1–4 (n = 11)

Teacher	Year 1	Year 2	Year 3	Year 4	Year 5
1	Adequate	--	--	--	--
2	Adequate	--	--	Adequate	--
3	Adequate	Moderate	Low	No evidence	Moderate
4	Moderate	Moderate	Moderate	Moderate	Moderate
5	No evidence	--	--	--	--
6	--	Low	Adequate	Adequate	Adequate
7	--	Low	Adequate	--	--
8	--	Low	--	--	--
9	--	--	Moderate	Moderate	--
10	--	--	--	--	Adequate
11	--	--	--	--	Moderate

Note. Implementation levels were defined as: 1 = No evidence (0–24%); 2 = Low (25–49%); 3 = Moderate (50–74%); and 4 = Adequate or High (75–100%).

For the classroom model, two of the five Xtreme Reading teachers received aggregate ratings of adequate or high in Year 5 (same as in Year 4) indicating fidelity of implementation as defined was achieved. Three of the five Xtreme Reading teachers were implementing with moderate fidelity.³ Overall, ratings for classroom fidelity increased in Year 5.

With the exception of one of the four returning Year 4 teachers, all had the same ratings for Years 3 and 4. Implementation results over time are difficult to interpret due to teacher turnover in Years 1, 2, and 5. Only three of the five teachers from Year 4 returned (the two teachers replacing Year 4 teachers in Year 5 had taught in the upper grades). The moderate or adequate aggregate ratings across time were largely due to the districts’ decision to replace teachers who were leaving with teachers who had previously taught the intervention in the upper grades. However, one of the two longer-term teachers had only ratings of moderate and the ratings were generally inconsistent for this teacher over time.

³ The moderate ratings for the two teachers in Year 4 were the result of these teachers being behind schedule as per the pacing calendar and not implementing core instructional strategies as defined. The teacher rated as having no evidence in Year 4 was not observed to be implementing Xtreme Reading content or instructional strategies.

Whole-School Intervention: Inputs, Classroom Model, and Context

SIM-CERT

The districts' training goals were set at 125 per year (25 teachers per school) for the SIM-CERT whole-school intervention. According to district records of professional development attendance, across the five grant years a total of 623 teachers were selected for inclusion in SIM-CERT cohorts and received some portion of SIM-CERT training.

Inputs and context. According to district records across Years 1 through 5 of SIM-CERT implementation, the majority of Chicopee teachers (70%) received the four required days of training during the first year of implementation compared to very few of Springfield teachers (4%).⁴ District variation was also observed for training rates of teachers in their second year of implementing SIM-CERT, with 78% of Chicopee teachers receiving the recommended two days of training compared with 46% of Springfield teachers (a slight decrease and increase overall, respectively). The timing and structure of the professional development schedule in Springfield accounts for the low percentage of *adequate* ratings for implementation of the professional development model. Refer to Exhibit 5 below.

Exhibit 5. Professional development days required: Percent of teachers receiving adequate ratings by district and cohort

District/ Cohort	Training for first year of implementation	Training for second year of implementation
	Four Days Required	Two Days Recommended
All SPS	4% (n = 14/352)	46% (n = 130/282)
All CPS	70% (n = 124/178)	78% (n = 112/144)
Total	26% (n = 138/530)	57% (n = 242/426)

When adequacy of professional development was assessed by the numbers of teachers *receiving the content* of training rather than by *the number of days* of training, professional

⁴ Springfield was operating on a different professional development training calendar and would only catch up to the original rates in the summer following each grant year.

development scores were high in both districts. This additional rating has been included since Year 3 when districts and developers provided information regarding required content and indicated teachers might have received training in all required topics, regardless of how many days it took to cover the material. Across districts, the majority of teachers (78%) received the training in required content, as illustrated by the following exhibit.

Exhibit 6. Percentage of teachers who received adequate levels of training in the required routines for the first year of implementation

	Receipt of all four core required routines (Unit Organizer, Framing, LINCing, Concept Mastery)
All SPS (n = 272)	74% (n = 202)
All CPS (n = 132)	86% (n = 113)
Total (n = 404)	78% (n = 315)

Over time, the minimum required number of training days set by developers decreased in Springfield. Originally, training would present one SIM-CERT routine and give teachers time to apply that routine to their course content in collaboration with colleagues from their departments. In the later years of the grant in Springfield, this collaborative work time was minimized. In Chicopee, the professional development plan, including the number of days, the content taught, and content delivery, remained consistent from Years 1–5.

In Years 2 and 3, the consensus among teachers and administrators was that the support provided by the literacy coaches had been instrumental in the classroom-level implementation of SIM-CERT. In Years 4 and 5, levels of teacher satisfaction with the training offered and received decreased from Years 2 and 3, and reports of satisfaction with coaching support were more mixed. District variation in teacher response was evident. The overall reduction in reported teacher satisfaction with professional development, in terms of the general amount and quality as well as coaching support, appears the result of several interrelated factors: consolidation of trainings; transfer of responsibility for trainings from developer to school staff; communication and lack of clarity about training requirements; and elimination of after-school training workshops in Springfield. Reported rates of teacher satisfaction for coaching in particular varied

within Springfield across schools, with lower levels of agreement for one school in particular as compared to the others.

Classroom model and context. Overall, approximately three-fourths teachers reported meeting minimum classroom model expectations, consisting of the use of the Unit Organizer and one other SIM-CERT routine during the course of the academic year (as indicated initially by developers). Across districts, approximately three-fourths of the group of teachers who received adequate scores for classroom model fidelity exceeded minimum requirements. These teachers implemented the minimum in addition to another routine of their choice during the school year. There was a minimal but steady decline over time in the percentage of teachers who reportedly met and/or exceeded classroom model requirements. Refer to the following exhibit.

Exhibit 7. Percentage of teachers who met and exceeded minimum requirements for classroom model implementation

Cohort	District	Met Minimum Usage Requirements	Exceeded Minimum Usage Requirements
		Unit Organizer + 1 additional routine	Unit Organizer + 2 or more additional routines
All	CPS (n = 124)	91 (73%)	58 (64%)
All	SPS (n = 172)	92 (52%)	71 (77%)
Total	(n = 296)	183 (62%)	129 (71%)

Across all years and cohorts, evidence of district variation was observed. A greater percentage of Chicopee teachers met and exceeded classroom model specifications than Springfield teachers in Years 2 through 5, but not in Year 5.

Impact

The evaluation of the Springfield-Chicopee’s Striving Readers Program had the primary goal of rigorously assessing the effectiveness of the interventions as implemented on reading achievement. The most rigorous design, a randomized controlled trial (RCT), was implemented for the targeted interventions to address the counterfactual (i.e., what would happen in the absence of treatment). Because such a design was not feasible to assess the

impact of the whole-school intervention, an interrupted time series (ITS) analysis of secondary data was proposed. In addition, comparison schools were included in the ITS analysis to more fully address the counterfactual. The primary outcome for the analysis of targeted student impacts is the Stanford Diagnostic Reading Test, version 4 (SDRT-4), and the Massachusetts Comprehensive Assessment System (MCAS) English Language Arts was used for assessing whole-school impact.

Targeted Interventions Impacts

Eligible, incoming, ninth-grade students were randomly assigned to one of three conditions: Control, READ 180, or Xtreme Reading. Each of the treatment group impact estimates—for READ 180 and Xtreme Reading—was assessed in comparison to the control group. Because students were randomly assigned to intervention groups, students were the primary unit of analysis. To answer the primary research question regarding the effectiveness of the interventions and to provide estimates of their “true” effects on reading achievement, average reading achievement scores of students in each of the two interventions were compared to the scores of students in control group classrooms, pooled across sites and study years.

Using criteria outlined by What Works Clearinghouse (WWC) for assessing the rigor of designs and analysis, baseline or pretest scores were assessed to identify pre-treatment differences among the groups. No significant baseline or pretest differences were observed. In addition, the numbers of “actual” exclusions were examined to identify differential attrition between groups (i.e., these exclusions would have been noted at the time of screening and assignment review but were not available to evaluators until late fall). No differences in attrition estimates among treatment groups were greater than 20%.

Patterns generally remained the same with the addition of Cohort 5 (Year 5) as in the past for baseline and outcome scores. No significant effects were observed for Xtreme Reading as compared to the control group. Significant effects were observed for READ180 as compared to the control group. READ 180 students scored significantly higher as compared to control students (1.5 points on average unadjusted NCE and 2.39 adjusted NCE), representing an effect size of .11. Although the unadjusted mean represents the true difference between

groups in this random assignment study, adjusted means were calculated in the event random assignment did not yield equivalent groups due to the smaller sample sizes. The mean scores at post-test, though higher than at pretest, represent less than grade level performance (approximately between a fifth and sixth grade reading level).

Targeted Interventions Impact and Classroom Implementation

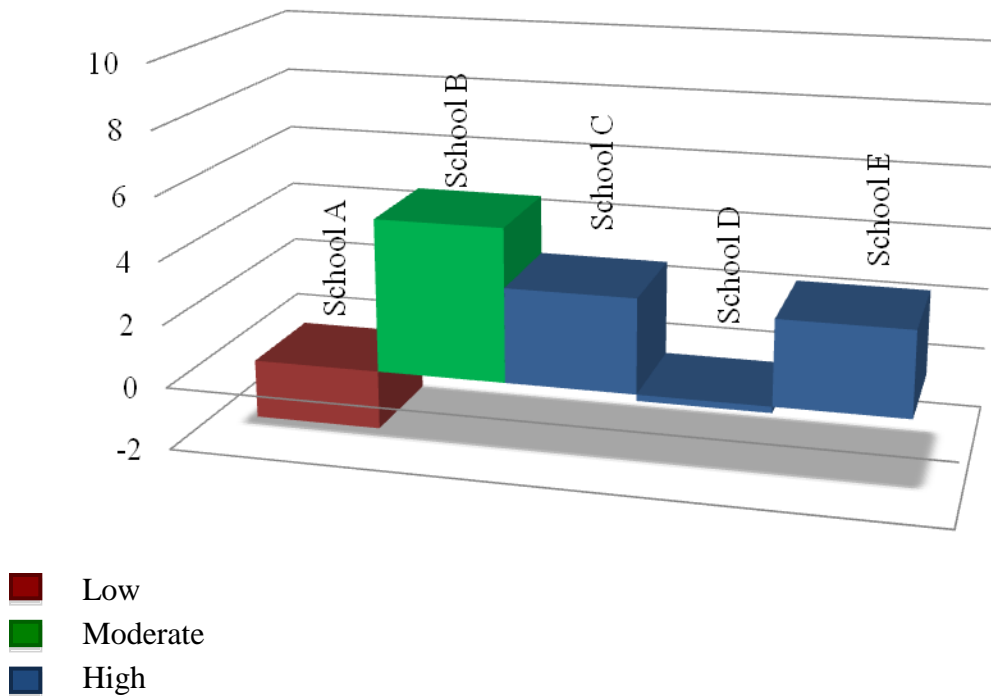
The goal of the targeted implementation study was to inform the interpretation of impact findings by describing the context in which the interventions were implemented. More specifically, implementation levels were established to characterize the context and its complexity and, as a result, to provide a gauge by which to judge any observed effects relative to the context. Therefore, the following analysis describing the relationship between classroom level implementation and impact scores was purely exploratory and not intended to predict the impact of the interventions.⁵

READ 180 Classroom Implementation and Impact

The comparison of classroom implementation and impact results for READ 180 is included in Exhibit 8 below. This exhibit illustrates that in schools where classroom implementation levels were observed to be moderate and high (as coded by color) the average reading scores of READ 180 students were higher relative to students in the control group (the difference represented on the Y axis in reading achievement scores or SDRT-4 NCEs).

⁵ The hypothesis that higher levels of implementation would be related to higher levels of observed impact was not empirically tested; analyses were purely illustrative. As described in the Enhanced Reading Opportunities Study, such analyses: "...are not able to establish causal links between these aspects of implementation and variation in program impacts across sites, because school characteristics and other implementation factors may confound the association between...impacts and the implementation factors included in the exploratory analysis" (Corrin, et al., 2008).

Exhibit 8. Impact of READ 180 by level of classroom implementation (Years 1–5)



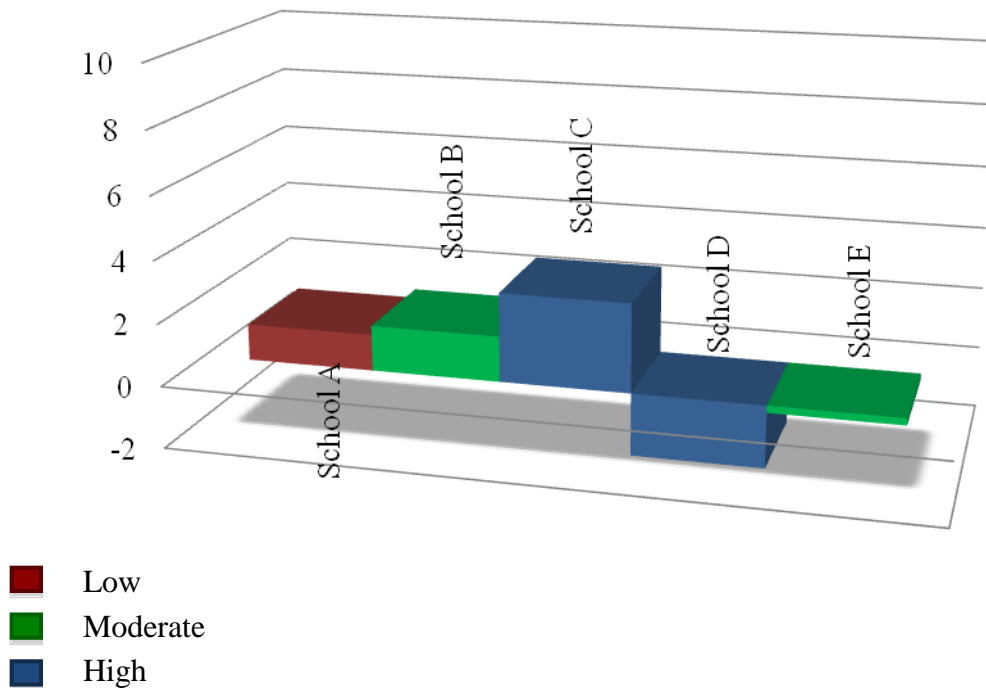
Note. Averages were calculated weighted by the total number of items across years. Implementation levels: No evidence (0–24%), Low (25–49%), Moderate (50–74%), and Adequate or High (75–100%).

READ 180 implementation levels were assessed in relationship to outcome scores for READ 180 students, and this relationship visually represented in the exhibit was significant. That is, higher levels of READ 180 implementation were associated with higher reading scores. Four of the five teachers with the highest classroom ratings had taught this intervention the longest, three for three years and one for four years. Results were more consistent over time for the majority of teachers especially those implementing at high levels over the entire study period. On average, READ 180 student scores were higher at post-test, controlling for pre-test scores and other student characteristics than control group student scores, and this difference was statistically significant.

Xtreme Reading Classroom Implementation and Impact

The comparison of classroom implementation and impact results for the Xtreme Reading intervention is included in the exhibit below.

Exhibit 9. Impact of Xtreme Reading by level of classroom implementation (Years 1–5)



Note. Averages were calculated weighted by the total number of items across years. Implementation levels: No evidence (0–24%), Low (25–49%), Moderate (50–74%), and Adequate or High (75–100%).

This exhibit illustrates that in schools where classroom implementation levels were observed to be moderate and high (as coded by color) the average reading scores of Xtreme Reading students were higher relative to students in the control group in only two of four schools (the difference represented on the Y axis in reading scores or SDRT-4 NCEs).

The pattern of prior teaching was not as easy to discern for Xtreme Reading; as noted in the prior scoring section, one of the two teachers with the lowest overall ratings had been implementing since the initial grant year.

Xtreme Reading implementation levels were assessed in relationship to outcome scores for Xtreme Reading students, and this relationship visually represented in the exhibit was not significant. That is, higher levels of Xtreme Reading implementation were not associated with higher reading achievement scores. On average, the Xtreme Reading student scores were approximately the same at post-test, controlling for pre-test scores and other student

characteristics than control group student scores; there was not a statistically significant difference observed between the two groups.

Implementation Patterns as Predictor

Despite the many complications related to implementation, particularly in Year 1 of the study, a pattern of medium (i.e., moderate) and high (i.e., adequate) targeted implementation levels and higher overall student reading scores was observed. This pattern was more pronounced for READ 180 and was significant when assessed in relationship to reading scores.

Over time, the targeted teachers had more experience, and the control classroom teachers had higher levels of education. As a result of teacher turnover, the backgrounds as compared to control classroom teachers changed. Background and experience, in addition to overall teaching quality (not directly measured), among other unmeasured factors could have influenced and moderated any observed results.

Although impact estimates were established across years, implementation levels and impact results varied by year, which itself has implications and at a minimum requires caution when interpreting any of these findings. It is important to note that these cautions should be exercised for both interventions, as there were differences in implementation between years for both Xtreme Reading and READ 180, including teacher turnover in earlier years.

Whole-School Intervention Impact

The impact of the whole-school intervention (SIM-CERT) on student achievement, specifically achievement in English language arts (ELA) inclusive of reading, was estimated over time.⁶ A quasi-experimental rigorous assessment of the impact utilized a short, interrupted, time-series analysis (SITS) inclusive of a comparison group.⁷ Student achievement trends at the Striving Readers high schools were compared to trends at other high schools in Massachusetts serving

⁶ Outcomes for teachers were not proposed as there were no secondary data available to assess teacher-level outcomes.

⁷ Refer to Bloom (2001). Source: <http://www.mdrc.org/>

similar student populations (see Exhibit 1). Aggregate student achievement scores as measured by the state ELA assessment (MCAS ELA, inclusive of reading) were obtained from both treatment and comparison schools. Aggregate scores were included for each cohort of 10th-grade students from each of the five years pre-treatment (2001–02 through 2005–06) and from each of the first four years during the treatment period (2006–07 through 2009–10).

In summary, the results from the pre-treatment years indicate the treatment and comparison schools were well-matched. On average, students' ELA achievement scores have increased by approximately 1 point per grant year, lower than the 2.3 point increase observed prior for three years of implementation. However, results from the current SITS analysis indicated the five Striving Readers schools were performing similarly to comparable schools in the state—in districts not participating in the Striving Readers grant—on the ELA portion of the MCAS. In conclusion, although the five Striving Readers schools implementing SIM-CERT increased their ELA achievement scores over time, there was no evidence that the increases were due to SIM-CERT as similar increases were observed for the comparison schools.

Any number of similar initiatives may have been implemented in the comparison group schools, which could explain a lack of observed impact results (i.e., no significant differences between the Striving Readers and non-Striving Readers schools on overall aggregate ELA achievement scores).⁸ Comparison schools may have been implementing an intervention or curricular changes with equal intensity to affect outcomes. In addition, a lack of observed impact results may be a function of a less than ideal sample size combined with less than ideal fidelity of implementation across treatment schools (refer to CERT implementation). That is, even if implementation was perfectly executed in one or two of the schools, the overall effect may not have been strong enough to illustrate differences in comparison to the other schools with a small sample size.

⁸ Especially in the context of schools in need of improvement and restructuring, this is likely to be the case. However, data were not readily available to assess this assumption.

Whole-School Impact and Implementation

A non-experimental assessment of the relationships between SIM-CERT training and implementation and school-level achievement scores over time were explored. Student achievement scores, as measured by the MCAS ELA, from each cohort of grade 10 students assessed in participating high schools were analyzed for the first four years of the treatment period (2006–07 through 2009–10).

Although the previously presented analysis of the impact of the whole-school intervention was conducted to assess a causal relationship, if one was present, the following analyses do not attempt the same.⁹ The previous analysis included a well-matched comparison group to address the counterfactual (i.e., what would happen in absence of treatment); the analyses presented here do not include a comparison group.

In summary, the results of this descriptive analysis (not implying causation) indicated that two of the four measures of SIM-CERT training and implementation levels were predictive of ELA achievement *between* schools. Three of the four SIM-CERT implementation variables were not measured in every program year, and therefore a potential association with the outcome may be underestimated. However, the results do not imply that higher implementation levels caused higher ELA achievement scores. Additional explanations for observed results include the possibility that higher performing schools, in terms of ELA achievement scores, may be more likely to implement SIM-CERT at higher levels. That is, schools performing at higher levels could be doing so as a result of factors unrelated to SIM-CERT such as less staff and administrative turnover, potentially resulting in more clearly defined leadership and stability as a result.

Additional results indicated implementation was not a significant predictor of the growth in ELA achievement scores in the treatment years, *within* schools. There was no evidence that when an individual school varied in implementation levels over time, ELA achievement

⁹ It is important to note the limitations of the prior analyses, as already described in the SIM-CERT impact section. However, even with a well-matched comparison group included, an assessment of aggregate school-level impacts like those reported here would not currently be considered for review by the What Works Clearinghouse (WWC).

scores were better in the years when implementation occurred at higher levels. However, three of the five schools *never* met adequate levels of professional development at any point over time. Delivering the complete training in the summer following the implementation school year meant that these schools always attempted to “catch up,” and this could explain a lack of observed results. Finally, there were a number of other interventions implemented school-wide in the treatment schools in Springfield over the course of the Striving Readers, making disentangling SIM-CERT results of further difficulty. Although attempts to assess the impact of the onset of these interventions versus SIM-CERT did not yield clear results, such an outcome could have been the result of an inability to define the onset more clearly rather than the mark of no influence at all.

Overall Summary

The evaluation of the Springfield-Chicopee’s Striving Readers Program had the primary goal of rigorously assessing the effectiveness of the interventions as implemented on reading achievement. In addition, implementation studies were included to present a broad picture of the overall level of implementation in context and a sense of the variability that may have occurred. Differing institutional contexts or constraints influenced the ways in which intervention components were implemented. Districts and schools possessed their own unique complexities, which may have supported or hindered implementation and, in turn, affected outcomes. Finally, implementation analysis indicated barriers faced and addressed throughout the grant period.

Final results from the implementation of Striving Readers interventions to date in Springfield and Chicopee school districts indicated a positive and significant impact on student reading achievement of one of the two targeted interventions. The impact of the whole-school intervention was not established. Implementation studies also indicated alignment of contextual results with outcomes observed.

The Springfield and Chicopee school districts have overcome many obstacles in the development, planning, and implementation of their Striving Readers grant. In particular, two dissimilar districts have implemented two targeted interventions (all other SR grantees

implemented only one) as well as one whole-school intervention. Implementation studies reported barriers in the implementation of the grant in Year 1 resulting from both contextual and contractual factors, which did not necessarily emerge from the intervention models but may have resulted from attempts to fit the models as required into this context. Some of the contextual factors included: the urban setting, population, and student needs; the various policies of the schools and districts addressing scheduling, and administrative issues; and general staffing and personnel matters. Contractual complexities specifically referred to the requirements for the grant implementation; the monitoring and oversight of the fidelity of implementation; and the observance of the rigorous research specifications.

Given the challenges inherent in creating a successful collaboration between two districts and implementing two interventions, it is not surprising that complexities arose that would not normally be encountered in a standard literacy program implementation.

An initial barrier related to the rigorous research requirements, for example, involved the cooperation, ability, and willingness of both districts to incorporate a “true” control group to address the counterfactual (i.e., what would happen in the absence of treatment). Additional challenges involved the need to standardize implementation across two very different district and school systems. Intervention plans necessitated consistent tailoring to accommodate rigorous research study requirements, and district staff and evaluators spent unanticipated time to ensure successful implementation. At the same time, districts faced turnover in lead program staff and administrators, challenges related to communication with stakeholders and participants, and complications in screening and placing the population of students who were randomly assigned to participate in the targeted interventions as well as the tracking of these students over time.

These difficulties have had some lasting influence but over time the districts have sought to address each one as presented in the evaluation reports. Progress was made in overcoming these barriers, particularly in Year 2, but also throughout Year 3. Districts implemented each of the targeted interventions while maintaining the integrity of the randomized controlled trial design and assignment to the best of their ability and repeatedly demonstrated their commitment to ensuring the success of the grant. District staff collaborated fully with evaluators in all phases

of the evaluation. Their serious consideration of any potential positive or negative influences on study outcomes as well as “full disclosure” has been commendable. Such diligence ensures that these final study results has produced information that can be used by policymakers, district administrators, and school staff to make confident choices regarding effective literacy interventions for their students.

I. Introduction and Study Background

This report presents implementation and impact findings to date based on district documentation and data gathered by The Education Alliance regarding the Striving Readers grant as implemented by the Springfield and Chicopee Public School Districts. Any questions regarding this final report should be directed to the Office of Elementary and Secondary Education (OESE) at the U.S. Department of Education.

The Striving Readers grant requires the implementation of both targeted and whole-school literacy interventions. In the Springfield and Chicopee Public School Districts, five high schools (three in Springfield and two in Chicopee) in collaboration with developers are implementing two targeted interventions—both developed using scientifically based research to promote the reading skills of struggling readers—as well as a whole-school intervention developed to promote reading skills throughout the student population.

The targeted interventions are: (1) READ 180 Enterprise Edition (Scholastic, Inc.) and (2) Strategic Instruction Model (SIM) Xtreme Reading (University of Kansas, Center for Research on Learning). Both targeted interventions have been provided as a supplement to the regular English Language Arts curriculum in the participating schools. The whole-school intervention is the Strategic Instruction Model Content Enhancement Routines for Teachers (SIM-CERT), which along with Xtreme Reading is a part of the University of Kansas’s Content Literacy Continuum (University of Kansas, Center for Research on Learning).

The U.S. Department of Education (ED) and its contracted Striving Readers technical assistance provider, Abt Associates, have made significant contributions to this report as has the Striving Readers district implementation team (SR district team) in its dedication to providing accurate information and documentation about implementation.

II. District Context

Located in western Massachusetts, the mid-sized city of Springfield was a community of 152,082 people at the onset of this grant (U.S. Census, 2006). Twenty-nine percent of Springfield’s population comprised children under the age of 18. Approximately 23% of the overall population and more than 75% of all public school students in Springfield lived in households at or below the poverty line.¹⁰ Chicopee is a neighboring community of Springfield. At the onset of the grant, Chicopee had 23,117 households, and 23% percent of the population comprised children under the age of 18. The median household income was \$35,672, and approximately 12% of the overall population lived below the poverty line (U.S. Census, 2006).

Characteristics of Districts and Student Population

Springfield Public Schools enrolled approximately 25,213 students in the 2010–11 school year (MADOE, 2011).¹¹ Springfield is the second largest school system and one of the lowest performing school districts in the state. A Title I District, Springfield has four high schools, three of which are participating in the Striving Readers Program.¹² Although the three high schools—High School of Commerce, Putnam Vocational-Technical High School, and the Springfield High School of Science and Technology (SciTech)—are non-Title I schools by designation, they qualify as eligible to receive Title 1 funds (MADOE, 2010).¹³ Additionally, all three high schools participate in the Metropolitan Council for Educational Opportunity (METCO), a state-funded program designed to address racial imbalances by busing children from urban to suburban areas (METCO, n.d.). A state-appointed financial control board has governed Springfield’s public schools as well as the City of Springfield. The financial

¹⁰ Local poverty statistics obtained from a district document downloaded from www.sps.springfield.ma.us, November 7, 2007 to reflect status prior to grant implementation.

¹¹ Data were obtained from the Massachusetts Department of Education’s District Profiles database, <http://profiles.doe.mass.edu/>, March 2011.

¹² This does not include the numerous alternative secondary schools and private secondary schools located in Springfield.

¹³ This is true of Chicopee High Schools as well. Eligibility relies upon what one Striving Readers program manager referred to as a “calculation of preponderance”; although the number of students registered for free/reduced lunch does not necessarily reflect a percentage that warrants Title I status, the preponderance of other factors (most notably, the Title I status of all middle schools) indicates that the number of known free/reduced lunches is lower than the number of students qualifying.

difficulties the city and district have faced, in addition to past teacher contract difficulties, have contributed to significant losses of teachers, other personnel, and services to the public schools.

Chicopee has two high schools, both of which are participating in the Striving Readers Program. Like Springfield, Chicopee is a Title I District with its two high schools eligible to receive Title I funds. Chicopee also participates in the METCO program. Chicopee Public Schools enrolled 7,875 students in the 2010–11 school year (MADDOE, 2011).

Descriptive information for every high school participating in the Striving Readers Program for Year 5 is presented in Exhibit 10.

Exhibit 10. Characteristics of participating schools (2010–11) ¹⁴

Characteristics	Chicopee Schools		Springfield Schools			State
	CHS	CCHS	Putnam	SciTech	Commerce	
	%	%	%	%	%	%
Non-White	35	25	88	89	92	32
First Language Not English	15	11	28	31	33	16
Limited English Proficient (LEP)	3	3	10	14	21	7
Low Income	51	44	80	84	81	34
Special Education	14	16	24	23	29	17
Total Number of Students	1209	1454	1545	1267	1286	–

Source: Massachusetts Department of Education. *School/District Profiles*. Retrieved March 2011 from <http://profiles.doe.mass.edu/>

Adequate Yearly Progress (AYP) Status

The five Springfield and Chicopee high schools operate in a high-stakes climate with strict, state-mandated graduation requirements. As required by the federal No Child Left Behind Act (NCLB), all schools and districts are expected to meet or exceed specific student performance standards in English Language Arts/Reading (ELA) by the year 2014. In order to monitor progress toward set performance goals, state departments of education issue adequate yearly

¹⁴ The characteristics of the participating schools were similar to those reported for the prior implementation years (2006-07, 2007-08, and 2008-09). Refer to prior reports posted on ed.gov.

progress (AYP) determinations. District accountability data trends demonstrate the need for literacy support for both middle school and high school students. Exhibit 11 depicts the performance history of the Springfield and Chicopee districts in ELA by providing a snapshot of AYP status for the year of the grant application and for the subsequent years of implementation of the Striving Readers Program to date (2006–10).

Exhibit 11. AYP determination for ELA by district (2006–10)

	Chicopee					Springfield				
	2006	2007	2008	2009	2010	2006	2007	2008	2009	2010
Grades 6–8										
Aggregate	Not met	<i>Met AYP</i>	Not met	<i>Met AYP</i>	Not met	Not met	Not met	Not met	Not met	Not met
Subgroup	Not met	Not met	Not met	Not met	Not met	Not met	Not met	Not met	Not met	Not met
Grades 9–12										
Aggregate	Not met	<i>Met AYP</i>	<i>Met AYP</i>	<i>Met AYP</i>	Not met	Not met	Not met	Not met	Not met	Not met
Subgroup	Not met	Not met	Not met	Not met	Not met	Not met	Not met	Not met	Not met	Not met

Source: Massachusetts Department of Education. *School/District Profiles*. Retrieved March 2011 from <http://profiles.doe.mass.edu/>

In Chicopee, at the high school level, aggregate scores met AYP criteria for three years, but subgroups continued to lag behind. Chicopee schools were designated as “Improvement Year 2” after three consecutive years of not making AYP requirements for subgroups. In such cases, the Massachusetts accountability system requires that the schools offer parents the option of sending their child to another school within the district that has made AYP, if space is available.

In Springfield, AYP benchmarks have not been met at the aggregate or subgroup level. As stated in the Year 2 report, the fact that these subgroups were not making AYP is particularly relevant given that a majority of district students were African American or living in poverty. In 2008, the Springfield schools were designated as “Restructuring.” The district’s only Chapter 74 approved, vocational-technical school was also designated by the state as “chronically under-performing” and was subsequently converted to a Commonwealth Pilot School in Year 2 of the Striving Readers grant.

III. Theoretical Rationale and Description of Interventions

Two targeted interventions, READ 180 and Xtreme Reading, were selected by the Springfield-Chicopee¹⁵ school districts to improve the reading skills of struggling readers. Both READ 180 and Xtreme Reading were implemented as “add-on” or supplemental interventions. That is, the interventions were conducted in addition to the regular ELA class required in the participating schools.¹⁶ The whole-school intervention model, SIM-CERT, was selected to improve literacy across content areas, and its implementation was phased in over the grant period. The following descriptions summarize key elements of the interventions, as planned and implemented, with any changes occurring in each year and over time.

READ 180 Targeted Intervention

The READ 180 program is an intensive literacy curriculum developed for struggling readers in grades 4 through 12 to bring their reading skills to grade-level standards and to promote reading comprehension. Initially developed in 1985 by Ted Hasselbring at Vanderbilt University, the program, then named the Peabody Literacy Lab, uses anchored instruction (Hasselbring & Goin, 2004). Anchored instruction is based on a philosophy of using authentic situations as anchors to “enable students to practice noticing and resolving problem situations” (p. 138). The READ 180 program also uses computer-assisted instructional (CAI) software to track individual student progress and to adjust reading instruction accordingly. Using the concept of anchored instruction, the CAI software has “an animated tutor who guides the student and provides feedback via a digitized human voice” (p. 133).¹⁷

The goal of READ 180 is to help struggling readers achieve proficiency in reading at grade level. Objectives of the program include targeting specific elements of phonics, fluency, vocabulary, comprehension, spelling, writing, and grammar, as well as promoting self-directed learning

¹⁵ Springfield-Chicopee is used as an abbreviation for the Springfield Public Schools and Chicopee Public Schools implementing their jointly proposed Striving Readers program.

¹⁶ As a result, students had to wait to take an elective, such as art, until the upper grades. Physical education, which is not an elective but required for one semester per year, was doubled-up in upper grades to fulfill this requirement.

¹⁷ After purchasing the rights to the Peabody Literacy Lab Program and changing its name to READ 180, Scholastic contributed significantly to the program’s further development (Scholastic, Inc., 2005a).

(Scholastic, Inc., 2005a). The reading materials contain content that is of interest to this particular age group and is connected to students' everyday experiences.

READ 180: Instructional Approach and Curriculum

The READ 180 instructional model provides structure to classroom activity. The model is based on a 90-minute block that blends whole-class instruction and small-group student work. The teacher begins with 20 minutes of whole-class instruction in which skills are explicitly taught in the areas of word analysis, vocabulary, and reading comprehension, and concludes with a 10-minute, whole-class wrap-up (Scholastic, Inc., 2005a). For the intervening 60 minutes, students break out into smaller groups and rotate among the following three stations:

1. Small-group direct instruction through which the teacher focuses on needs specific to the selected group of students;
2. Independent student work using READ 180's CAI software; and
3. Modeled or independent reading from paperbacks and/or audio books.

READ 180 provides content through specific teacher resources (e.g., rBook Teacher's Edition, Anchor videos) and student materials for the whole-class and small-group sessions. The rBook Teacher's Edition contains content and instructional routines to encourage active participation and further develop students' reading comprehension, vocabulary, writing, and grammar skills.¹⁸ Anchor videos jump-start the activity during the whole-class direct instruction segment of the class, provide background information, and are designed to capture student interest by raising provocative questions. The rBook's nine workshops are estimated to require one school-year of instruction (approximately eight months or between 125 and 145 days in addition to the two weeks at the beginning of the school year for start-up). In addition, students are provided with their own rBooks, which are interactive work texts.

Teachers use specific READ 180 instructional strategies during READ 180 teacher-directed activities in whole and small groups. In small-group segments, teachers can use many of the

¹⁸ Instructional routines covered include: teaching vocabulary, oral cloze, think (write)-pair-share, idea wave, numbered heads, the writing process, and peer feedback.

whole-class strategies and also offer differentiated instruction in phonics, fluency, vocabulary, word study, spelling, and comprehension. They can provide fluency assessment and practice or conduct teacher conferences to set goals, check reports, reflect on books, and review rBooks (Scholastic, 2005e).

READ 180's professional development is designed "to help teachers be successful and to foster and sustain best teaching practices in the classroom" (Scholastic communication, 2007). Accordingly, READ 180 offers a variety of professional development opportunities and support, ranging from trainings, seminars, in-classroom support, web-based instructional support, and online courses (referred to by Scholastic as RED courses) focused on aspects of reading instruction. A logic model depicting the key components of the READ 180 intervention (as planned and expected outcomes) is depicted in Exhibit 12.

READ 180: Over Time

Scholastic provided updated documentation in Year 3 specifying the number of required in-classroom coaching visits, seminars, and RED online courses based on teachers' years of experience in the READ 180 program.¹⁹ Teachers with either a year or two of experience teaching READ 180 were required to complete an additional Scholastic online course (6 hours total), equal to the hours required of a teacher with no prior experience. Teachers with a year of READ 180 teaching experience were not required to attend seminars and those with two years of experience were required to attend two seminars as compared to the six required for teachers in their first year of teaching READ 180. Finally, teachers with a year of prior READ 180 teaching experience were to be provided with the eight monthly coaching sessions over the school year as was true for teachers with no prior READ 180 teaching experience (approximately 2 hours each). In comparison, teachers with two to three years of prior READ 180 teaching experience were

¹⁹ This document was dated April 6, 2009 and provided to evaluators following the developer interview. The Scholastic online RED 180 course differed based on the number of years a teacher had participated: teachers new to READ 180 received "Read 180: Best Practices in Reading Intervention"; teachers in their second year of teaching READ 180 received Teaching Striving Readers"; and teachers in their third year of teaching READ 180 received "High School Literacy Comprehension Through Active Strategic Reading."

provided with half the monthly coaching sessions, or four total over the school year. Finally, teachers with four years of prior experience were not required to participate in professional development.

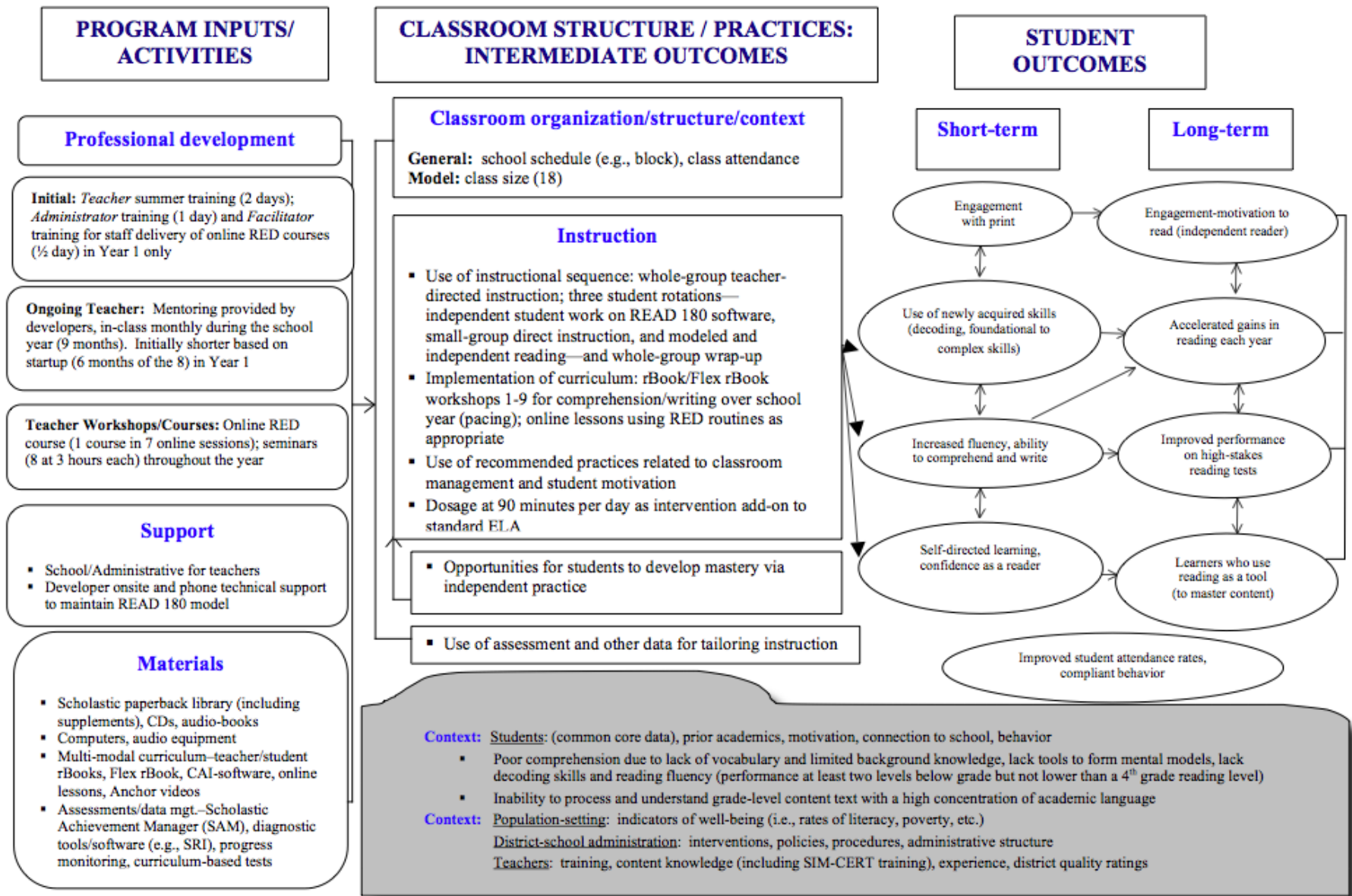
First-year teachers were required to complete a total of 36 hours of group training, which included two initial training sessions (6 hours per session), six follow-up seminars (3 hours per seminar), and Scholastic online training (6 hours in a seven-session course). In addition, first-year teachers were to receive a total of 16 hours of ongoing and individual training and support provided by developers, consisting of eight monthly mentoring sessions over the school-year (approximately 2 hours per session) for a total of 52 hours of professional development training.

As planned and as occurred for prior cohorts, students who received one year of READ 180 in 2008–09 but did not have outcome test scores (SDRT-4) that met grade-level expectations were to be provided with a second year of READ 180.²⁰ These students worked from the already developed Flex rBook that parallels the content of the rBook (the student resource for whole-class and small-group instruction) without duplicating the same texts.²¹ Additionally, per the SR district team, more complex texts were introduced to the students receiving a second year of READ 180 in Years 2 and 3. Developers provided books with more challenging reading for those at higher levels as well as additional titles at the lower Lexile levels for greater variety.

²⁰ Although there was a review of the same skills in the second year of READ 180 participation, including summarizing for comprehension, teachers were to use differentiation to address student needs and to increase the level of sophistication of the skills learned so that these literacy skills could be applied to different content areas/subjects. Information provided by Karen Burke, Scholastic, November 2008.

²¹ These texts are not sequential, so a whole class may start in either the rBook or the Flex rBook and then alternate to the other text the following year, when needed.

Exhibit 12. READ 180 logic model



Xtreme Reading Targeted Intervention

The Xtreme Reading Program of the Strategic Instruction Model (SIM) was developed for adolescents who struggle with reading and writing by the University of Kansas Center for Research on Learning (KU-CRL). Whereas READ 180 focuses on the fundamentals of reading, Xtreme Reading has a meta-cognitive approach focusing heavily on explicit strategy instruction. The Strategic Instruction Model is based on research indicating that content literacy occurs not only when students have mastered the critical content as determined by teachers, but also when students can manipulate and generalize this content to other learning situations (Content Learning Center, 2007).

The SIM Content Literacy Continuum comprises three levels: the SIM-CERT or Content Enhancement Routines for Teachers (Levels 1 and 2) and Xtreme Reading (Level 3) (refer to Exhibit 13).

Exhibit 13. SIM Content Literacy Continuum (CLC)

Level	Purpose	Instruction
1	Master critical content	Enhanced content instruction (strategic teaching to ensure mastery of critical content for all students)
2	Use learning strategies across classes	Embedded strategy instruction (teachers embed selected learning strategies in core curriculum courses)
3	Master specific reading strategies (e.g., self-questioning, visual imagery, paraphrasing)	Explicit strategy instruction (Xtreme Reading)

Source: Dr. Faddis (personal communication, November 2007), RMC Research Corporation, Portland, Oregon, based on information provided by Susan Robinson, University of Kansas, Center for Research on Learning.

More specifically, Xtreme Reading targets students reading at least two years below grade level but who read at or above the fourth grade level. Intensive strategy instruction addresses the skills needed to bring meaning to reading, particularly reading instruction that helps students to

develop accurate word recognition and increased fluency and comprehension. The approach to instruction involves intensive, carefully tailored lessons in which students have numerous opportunities to practice targeted learning strategies that will help them succeed in their classes.

Developers train teachers in all aspects of what are called “Learning Strategies” for students. The professional development model includes initial training, ongoing in-class mentoring by providers, and workshops on specific routines. These strategies prompt teachers to organize, clarify, and standardize student approaches to engaging with and mastering content.

Xtreme Reading: Instructional Approach and Curriculum

The year begins with units addressing behavior (ACHIEVE, Talking Together, SCORE) and motivation (Possible Selves) in which students learn about what is expected of them in the classroom and how to create a productive learning environment. Students are explicitly taught the appropriate behaviors for classroom situations including lectures, discussions, independent study, and small-group work. The Possible Selves unit focuses specifically on student motivation and involves having students analyze their current lives and then set goals to enhance their futures.²² The behavioral and motivational portion of Xtreme Reading takes approximately four weeks to implement. These units changed in Year 2, as noted on the following pages.

The Xtreme Reading program then shifts to the seven reading strategies: LINC'S Vocabulary, Word Mapping, Word Identification, Self-Questioning, Visual Imagery, Paraphrasing, and Inference. The first three strategies focus on vocabulary development (although the LINC'S model focuses on learning the meaning of new words through memorization, as well as on advanced phonics and decoding for multi-syllabic words). The remaining four strategies target reading comprehension using strategies such as imagery (i.e., teaching students to create mental pictures as they read), paraphrasing (i.e., teaching students to identify and restate the main points of a paragraph in their own words), prediction, and questioning. The program also encourages teachers to support reading fluency through explicit teaching and modeling for students. In addition to the reading strategies, Xtreme Reading integrates writing strategies (such as

²² Data were obtained from the KU-CRL website <http://www.Xtremereading.com>, February 2010.

Paragraph Writing and Theme Writing) with reading instruction. These writing strategies focus on the writing process and thus emphasize planning, writing, providing or accepting feedback, and editing.²³

The Xtreme Reading model uses an instructional approach that involves both teacher-directed whole-group discussions, teacher modeling of strategies, guided practice activities, paired-student practice, and independent practice. Xtreme Reading teachers receive direct training in the Learning Strategies and SIM-CERT strategies as well as ongoing consultation services from the SIM developers (KU-CRL staff). Xtreme Reading instructional strategies fall into six categories: (1) reading, (2) storing and remembering information, (3) expressing information (writing), (4) demonstrating competence, (5) effectively interacting with others, and (6) motivation. These strategies include components of reading as well as class participation. A logic model depicting the key components of the Xtreme Reading intervention as planned and expected outcomes are depicted in Exhibit 14.

Xtreme Reading: Over Time

In Year 4, developers continued to make changes to the professional development model, Xtreme Reading materials, and required assessments (refer to Appendix A, sections A9 and A10, for more information about the professional development received and intervention changes over time). According to the developer, the framework for assessing fidelity of professional development in Year 4 was not based on a defined amount of time, as in Years 1–3, and as required for federal reporting (i.e., numbers of professional development hours as planned and as delivered are required on Annual Performance Reporting or the APR for this grant).

Professional development was administered as needed, based on outcomes as defined by SIM, and not on a specified amount of training time.

Previously, in Years 1 and 2, teachers in their second year of implementation were expected to attend a one-day workshop on Strategy Integration, but second-year teachers had already received training in this content in their first year of implementation. Teachers were expected to

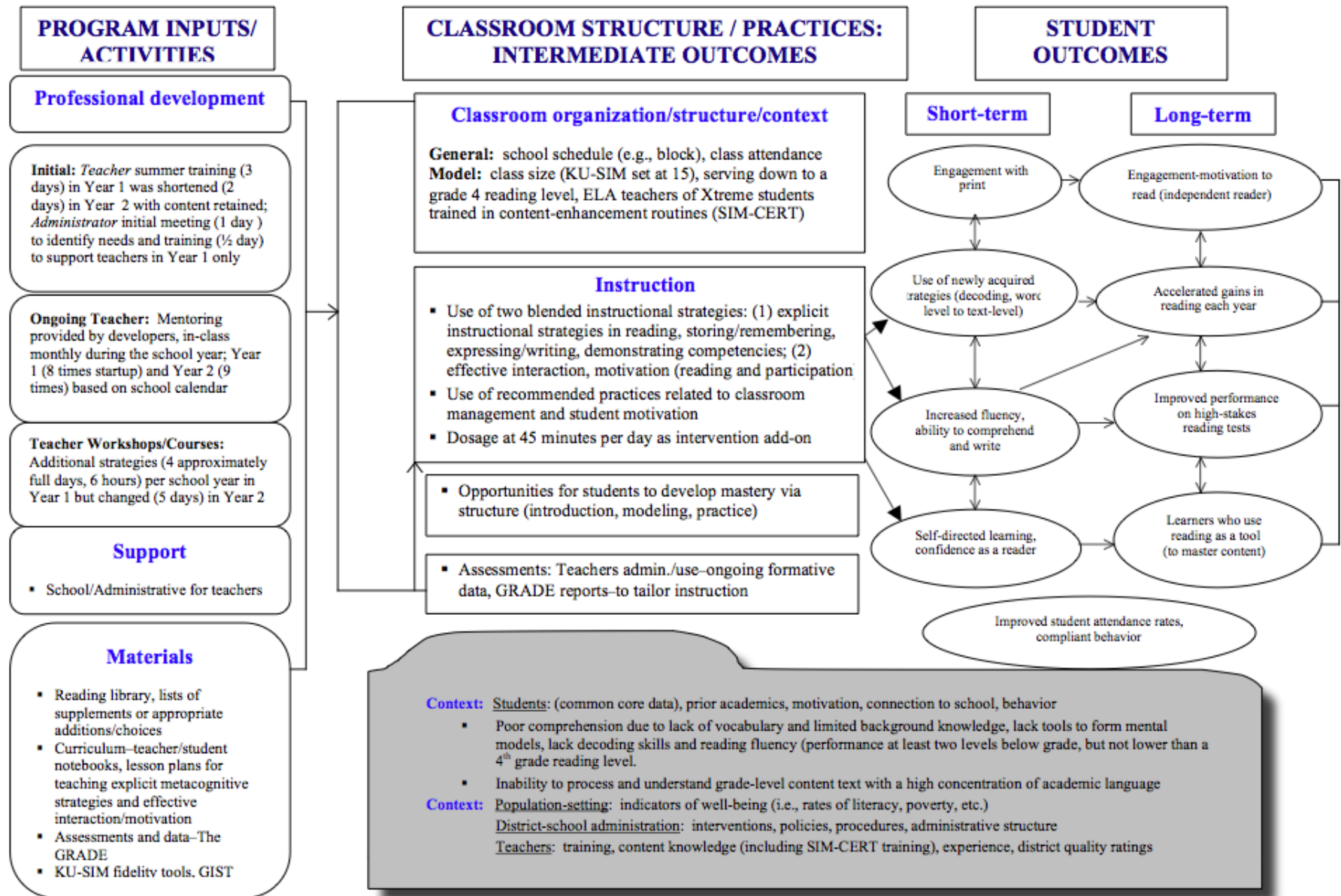
²³ Data were obtained from the KU-CRL website <http://www.Xtremereading.com>, February 2010.

receive professional development inputs during their first year only, with the assumption that this was sufficient to implement the classroom model with fidelity. In Year 3, developers determined that second- and third-year teachers should have ongoing mentoring visits (for a minimum of nine visits per academic school year, or in the case of Chicopee teachers, seven times).²⁴ In Year 4, as in Year 3, Xtreme Reading teachers did not participate in any subsequent SIM-CERT activities. Any necessary SIM-CERT training was embedded in Xtreme Reading sessions or monthly coaching.

Toward the end of Year 2, developers modified Xtreme Reading materials and changed the yearly pacing calendar in response to teacher requests.²⁵ The initial units on student behavior and motivation were abbreviated or covered as needed. In addition to changes in the pacing calendar, more titles were offered in the Xtreme Reading library to address higher reading levels and to provide more variety for students, per the SR district team. SIM-CERT does provide Lexile levels on selections included in the libraries. According to teacher interview data, developers continued to revise teacher and student materials in Year 3. In Year 3, alterations to assessment requirements also changed. Developers required, and then subsequently discontinued, the use of MAZE in Year 3. In Year 4, teachers were asked to submit an additional monthly calendar, which was not aligned to the pacing calendar used since Year 2.

²⁵ SIM-CERT developers reiterated that Xtreme Reading is an experimental version, and revisions have been ongoing during the Striving Readers studies.

Exhibit 14. Xtreme Reading logic model



Whole-School Intervention

As a whole-school intervention, SIM-CERT provides reading strategies to improve literacy instruction across all disciplines. KU-CRL developed these strategies based on over 20 years of reading research. The intervention comprises Levels 1 and 2 of the Content Literacy Continuum (CLC) and is designed to help students understand critical course content (refer to Exhibit 4). The overarching goal of SIM-CERT implementation is to empower teachers to facilitate and students to develop content literacy. Content literacy is defined as the engagement skills and strategies (including listening, speaking, reading, and writing) necessary to process, understand, and master material across a range of academic disciplines.

SIM-CERT: Instructional Approach and Learning Strategies

The approach centers on the provision of meta-cognitive strategies for teachers to evaluate and therefore improve their practice. The developers of SIM-CERT identified three key activities for teachers to enhance their students' understanding of content: evaluate the content, determine the necessary approaches to learning for student success, and teach with routines and instructional supports that assist students as they apply appropriate techniques. By following these steps, teachers identify and demonstrate for students the goal or product of learning while modeling the method by which learning occurs. Teachers assess student characteristics such as intellectual curiosity, interest in the subject matter, and general motivation to learn. Teachers also choose appropriate and customized instructional strategies or routines. By matching instructional approaches with the learning characteristics of students, teachers can differentiate their instruction to meet individual student needs.

KU-CRL noted that the explicit instruction of the strategies is critical for two reasons. First, specificity helps teachers to impart the details of given approaches to students and to be sure students understand. Second, students understand *how* they are learning, in addition to *what* they are learning. There are four categories of strategies, termed Enhancement Routines, which teachers can use in the following areas: planning and leading learning; exploring text, topics, and details; teaching concepts; and, increasing student performance (refer to Exhibit 15).

Exhibit 15. SIM Content Enhancement Routines for Teaching (SIM-CERT)

Planning and Leading Learning	Teaching Concepts
▪ Course Organizer	▪ Concept Mastery Routine
▪ Unit Organizer	▪ Concept Anchoring Routine
▪ Lesson Organizer	▪ Concept Comparison Routine

Exploring Text, Topics, and Details	Increasing Performance
▪ Framing Routine	▪ Quality Assignment Routine
▪ Survey Routine	▪ Question Exploration Routine
▪ Clarifying Routine	▪ Recall Enhancement Routine
▪ Order Routine	▪ LINCing Routine

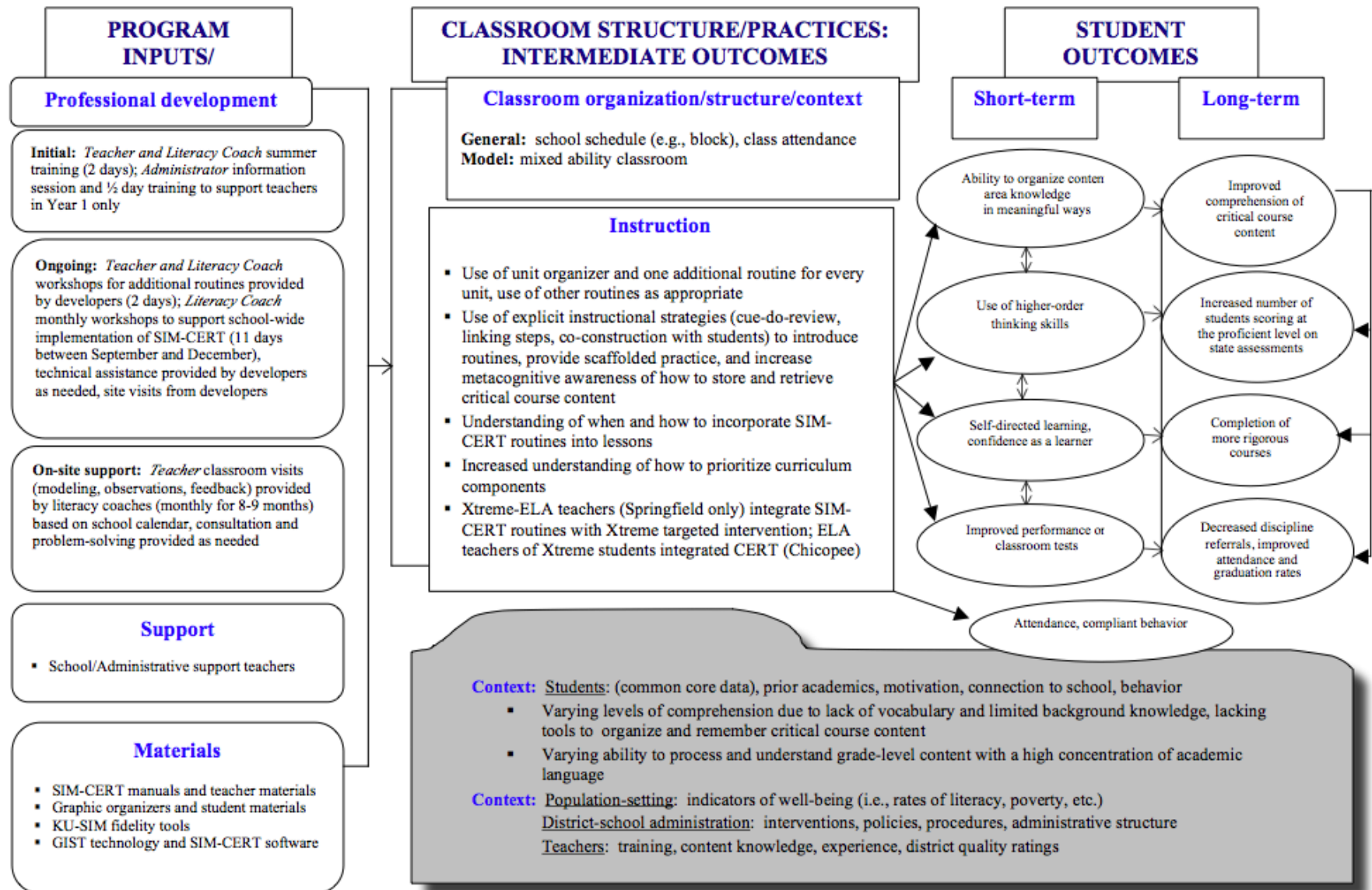
Note. Information provided by Dr. Robinson, University of Kansas, Center for Research on Learning, November, 2007 (Source: Dr. Faddis, RMC Research, Portland, Oregon).

These categories represent the four general task areas that teachers engage in as they evaluate, organize, prepare, deliver, and enhance content delivery for students. Each Enhancement Routine has several subcategories. For example, the first category, “Teaching Routines for Planning and Leading Learning,” has three “Organizer” subcategories—for the whole Course, Units, and Lessons. Teachers choose routines depending on the relevance to the content taught, their needs, and the needs of their department. A school-embedded literacy coach, trained intensively by the SIM-CERT network of trainers, provides ongoing on-site support to teachers as they implement the intervention.

A nationwide SIM-CERT trainer network, overseen by KU-CRL, works directly with teachers and districts to teach, promote, and support the use of these strategies in the classroom in a manner that is customized to school needs. Prior to implementation, individual interviews with teachers allow SIM-CERT trainers to gather information about teacher challenges, student needs, and cultural norms specific to the school. During interviews, trainers explain the content and process of upcoming trainings. Moreover, information from the interviews becomes the basis for vignettes and themes for whole-class training.

Exhibit 16 presents a logic model that depicts the key components of the SIM-CERT intervention (as planned and expected outcomes). Changes from Year 2 to Year 3 are described below.

Exhibit 16. SIM-CERT logic model



SIM-CERT Inclusion Criteria

The Springfield-Chicopee whole-school implementation plans required that all teachers be trained eventually over time. Initial specifications were set for Cohorts 1 and 2 by districts in collaboration with evaluators, to observe training requirements while avoiding confounding targeted study results with the whole-school study results.²⁶ Therefore, teachers in the upper grades (beyond ninth grade where the targeted study—randomized controlled trial or RCT—was implemented) would be given priority in the selection process.²⁷ Participants would be randomly selected from the pool of priority groups (within a discipline so all were trained at the same time).

Inclusion in both SIM-CERT cohorts, based on these criteria, was not planned to occur on a volunteer basis.²⁸ The plan was to randomly select participants from the priority groups. This would be a more equitable process that avoided complications in the interpretation of outcomes given that *all* teachers were eventually obligated to participate in SIM-CERT training over the grant period. In addition, this process would avoid the complications that voluntary enrollment would present for the interpretation of outcomes.²⁹ That is, if only teachers motivated to participate were included, observed outcomes could be the result of this motivation rather than the result (or solely the result) of participation in the program itself. In addition, mandatory district professional development has been the normal context for the SIM-CERT or any whole-school initiative.

²⁷ Refer to Appendix B for the Year 1 and Year 2 criteria (or the Year 2 report).

²⁸ If only those teachers who were motivated to participate were included, observed outcomes could be the result of such motivation. This selection bias would be a threat to the validity of the whole-school study, implemented over time.

²⁹ Selecting from the pool of all required participants, or those identified in groups first, is a method for avoiding selection bias and is often understood to be a more equitable way of including all teachers since all teachers were required to be trained by the conclusion of the grant.

The development of criteria was complicated because developer requirements and research design considerations had to be balanced.³⁰ Other complications in the establishment of criteria for SIM-CERT inclusion over time were: (1) the same teachers delivered both Xtreme Reading and ELA in Springfield, necessitating more individual training in a very tight professional development schedule; (2) professional development was offered and conducted differently in each district; and (3) buy-in for the whole-school intervention and plans continued to present a challenge.

SIM-CERT: Over Time

The Springfield-Chicopee whole-school implementation plans required all teachers be trained eventually over time. Approximately 25 content-area teachers per school would attend SIM-CERT professional development during the first and second year of implementation, a total of 125 teachers per year; 130 inclusive of the five school-based coaches.

Multiple data sources (district and developer documents, literacy coach, district and developer staff interviews, teacher focus groups) suggest that SIM-CERT specifications have evolved over time, reportedly in an effort to be responsive to district context and needs. However, not all changes made were requested by the districts. In fact, the districts often reported challenges resulting from the lack of developer specifications and the changes they made. Throughout the duration of the grant, the developer made adjustments to the program model via their continuous improvement philosophy, altering implementation specifications for both the professional development and classroom models, modifying tools for assessing fidelity to model, and adding a cadre of in-house professional development coaching apprentices (refer to Appendix B for more information about the professional development received).

³⁰ For example, developers initially required ELA teachers of Xtreme Reading students to be included in the SIM-CERT training, adding content to Xtreme Reading teachers' professional development. Subsequently, developers and districts determined that Xtreme Reading teachers should not receive separate training in SIM-CERT to better meet district and teacher professional development needs. In addition, some content units were not yet created for delivery.

Changes to the Professional Development Model

According to grant requirements, adherence to the planned professional development model was measured solely by number of days in attendance at training sessions based on original model specifications.³¹ The following exhibit presents the plans for the delivery of professional development.

Exhibit 17. SIM-CERT delivery of professional development (As planned, Years 1–4)

	2006–07 school year (Year 1)	2007–08 school year (Year 2)	2008–09 school year (Year 3)	2009–10 school year (Year 4)	Total
Cohort 1	Total = 4 days Routines covered: Unit organizer, Framing, LINCing, Concept Mastery	Total = 2 days Routines covered: Course Organizer, Concept Comparison, Integrated Units			6 days
Cohort 2		Total = 4 days Routines covered: Unit organizer, Framing, LINCing, Concept Mastery	Total = 2 days Routines covered: Course Organizer, Concept Comparison, Integrated Units		6 days
Cohort 3			Total = 4 days Routines covered: Unit organizer, Framing, LINCing, Concept Mastery	Total = 2 days Routines covered: Course Organizer, Concept Comparison, Integrated Units	6 days

Note. The plans for Year 3 delivery were last updated November 19, 2009 based on district and developer information including documentation and additional clarifications. Data vary by time period and source.

³¹ A second component of the professional development model, in addition to workshops for teachers, includes mentoring from school-based literacy coaches. As in prior years, developers have indicated there are no minimum requirements for mentoring sessions given these are individualized based on a teacher’s needs (i.e., how often coaches should meet with individual teachers, what activities should be included in each mentoring session). Furthermore, coaching visits were reported by the SR district team to occur often and reported by coaches to occur monthly. Complete documentation indicating the number of visits conducted by literacy coaches and to whom or what individualized instruction was provided during these visits was not received. Therefore, evaluators could not measure fidelity to the coaching component of the model for the whole-school literacy intervention.

Previously, two subcomponents were included in the overall rating of the level and adequacy of planned SIM-CERT professional development: (1) receipt of the *initial training workshops before the first year* of each cohort's implementation of the intervention and (2) receipt of *ongoing training workshops within the academic school year* that built upon the initial training provided.

Two initial and two ongoing full-day training sessions were required for teachers during their first year of teaching SIM-CERT. During a second year of teaching SIM-CERT, teachers were required to participate in two additional ongoing training sessions. In the third year of implementation, the distinction between initial and ongoing training was no longer made given the evolution of training schedules and the second year of training was recommended but not required (reported by the districts as SIM-CERT-initiated and reported by developers as district-initiated). Refer to Appendix B for the most recent professional development plans provided by the SR district team.

In Year 3, districts requested that the content be rated in addition to the training hours to provide a more accurate picture of the provision and receipt of training, especially since training plans had varied over time (e.g., number of days, amount of training per day). Developers supported the district's assertion that the content (e.g., SIM-CERT routines such as Unit Organizer, Framing, and Concept Comparison) should receive greater emphasis than the number of days in which training was delivered, but these data were not available in prior years. Districts compiled these data and provided them to evaluators for the Year 3 and 4 reports.

Changes to the SIM-CERT Classroom Model

Teachers were to be provided with explicit instruction on the routines; to integrate other Enhancement Routines as appropriate into their daily lesson plans; and to co-construct routines with students to encourage and develop active learning, engagement with the subject matter, and independent mastery of the routines. Coaches were to support these efforts to promote implementation and ongoing use.

The following coaching responsibilities remained consistent across the years of the grant: working with SIM-CERT-trained teachers to co-plan, model, and co-teach lessons; co-creating SIM-CERT devices; conducting classroom observations of SIM-CERT implementation; and providing feedback in debriefing sessions. At the beginning of the grant, coaches were to provide additional, voluntary, after-school trainings in SIM-CERT based on teacher interest and need. In Year 3, the frequency of after-school training workshops and observations varied considerably by district and school based on a review of district and developer records. Chicopee coaches reported conducting monthly training sessions, whereas the reported frequency of training sessions was not as high in Springfield. In Year 4, after-school training workshops altogether stopped in Springfield, whereas Chicopee coaches continued providing this form of support. In Year 5, after-school support was not offered in either district.

Similar to the professional development model, specifications for implementation of the classroom model have evolved over time. This evolution complicated district attempts to accurately monitor classroom implementation and provide support. The districts requested more explicit guidelines and measurable expectations for classroom implementation and the use of routines; the developer has reported providing such guidelines in Year 2. However, there was disagreement regarding the appropriateness of the guidelines and materials provided.

Efforts were initiated in Years 3 and 4 to promote sustainability, such as identifying teachers who could assist with training, training new school administrators, and working with instructional leaders to integrate SIM-CERT with other district initiatives. However, many of these efforts were discontinued in the final year of the grant.

IV. Evaluation of the Implementation of the Targeted Interventions

The goals of the targeted implementation study were to present a broad picture of the overall level of implementation in context and a sense of the variability that may have occurred. Differing institutional contexts or constraints influenced the ways in which intervention components were implemented. Districts and schools possessed their own unique complexities, which may have supported or hindered implementation and, in turn, affected outcomes.

The evaluation of the Springfield-Chicopee’s Striving Readers Program implementation focused on *the extent to which the intensive targeted and school-wide interventions were implemented on-model* and also sought to describe the general context of implementation for the interpretation of outcomes. For this study, the extent to which an intervention was “on-model” was the extent to which the targeted intervention was implemented according to the developers’ and districts’ specifications and plans.³² Implementation was evaluated within and across years.

Targeted Implementation Research Questions and Methods

The implementation research questions were developed based on the program models and their intended activities, methods, objectives, and outcome goals. The primary research questions are:

1. What was the level of implementation and variability of professional development/support for teachers/administrators?
2. What was the level of implementation and variability of classroom instruction?
3. What was the context of implementation (e.g., potential influences on implementation)?³³
4. Non-implementation question: What characterized the counterfactual? How did the counterfactual compare to the treatment?

³² Project Officer Communication, November 15, 2006.

³³ This question has been implicit in the evaluation of implementation across years, and data have been collected, analyzed, and reported regarding the general context of implementation but is now explicitly included in this section.

Refer to Appendix A for exhibits including specific implementation research questions within each primary question listed above. Across the areas of implementation, data collection served multiple purposes: (1) to document and assess fidelity of implementation, (2) to determine the level of program implementation, (3) to document variation in program implementation, and (4) to examine variation in program implementation as a potential influence on observed outcomes. Data were collected to assess the presence of relevant contextual factors for both groups of targeted intervention teachers. Finally, data were collected to characterize the counterfactual (i.e., what happens in the absence of a targeted intervention treatment). Although not related to the implementation of the targeted interventions, the assessment of the counterfactual—or rather what occurs as business as usual (e.g., ELA and supplemental reading supports)—provides contextual information for consideration in the characterization of impacts.

Evaluators collected primary data twice per year based on the schedule established in the initial year. District agreements were made with teaching staff (supported by Striving Readers funds) to provide the necessary evaluation data. In addition, districts required other staff with knowledge of Striving Readers implementation or knowledge of the “counterfactual” to participate in data collection activities. The SR district team supported evaluator efforts to obtain complete data and provided secondary data collected while documenting implementation activities. Appendix C includes the multiple measures and data collection methods used for the evaluation of the targeted interventions.

Targeted Implementation Teachers

Random assignment was employed to help ensure that teacher quality would be as equally distributed among the conditions as possible (refer to prior reports for more information as well as Appendix A). A total of 15 teachers study teachers were assigned: five READ 180 teachers, five Xtreme Reading teachers, and five control classroom teachers.³⁴ Teachers also delivered the intervention in upper-grades (10th and above), but control groups were not included in these

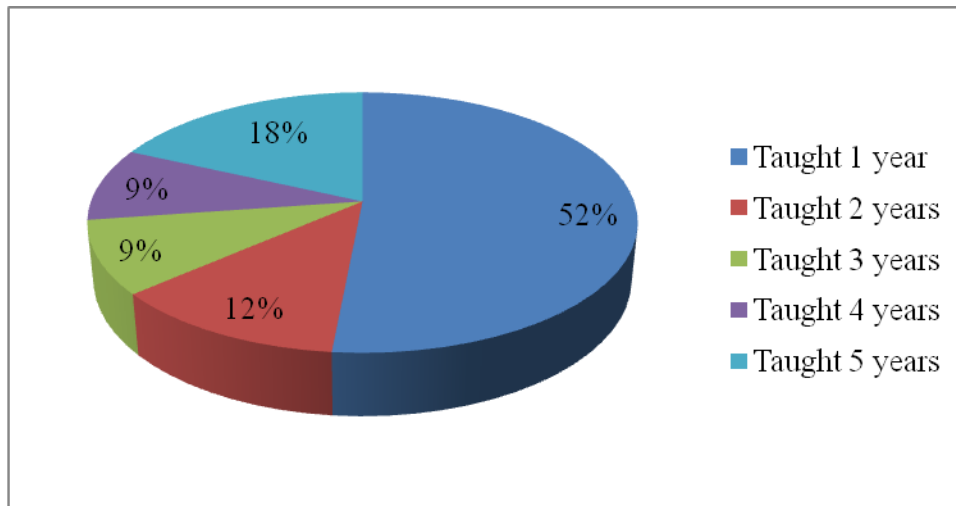
³⁴ As reported in prior years, an additional teacher co-taught READ 180 in one school in the initial grant year. This teacher was not the primary classroom teacher and so is not included in summary study numbers.

grades as per district plans. The numbers of teachers implementing upper-grades were not reported here because they were not a part of the study, i.e., the randomized controlled trial in ninth grade. The districts initially projected larger numbers of teachers to be hired and assigned based on the numbers of students projected to be eligible for assignment. Because the total number of teachers each year was small, differences may be present in unmeasured characteristics among these three groups (e.g., teacher quality).

Characteristics of Teachers: Prior Study Participation

As reported via surveys, none of the teachers Years 1–5 had teaching experience with the reading invention programs prior to participating in the Striving Readers Program. In the case of teacher attrition, the district replaced vacancies with teachers who had experience teaching the intervention in the upper grades. Exhibit 18 below displays the percentage of teachers who taught all study years and 1 through 5 study years in total.

Exhibit 18. Intervention teaching experience by year (As planned, Years 1–5)



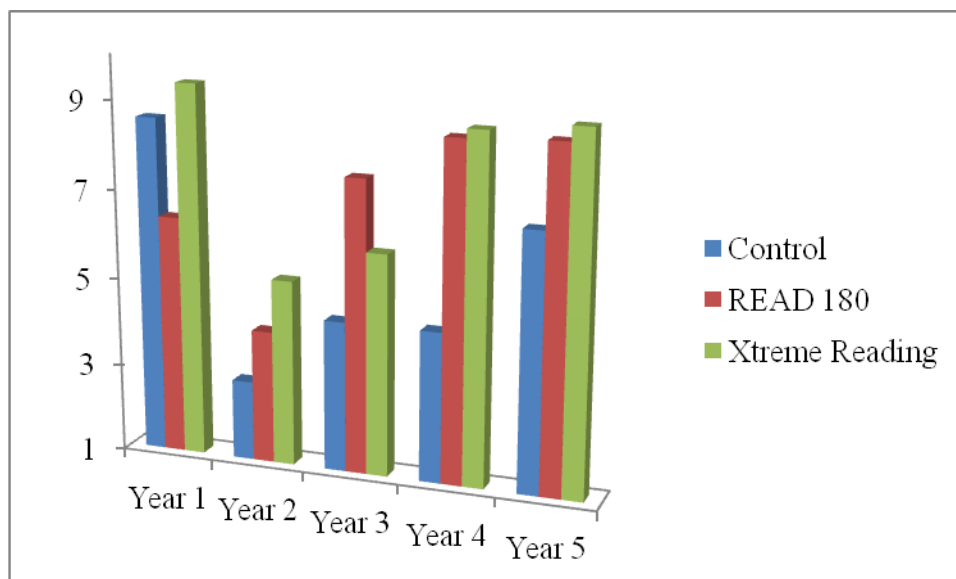
Across the grant years, a total of 33 teachers (13 READ 180 teachers, 11 Xtreme Reading teachers, and 9 control teachers) filled the 15 yearly study positions maintained by the districts after the initial study year. Of the 20 teaching positions replaced, 10 remained in the

district and taught post-study experience and 2 of these teachers returned to the study to teach in subsequent study years. Of the 33 total in the study, 6 taught for all grant years while 17 taught for only one year of the grant implementation. The majority of the 17 teachers leaving the study after one year did so in the first and second year of grant implementation, 8 and 6, respectively.

Characteristics of Teachers: Over Time and Across Groups

Over time, the targeted teachers had more teaching experience and the control classroom teachers had higher levels of education. As a result of teacher turnover, the picture of teacher experience and backgrounds over time is difficult to interpret as a whole, given the changes in teaching staff over time. In part, more overall experience was observed for the treatment teachers given the higher levels of experience in general for Year 1 teachers in comparison to later years and the lower attrition rates for the control classroom teachers over time. The higher experience was likely in part due to the hiring of teachers laid off in the previous year and the lower attrition was likely in part due to fewer requirements and restrictions for the intervention positions in comparison. Refer to Exhibit 19 below.

Exhibit 19. Average years of teaching experience across study years, by group

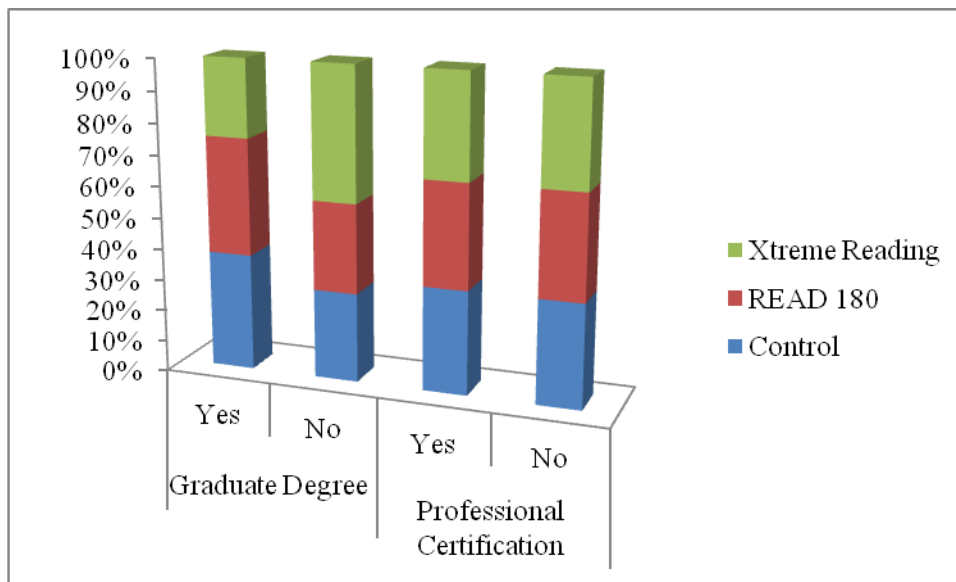


Note. When survey and resume data conflicted, resume data were used for analysis and reporting.

The average number of years of teaching experience for teachers across the five study years was 7.6. Xtreme Reading teachers had the highest average years of teaching experience (7.7) across the five years of the study followed by READ 180 teachers (7.1) and control teachers (5.4). Across the five study years, the average number of years teachers taught at their current school was 3.4.

Xtreme Reading teachers had the highest average years of teaching experience at their current school (4.2), followed by control teachers (3) and then READ180 teachers (2.9). A total of seven first-year teachers participated in the study. Years 1 and 2 had the largest number of first-year teachers (three in total). By Year 4 and continuing in Year 5, there were no first-year teachers participating in the study, based on the replacement strategy used to pull from upper grades and existing teaching staff. Exhibit 20 presents information regarding the highest degree earned by teachers and levels of certification.

Exhibit 20. Percentage of teachers with highest degree and certification by group



Note. Certification information was not consistently collected until Year 3. In Year 5, one teacher did not provide information regarding certification. When survey and resume data conflicted, resume data was used for data analysis. Degree information was not consistently collected until Year 3. In Year 5, one teacher did not provide information regarding highest degree earned. When survey and resume data conflicted, resume data was used for data analysis.

Business as Usual

The counterfactual is addressed by the inclusion of a control group to answer the question, “What would happen in the absence of treatment?” The two components of business as usual for the Striving Readers study included (1) the supplemental services ordinarily available to students in need of additional reading support and (2) the standard ELA courses for all students inclusive of any normally provided reading instruction.³⁵ The first component is the true counterfactual because the supplemental services were to be provided in addition to required ELA courses, as per cross-district plans to ensure consistency of implementation. Therefore, all students in the study, treatment included, were to receive the standard ELA course.³⁶

Standard ELA courses were also examined because control students may receive supplemental supports in this context. An analysis of data collected from district documents, interviews of control classroom teachers and administrative staff, and observations of control classrooms allowed evaluators to note how course content was planned and delivered; what instructional strategies were employed by control teachers; and which instructional supports were provided to struggling readers during, and in addition to, the standard ELA class period. Finally, these data were used to determine any potential study contamination (i.e., the incorporation of targeted intervention materials in class or reported training experiences similar to those of targeted intervention teachers).

In Chicopee, there was little change in the ELA curriculum from Years 1 to 5. In Springfield, the curriculum underwent significant changes in Year 2 in an effort to increase curricular consistency across schools. An analysis of data in Year 3 suggests that these changes included standardized reading selections and assessments, although many teachers continued to implement their own lessons and strategies.

³⁵ Note that business as usual globally consists of all course requirements for graduation as well as exposure to school- and district-wide initiatives. Only those courses and initiatives implemented specifically to enhance literacy are described in this report given the purpose of this initiative.

³⁶ Students identified as struggling readers included Students with Disabilities (SWDs) and English Language Learners (ELLs).

Although various supports were provided to struggling readers across the districts, there was no systematic district-wide approach to identifying and delivering supports to Striving Readers.

In general, students classified as “special education” students had the most access to additional literacy supports outside of standard ELA classes. In the absence of such designation, however, the availability of supplemental supports for students was minimal. Additional reading support was not provided aside from occasional test preparation, teacher tutoring, and a special education English class that was reportedly open to a few non-special education students in two schools. As such, the majority of students in the control group took regular ELA and enrolled in elective courses in lieu of receiving additional reading supports. Teachers reported adapting the general pace and content of lessons to the lower-level reading skills of many of their students. However, they had not received formal training in reading instruction and were not observed to be teaching explicit decoding or comprehension strategies, with the exception of one teacher during one observation conducted to date.

Contamination of Control Condition

As in the past, teacher interviews, surveys, and classroom observations all confirmed a lack of contamination between the reading interventions and the control classrooms for Year 5. None of the control teachers reported experience teaching the interventions in prior years nor had they engaged in SIM-CERT or targeted trainings. In Springfield, however, one teacher reported familiarity with a routine in the context of teaching but may have been referring to general strategy rather than the specific SIM-CERT routine. Control teachers were not observed using the current READ 180 or Xtreme Reading materials, technology, or model-specific instructional strategies nor did they report using these materials, technology, or strategies. Likewise, the unique characteristics of the interventions were not found to be incorporated in the supplemental services that the few control students received. In one district, some of the control students with special needs received instruction with an earlier version of READ 180, version 1.6, per their individualized education plans (this was business-as-usual prior to grant implementation). In addition, prior to entering high school, a small percentage of students received READ 180 version 1.6 services (approximately 15% as reported by the district).

Control Teacher Professional Development

All five control teachers taught ELA courses in grade 9 and above as well as courses. According to survey and interview data, they attended professional development sessions related to the content areas in which they taught. More specifically, session topics were either specific to instruction (e.g., state assessment prompts, ELLs, advanced placement) or more general (e.g., school goals, motivating students). However, the control teachers received no formal professional development in literacy instruction unrelated to the state assessment prompts. Two of the five teachers received support in teaching reading or writing. One teacher had the Department Chair observe her class, present information and co-plan lessons. The other teacher had the Department Chair model lessons for her.

Control Teacher Supports

Instructional materials for lesson planning varied from site to site, although no teacher specifically mentioned reading support materials in reference to their lesson planning. Teachers reportedly sought resources based on personal preference, including prepackaged lesson materials, grammar manuals, the MCAS, and state academic standards. According to teacher surveys and classroom observations, technology use in the classroom was limited to videos, instruction on the overheads, or teacher-led PowerPoint presentations and did not resemble technology used in READ 180 classrooms.

V. Targeted Interventions: Results and Implications

The goal of an implementation study is to gain an understanding of ways in which context may influence study outcomes. It is important for model specifications and implementation plans to be clearly defined to allow for a systematic assessment of implementation levels.

Implementation levels characterize the complexity of the context in a meaningful and understandable way. In addition, defining levels of implementation provides a way to gauge the magnitude of an identified influence on study outcomes. Therefore, this study used a systematic approach to define measurable facets of the interventions and to rate these in comparison to proposed specifications for implementing the Striving Readers Program.

Ratings serve the purpose of providing a snapshot of the implementation level rather than an accounting of every nuance of implementation.³⁷ Implementation scoring is a descriptive process and is not intended to predict (or directly connect to) the impact of the interventions, which are being studied because those impacts under the described conditions are unknown. In addition, data were collected in snapshots and by definition represent only a picture at that point in time. This applies to those teaching in multiple years (i.e., these teachers have a series of snapshots over time). Finally, it is important to note that the interventions were not equivalent, and therefore their ratings should not be compared.

Targeted Implementation Components

Intervention logic models provide the necessary framework for identifying the key components of the targeted interventions to be assessed for implementation fidelity. The logic models reflect what was “planned” by the districts in conjunction with the model developers and thus what was “required” for adequate implementation.³⁸

³⁷ These nuances, though difficult to measure or document, represent potentially important aspects of the interventions.

³⁸ Note that the terms planned and required are used interchangeably in this report.

As per the logic models, each intervention encompassed both specifications related to classroom model implementation and specifications related to the necessary inputs that support delivery of the intervention in the classroom.

Five components were identified to assess the fidelity of implementation of the targeted interventions. The components are as follows:

1. Professional development
2. Materials, technology, assessments
3. Classroom organization, structure, context
4. Classroom model including instructional practice, pacing/dosage, use of materials/assessments
5. Behavior – student ³⁹

Targeted Implementation Component Ratings

The overall rating of adequacy of implementation for the five components was based on subcomponent and indicator scores. Adequacy was defined as the required implementation of intervention components as specified by the developers and planned by the districts. As described previously, the assumption has been that all model components were specified by the developers at the level necessary to promote student improvement in reading skills based on their own research. Therefore, overall quality of implementation was assessed by the overall rating of adequacy of implementation. Each specified subcomponent and indicator were scored based on criteria provided by developers. Fidelity ratings for each subcomponent were then assigned using a binary scoring method. Individual ratings were calculated based on the presence or absence of the subcomponent/indicator (1 = yes, present; 0 = no, not present) or

³⁹ Although student behavior is referenced in developer materials and the logic models, this component was not specified in measurable ways especially given it is both a potential mediator and outcome of the targeted interventions. Therefore, student on-task behavior was included as a separate and *indirect* model component, and not included in the overall implementation scores.

based on whether specific criteria were met (1 = yes, adequate; 0 = no, not adequate).⁴⁰ A score range and percentage were calculated for each primary component based on these subcomponent ratings for each teacher. Refer to Appendix A for a presentation of identified model components, subcomponent indicators, binary codes used for scoring, possible score ranges for each component, and criteria use for scoring.⁴¹

This level-of-implementation rating system is rudimentary and as such captures the adequacy of implementation only and not the quality of implementation. For example, the amount of mentoring provided may have exceeded the amount specified by the model, yet the rating would still be designated as “adequate.” Conversely, if some amount of professional development (e.g., ongoing mentoring) was received but not the model-specified amount, the ongoing mentoring training subcomponent of professional development would not be given a rating of adequate or the highest rating to be obtained.

Targeted Implementation Overall Ratings

The final phase in establishing an overall implementation rating for each of the targeted interventions involved compiling the primary component ratings by teacher and indicating the numbers of teachers achieving the highest level (adequacy). To reiterate, a rating of adequate has been defined as implementation of the intervention *at the expected level given reported model specifications, representing the highest level of implementation*. Composite ratings were created (ranging from 1 to 4) for each primary component.

The overall ratings for inputs consisted of three primary components: (1) professional development participation, (2) provision of materials/technology/assessments, and (3) classroom organization/structure. The overall classroom model rating, as a primary component itself, consisted of the four subcomponents: (1) instructional practices including use of structured

⁴⁰ Two observations were used to increase reliability (an over 85% rate of item-level agreement). The scores were based on the observed occurrence of specific subcomponents in *both* instances. That is, when two observations were conducted for a single teacher, a score of 1 was only assigned if the teacher received a score of 1 for both observations.

⁴¹ Each subcomponent and indicator listed may include more than one item from the data sources used (e.g., observation and survey data) to calculate the rating as previously described.

content, research-based instructional methods, and responsive teaching; (2) dosage, including use of rotations, pacing for the year, and amount of instructional time;⁴² (3) use of materials and/or technology; and (4) use of assessments to inform instruction. Refer to Appendix A for more information regarding components and subcomponents. Summary input and classroom model ratings were created by averaging to calculate overall implementation percentages and associated implementation levels: 1 = no evidence (0–24%); 2 = low (25–49%); 3 = moderate (50–74%); and 4 = adequate or high (75–100%). These summary implementation ratings are presented for both interventions below.

READ 180: Implementation Ratings

The summary of input ratings for READ 180 model implementation is presented by teacher, over time, in the Exhibit 21. For the inputs, all READ 180 teachers received aggregate ratings of adequate or high in Year 5, indicating that the professional development, materials, and classroom structure required for implementation had been provided for the majority of teachers.

Exhibit 21. Summary READ 180 input ratings Years 1–5 (n = 14)

Teacher	Year 1	Year 2	Year 3	Year 4	Year 5
1	Adequate	--	--	--	--
2	Moderate	--	--	--	--
3	Moderate	--	--	--	--
4	Adequate	Adequate	--	--	--
5	Adequate	--	--	--	--
6	Adequate	--	Adequate	Adequate	--
7	--	Adequate	Adequate	Adequate	Adequate
8	--	Adequate	--	--	--
9	--	Moderate	--	--	--
10	--	Adequate	--	--	--
11	--	--	Moderate	Adequate	Adequate
12	--	--	Moderate	Moderate	Adequate
13	--	--	Moderate	Adequate	Adequate
14	--	--	--	--	Adequate

Note. Implementation levels were defined as: 1 = No evidence (0–24%); 2 = Low (25–49%); 3 = Moderate (50–74%); and 4 = Adequate or High (75–100%).

⁴² In Year 1, for Xtreme Reading, dosage was measured in terms of weekly lesson plans but not in terms of units completed over the course of the academic year. In Year 1, several Xtreme Reading teachers did not cover all the units as planned for the year; however, this was not captured in the Year 1 scores. Evaluators added pacing in Year 2.

Only because the ratings were aggregated for professional development, materials, and classroom structure did one of the teachers, new to READ 180, receive an adequate rating as this teacher had not received all of the professional development. All teachers indicated they had enough teacher materials and were provided with the required 90-minute daily class period. Input scores increased from prior years when fewer teachers received moderate scores.

The summary of classroom ratings for READ 180 model implementation is presented by teacher, over time, in Exhibit 22. For the classroom model, four of the five READ 180 teachers received aggregate ratings of adequate or high in Year 5, indicating that fidelity of implementation as defined was achieved. The remaining READ 180 teacher (one of the five) was implementing with a low level of fidelity. Overall, ratings for classroom fidelity increased in Year 5.⁴³

Exhibit 22. Summary READ 180 classroom model ratings Years 1–5 (n = 14)

Teacher	Year 1	Year 2	Year 3	Year 4	Year 5
1	Adequate	--	--	--	--
2	Adequate	--	--	--	--
3	No evidence	--	--	--	--
4	Adequate	Low	--	--	--
5	No evidence	--	--	--	--
6	Low	--	Low	Adequate	--
7	--	Moderate	Adequate	Adequate	Adequate
8	--	Adequate	--	--	--
9	--	Moderate	--	--	--
10	--	Adequate	--	--	--
11	--	--	Moderate	Moderate	Adequate
12	--	--	Moderate	Moderate	Adequate
13	--	--	Moderate	Moderate	Adequate
14	--	--	--	--	Low

Note. Implementation levels were defined as: 1 = No evidence (0–24%); 2 = Low (25–49%); 3 = Moderate (50–74%); and 4 = Adequate or High (75–100%).

⁴³ Overall, ratings for classroom fidelity remained the same in Year 4 as compared to Year 3 with the exception of one teacher (a rating of low changed to a rating of high). In both Years 3 and 4, teachers received moderate scores rather than adequate because they were observed to be behind schedule as per the pacing calendar and did not devote the full 90-minute class period to READ 180 instruction.

Patterns over time were difficult to discern because, with the exception of one teacher, different teachers implemented in Years 1 and 2 as compared to Years 3 and 4. However, ratings remained consistent over time despite teacher turnover in Years 1 and 2, likely due to the district decision to replace these teachers with those experienced in teaching the intervention in the upper grades when new hires and random assignment were not possible. Teachers who continued teaching READ 180 over time had higher classroom implementation ratings over time. Four of the five READ 180 teachers had implemented the intervention in the prior year; one of the four teachers with the highest ratings had taught READ 180 longest (four years as compared to three years for the remaining three teachers).

Xtreme Reading: Implementation Ratings

The summary of input ratings for the Xtreme Reading model implementation is presented by teacher, over time, in Exhibit 23. For the inputs, all Xtreme Reading teachers received aggregate ratings of adequate or high in Year 5, with the exception of one teacher.⁴⁴ The teacher with a rating of moderate for implementation was new to teaching Xtreme Reading for this grade level.

Exhibit 23. Summary Xtreme Reading input ratings Years 1–5 (n = 11)

Teacher	Year 1	Year 2	Year 3	Year 4	Year 5
1	Adequate	--	--	--	--
2	Adequate	--	--	Adequate	--
3	Adequate	Moderate	Adequate	Adequate	Adequate
4	Moderate	Moderate	Adequate	Adequate	Adequate
5	Adequate	--	--	--	--
6	--	Low	Adequate	Adequate	Adequate
7	--	Adequate	Adequate	--	--
8	--	Adequate	--	--	--
9	--	--	Adequate	Adequate	--
10	--	--	--	--	Adequate
11	--	--	--	--	Moderate

Note. Implementation levels were defined as: 1 = No evidence (0–24%); 2 = Low (25–49 %); 3 = Moderate (50–74%); and 4 = Adequate or High (75–100%).

⁴⁴ For the inputs, all Xtreme Reading teachers received ratings of adequate or high in Year 4, as in Year 3. Ratings were lower in Year 2 (two teachers with moderate ratings and one teacher with a low rating), primarily due to the teacher-reported lack of receipt of all instructional materials and, for one teacher, insufficient provision of professional development.

No professional development was required because the two teachers replaced in Year 5 had been teaching the intervention previously in the upper-grades. The lower rating for one of the teachers was due to lower ratings for materials received.

The summary of classroom ratings of Xtreme Reading model implementation is presented by teacher, over time, in the following exhibit.

Exhibit 24. Summary Xtreme Reading classroom ratings Years 1–5 (n = 11)

Teacher	Year 1	Year 2	Year 3	Year 4	Year 5
1	Adequate	--	--	--	--
2	Adequate	--	--	Adequate	--
3	Adequate	Moderate	Low	No evidence	Moderate
4	Moderate	Moderate	Moderate	Moderate	Moderate
5	No evidence	--	--	--	--
6	--	Low	Adequate	Adequate	Adequate
7	--	Low	Adequate	--	--
8	--	Low	--	--	--
9	--	--	Moderate	Moderate	--
10	--	--	--	--	Adequate
11	--	--	--	--	Moderate

Note. Implementation levels were defined as: 1 = No evidence (0–24%); 2 = Low (25–49%); 3 = Moderate (50–74%); and 4 = Adequate or High (75–100%).

For the classroom model, two of the five Xtreme Reading teachers received aggregate ratings of adequate or high in Year 5 (same as in Year 4) indicating fidelity of implementation as defined was achieved. Three of the five Xtreme Reading teachers were implementing with moderate fidelity.⁴⁵ Overall, ratings for classroom fidelity increased in Year 5.

With the exception of one of the four returning Year 4 teachers, all had the same ratings for Years 3 and 4. Implementation results over time are difficult to interpret over time due to teacher turnover in Years 1, 2, and 5. Only three of the five teachers from Year 4 returned (the

⁴⁵ The moderate ratings for the two teachers in Year 4 were the result of these teachers being behind schedule as per the pacing calendar and not implementing core instructional strategies as defined. The teacher who was rated as having no evidence in Year 4 was not observed to be implementing Xtreme Reading content or instructional strategies.

two teachers replacing Year 4 teachers in Year 5 had taught in the upper grades). The moderate or adequate aggregate ratings across time were largely due to the district decision to replace teachers who had left their positions with experienced teachers who had taught previously in the upper grades. However, one of the two longer term teachers had only ratings of moderate, and the ratings were generally inconsistent for this teacher over time.

Targeted Intervention Implications: What Ratings Do Not Illuminate

The goal of the implementation study was to present a broad picture of the overall level of implementation for both of the targeted interventions: READ 180 and Xtreme Reading. Implementation was assessed for each study year, and findings provide contextual information to inform the interpretation of the results from the impact analyses. The implementation study entailed assigning ratings for adequacy based on the presence of observed and reported model components as defined by the developers and the districts prior to implementation. Additional data sources (e.g., documents, interviews, surveys) provided a broad picture of the context of study implementation including teacher perception data regarding satisfaction with the training and support. A summary of the findings is presented in the following pages.

READ 180 Inputs

Professional Development

In Year 5, four of the five teachers were not required to attend any READ 180 professional development hours. These teachers had been implementing READ 180 for four years already, either in ninth grade or upper grades, and did not require further professional development.⁴⁶ The remaining teacher was required to receive professional development as a new teacher, but did not receive all of the training.

In surveys and interviews, Year 5 teachers were generally positive about the on-site support received over the years, but were skeptical regarding the usefulness of the online RED

⁴⁶ Training requirements were specified by the developer.

course. As in prior years, teachers had either mixed opinions or were undecided regarding the quality of workshops and seminars. Year 5 teachers characterized the quality of READ 180 workshops as “hit or miss” and expressed a range of opinions about workshop components. Overall, survey and interview data among Year 5 teachers highlight a decrease in satisfaction levels with professional development received upon reflection and over time (in prior years, these ratings had been higher).

Over time, the 14 READ 180 teachers expressed similar levels of satisfaction with mentoring (high) and the online RED course (low).

For example, most were in agreement that the mentoring, including feedback provided after observations and lesson modeling, were instrumental in helping them to implement the program on their own. Conversely, teachers tended to strongly disagree that the online training component was useful, with only 2 of the 14 teachers mentioning use of the online resources available via Scholastic.

Over time, teachers indicated that they were actually provided with too much training on the whole or that the training was not focused enough on the support of interest to them.

When teachers were asked about the sufficiency of the amount of professional development delivered (including initial training, seminars, coaching, and online RED courses), responses were mixed. Several teachers, particularly those in Years 1 and 2, were critical of the redundancy they reported experiencing in seminars and commented that the same content could have been covered in less time. Teachers suggested that training workshops could have been improved by offering more information regarding which adaptations to the program were permissible while retaining fidelity to model as well as by providing more guidance on assessment techniques, grading, and the use of SAM reports at initial trainings rather than later in the year.

Over time, teacher opinions were mixed regarding the training they received but opinions tended to be related to their level of buy-in with the READ 180 program overall (i.e., the more satisfied with the program, the more satisfied with the training).

Teachers who rated their level of satisfaction as high with the program also tended to be more satisfied with the overall quality of the training they received, while teachers with low levels of satisfaction about the program tended to be more critical of their training. Thus, self-report data on satisfaction with READ 180 professional development may have less to do with the training provided and more to do with teacher satisfaction with delivering the intervention.

Receipt of and Satisfaction with Materials

Two of the four Year 5 teachers who responded to the survey received adequate ratings for provision of intervention materials.⁴⁷ The two teachers with ratings lower than adequate for materials, both in Springfield, reported faulty technology components (i.e., computers and CD players). In an interview, the fifth teacher, who did not respond to the survey and could not be scored, also noted that several technology components were dysfunctional.

Over time, teachers in both districts expressed discouragement regarding the functionality of select items, but confirmed they had enough materials to implement the program.

Teachers commented in interviews and surveys across the years that materials, specifically the headsets and microphones, broke easily and hindered their ability to fully implement the computer rotation as planned. Yet, despite these comments, the majority of teachers over time confirmed that they had received enough materials to implement the model overall.

⁴⁷ Materials include student eBooks, materials for the classroom library, working computers and related gear, working CD players, and topic CDs.

Over time, the majority of teachers reported that students enjoy and benefit from the READ 180 software programs, with a few of the teachers extremely enthusiastic about this program component.

According to the more enthusiastic teachers, the computer rotation created engagement and motivation in that students can observe and track their own reading progress over time. A few teachers strongly disagreed that students enjoyed the software; these teachers were also critical of the READ 180 program as a whole and reported low levels of buy-in.

Over time, the 14 teachers who taught READ 180 Years 1–5 had mixed opinions on the student library materials.

About half of the teachers remarked that students enjoyed choosing and reading books from the READ 180 and Stretch libraries. In Years 1 and 2, some teachers expressed discontent with the limited range of books from which students could select; in terms of the subject matter and the Lexile levels contained within the library.⁴⁸ In Year 3, the developer supplemented the READ 180 library with the Stretch library and later Grolier Online to address teacher comments.⁴⁹ Teacher self-reported levels of satisfaction with classroom library materials increased subsequent to the changes made.

Classroom Structure/Organization

As in Year 4, all teachers in Year 5 received adequate ratings for classroom structure and organization. That is, according to district documents, all classes were scheduled for 90

⁴⁸ Many of these comments regarding the limited range of leveled books coincided with teachers' sentiments that students at higher reading levels were improperly placed in their classes. Because teachers perceived student reading comprehension and associated abilities such as fluency to be much higher, many basing their perceptions on SRI assessments, they may have also perceived the classroom library book Lexile levels to be too low. Refer to the concerns regarding the SRI and error margins presented in prior reports and the impetus for the use of the state assessment as the pretest covariate. A lack of precision in the SRI assessment results may have led to teachers to believe students were higher performing than they actually were; this was confirmed with MCAS and other verification process data prior to placement.

⁴⁹ In Year 2, more complex texts in the classroom library were introduced to students receiving a second year of READ 180 instruction, but not to students receiving READ 180 instruction for the first time. Grolier was introduced in Year 4.

minutes each day and enrolled fewer than 18 students.⁵⁰ In prior years, READ 180 classes at one of the vocational-technical school were scheduled for a double block of 90 minutes every other week, with a single block of 45 minutes during the other week, and did not meet model requirements.

Over time, data from multiple sources suggest READ 180 classes in one of the vocational-technical schools did not occur as planned, and were blended with regular ELA content.

Although Springfield originally intended to blend READ 180 and ELA, they later agreed to the implementation of READ 180 as an add-on to remain consistent with Chicopee.⁵¹ In Springfield, block scheduling provided for an opportunity to blend that Chicopee did not benefit from, so blending was not to occur. Therefore, as planned, READ 180 was intended to be an add-on to the regular ELA instruction that all students received.

READ 180 teachers across years consistently reported that they were required to implement district ELA curriculum within the 90-minute block of time reserved for READ 180 classes. That is, READ 180 at one of the vocational-technical school was inclusive of standard ELA coursework and not implemented as an add-on intervention as originally planned. Teachers at this particular school reported these competing curricular demands as the greatest barrier to implementation, limiting their ability to deliver the appropriate dosage and cover all intervention content. Teachers were encouraged by developers to “blend” or incorporate district ELA objectives into READ 180 instruction so as to meet pacing expectations.⁵²

Over time, multiple data sources indicate that lower than expected attendance rates negatively influenced teachers’ ability to implement the classroom model as planned.

⁵⁰ Or the equivalent number of minutes compiled across a week-long schedule. See Appendix A for more information regarding the Year 5 scores.

⁵¹ Administrators and teachers in Chicopee conducted a review of READ 180 following the award of the grant and commented that it could not meet curricular standards for ELA. Whether it made pedagogical sense to blend the two was not of issue; the districts agreed to implement the intervention consistently as an add-on following the award.

⁵² Although developers encouraged teachers to blend, they noted the blending of ELA as a barrier to achieving fidelity to model in debrief notes and observation protocols submitted in Years 2-5. In Year 5, the READ 180 teacher at this school received low classroom model ratings as a result of this blending of ELA and READ 180 within the 90-minute instructional block.

Although the maximum class size was not exceeded, many sources including READ 180 teacher interviews and surveys, developer observation notes and debriefs, interviews with other key stakeholders within the school and district, and evaluator observations noted the difficulties imposed on teachers attempting to implement with fewer than anticipated students in attendance. An analysis of attendance records and rosters indicated far fewer attending than originally assigned or later placed. In addition, overall class size enrollment was fewer than the maximum in several instances, as many classes were not consolidated based on final enrollment numbers in the fall.⁵³

READ 180 Classroom Model

Instruction Practices: Adaptations to the curriculum, lesson plans, and instruction

In Year 5, four of the five READ 180 teachers received adequate scores for implementing READ 180 with fidelity in the classroom. These teachers implemented content from READ 180 workshops and used program materials to teach this content.⁵⁴

When teachers were interviewed and surveyed, they were asked to reflect on various adaptations they made to the curriculum, lesson plans, and materials. Overall, in Year 5, three teachers reported adaptations to the READ 180 program materials, curriculum, and lesson plans that supplemented and elaborated, rather than reduced, coverage of essential READ 180 content. Some of these changes illustrate district or school-wide initiatives.⁵⁵ Other adaptations were made based on individual teacher discretion and were not consistent across classrooms.

In Year 5, individual-teacher adaptations included: expedited coverage of workshops; use of alternative texts for student independent reading selections; extended writing assignments;

⁵³ Attrition between assignment and enrollment in the subsequent fall was a factor in final class size numbers, for those classes not consolidated in the fall. Complete rosters were provided in the final year of the study.

⁵⁴ As noted in Appendix A, the remaining, new teacher was observed by the evaluator and the developer to be implementing a lesson from district ELA curriculum, using a district-required ninth-grade text, and did not utilize any of the READ 180 workshop materials or content from lesson plans.

⁵⁵ Initiatives included a targeted focus on MCAS prep and open response at the three Springfield schools, and vocabulary through context strategies and writing across the curriculum school-based initiatives at one of the Chicopee schools.

altered sequencing of workshops; and teacher-created workshops to take the place of READ 180 workshops in the curriculum. Two teachers also mentioned adding novels and/or their own instructional materials to the program. Reasons for the adaptations also differed by individual teacher. For example, one teacher reported altering materials and lessons due to his own boredom while another teacher described elaborating on workshop topics to capitalize on student interest and engagement. Two teachers, one of whom reported pressure to prioritize the district ELA curriculum, made adaptations that reduced the extent to which students had exposure to READ 180 program materials, workshops, and lesson plans.

Over time, the types of adaptations made to the curriculum and materials remained consistent, with teachers reporting adaptations that supplemented and elaborated upon the program rather than reducing essential content.

Some of these changes illustrate district or school-wide initiatives while others were based on individual teacher discretion and were not consistent across classrooms. Adaptations based on district or school-related contexts remained fairly consistent over the years, with MCAS prep being a primary focus across districts. However, within the technical-vocational high school, teachers across all five study years reported and/or were observed to be, at times, providing instruction in ELA content at the expense of READ 180 curricular coverage.⁵⁶ In Years 1 and 2, teachers at other schools also reported and/or were observed to be integrating ELA content such as literacy terms into READ 180 instructional time.

Over time, adaptations made at the discretion of individual teachers ranged in the extent to which teachers strayed from the READ 180 curriculum and structure.⁵⁷

Generally speaking, individual teacher adherence to and coverage of READ 180 lessons and use of materials coincided with levels of teacher buy-in with the program. Teachers who

⁵⁶ As observed by the developer, the district, the evaluator, or a combination of the three. Teachers at one of the vocational schools consistently cited the pressure to follow ELA curricular demands as one of the greatest barriers to implementation of the program. As mentioned previously, inadequate scores for classroom model in Year 5 for one teacher were a direct result of ELA requirements and challenges in scheduling READ 180 in addition to regular ELA classes.

⁵⁷ This conclusion was based on the analysis of data from observations across sources as well as interview and survey data and district-provided documents.

perceived that students were benefiting from the program also tended to follow the program structure more closely, while teachers who were more critical of the program tended to alter and add their own materials as substitutions for the lesson plans rather than as supplements. Some of the ways teachers elaborated on the core READ 180 workshops included expanded writing assignments such as five-paragraph essays or double entry journal writing, and the use of alternative texts or reading passages to apply reading strategies and techniques learned in READ 180 workshops.

Rotations, pacing, and amount of instructional time

In Year 5, three of the five teachers were observed to be achieving fidelity in use of rotations. In Year 5, one of the teachers who received a lower rating for rotations was observed to apply more flexibility to individual students switching rotations; the other implemented only one of the three basic rotations in the model, spending the remainder of class time on ELA instruction. Only one of the teachers was aligned with the pacing calendar in Year 5 and two teachers were observed to use the full 90-minute class period for READ 180 instruction.

Over time, ratings for all components increased, but there was variation in how closely teachers adhered to the rotational increments.

Across years and districts, teachers reported collapsing their small- and whole-group rotations, a deviation from the model, due to low attendance and/or low enrollment numbers. However, this modification was later approved by developers.⁵⁸ Because the developer had not yet approved this change in Years 1-3, subcomponent scores for rotations were lower as compared to the same ratings in Years 4 and 5. Apart from modifications to whole- and

⁵⁸ Teachers across years reported in interviews that they encountered difficulties in dividing students into three separate groups for rotations when less than seven students would attend class on a given day. In Year 4, the developer, having also observed the challenges of implementing whole- and small-group as distinct rotations with few students, approved the modification to small and whole group instruction.

small-group rotations, teachers across years tended to eliminate the wrap-up during the final 5-10 minutes of class.⁵⁹

Variation in teacher adherence to the 20-minute increments of rotations was observed. Over time, 7 of the 14 teachers were observed to have shortened or elongated certain rotations. Five of the 14 teachers, including all four teachers at one of the technical-vocational schools, eliminated rotations altogether during one or more observations conducted by the evaluator, district, and/or developer.

Over time, ratings for pacing and use of instructional time were consistently lower than ratings for implementation of rotations and adherence to program content and instructional practices.

Only one teacher received high ratings for pacing in Years 2, 3, and 5, and none of the teachers were on target for READ 180 coverage according to the pacing guide in Year 4. Teachers, districts, the developer, and other stakeholders attributed issues related to pacing, at least partially, to low student attendance.⁶⁰ However, observations indicated that the teachers who were significantly behind in pacing were either those teachers responsible for covering the ELA curriculum or the teachers who liberally added content and material to the rBook as a supplement to workshop coverage.

Finally, although not reported by teachers, observations indicated time management in the classroom may have negatively impacted pacing for the year. These factors related to how

⁵⁹ The elimination of wrap-up was approved by developers in Year 2.

⁶⁰ Teachers explained that it was difficult to cover a workshop when students would be absent and/or late to class. Teachers in prior years reported that additions to the READ 180 course, such as ELA requirements, MCAS preparation, and teacher-created materials and assignments, influenced the pace with which they covered the rBook instructional material.

instructional time was used in the READ 180 class period likely influenced lower individual ratings for pacing over the years.⁶¹

Use of Assessments

In Year 5, three of the four teachers who could be scored for use of assessments received high ratings, indicating that they had: utilized the SAM data reports; administered the formative assessment (i.e., SRI) at least twice during the year; and administered the rSkills tests at least three times, which measure understanding of material covered in READ 180 workshops.⁶² The teacher with a lower rating for use of assessments also received lower scores for coverage of content within the READ 180 curriculum and implementation of rotations as per the model. This teacher was observed by the evaluator and the developer to be implementing content from the district ELA curriculum.⁶³

Over time, teachers generally were in agreement that SAM data and the use of rSkills tests assisted them in their classroom instruction and planning, measuring student progress, differentiating instruction, and implementing READ 180 overall.

Based on survey responses, teachers were in agreement SAM and rSkills tests helped them in their implementation of READ 180 and overall. However, some teachers expressed more

⁶¹ For example, 5 of the 14 teachers were observed to use less of the time allocated for instruction than the other 9 (e.g., having personal conversations with students, assisting students with unrelated subject matter). In some of the more extreme cases, teachers were observed using less than half of the 90-minute instructional period for READ 180 coverage. Reduced READ 180 instructional time was also the result of challenges observed maintaining control of the classroom, especially during transition between rotations.

⁶² In the interviews and/or surveys, all four teachers reported administering the SRI and using SAM data. Three teachers confirmed that they had administered rSkills tests at least three times throughout the year, corresponding with coverage of the READ 180 workshops.

⁶³ This teacher also reported in interviews that one of the primary reasons READ 180 assessments were not administered was that this teacher was responsible for covering both ELA and READ 180 within the same 90-minute instructional block of time.

enthusiasm and reported more frequent use than other teachers.⁶⁴ Over time, variation in teachers' self-reported use of assessments was evident (i.e., types of assessments used and frequency of use).

Xtreme Reading Inputs

Professional Development

According to the developer, the provision of professional development, including mentoring visits, was dependent upon the needs of individual teachers and based on three objectives: (1) new teachers learn the program, (2) teachers get the coaching support they need to improve implementation, and (3) district capacity is built so that professional development can be provided internally. In both Years 4 and 5, the developer did not submit specific guidelines or documentation for the evaluator to measure whether these newly defined objectives were met. Therefore, fidelity to the Xtreme Reading professional development model was no longer assessed. In interviews and surveys, all five teachers reported that they had not participated in any training in Year 5. However, three of the five teachers mentioned that SIM-CERT coaches had visited their classrooms one or two times during the year. According to district and teacher reports, the provision of professional development in the final years was minimal since all teachers had been delivering the intervention for two or more years.⁶⁵

⁶⁴ For example, two teachers specifically mentioned using SAM reports as a method to increase student engagement and ownership of their own reading progress, and utilizing SAM reports to plan lessons and individualize instruction. In contrast, two teachers reported limited use of SAM. Similar variation of frequency of use can be observed in teacher reports of rSkills tests, intended to be implemented every other workshop. In the case of one teacher, students had only completed one rSkills test because READ 180 workshops were not the emphasis of classroom instruction, and two teachers reported creating their own quizzes at the end of workshops. Only one teacher indicated use of Reading Counts quizzes to assess comprehension of independent reading books, and two of the five teachers reported checking fluency, though the frequency with which they did this varied from once per month to twice per month.

⁶⁵ In Year 5, the developer did not provide any documentation of classroom visits to support classroom implementation, nor did they provide a record of meetings with administrators and other staff to promote sustainability of the program at the end of the grant.

Although implementation scores for the professional development model could not be assigned in the last two years of the project, self-report data via interviews and surveys, as well as district-provided documentation of professional development sessions and mentoring visits, were collected and analyzed to provide contextual information about the provision of training for Xtreme Reading and teacher levels of satisfaction with training provided over time.

Over time, the amount and purpose of on-site support varied greatly from teacher to teacher and was not related to levels of classroom implementation.

The amount of on-site support did not have a clear association with either the level of developer-assigned ratings of classroom implementation or the number of years teachers had been implementing the program, as observed by both developers and evaluators. For example, the teacher with the highest level of implementation across time reported the greatest number of coaching visits in Year 5, while the teacher with the lowest levels of implementation reported no visits at all.⁶⁶ Districts reported inconsistency in receiving documentation regarding when visits occurred, for whom, and what took place (e.g., what specifically was done to support the teacher).

Over time, the majority of teachers reported satisfaction with the quality of training and mentor-coaching they received.

Of the 11 teachers who have taught Xtreme Reading throughout the five years of the grant, the majority reported that they were satisfied with the quality of the training workshops and, especially, with the on-site support they received from their SIM-CERT mentor. In fact, five of these teachers explicitly stated that the mentor-coaching was the greatest support they had received for implementation across the years of teaching Xtreme Reading. Teachers explained that the modeling of strategy instruction and the provision of personalized and immediate feedback from the SIM-CERT mentors was the most helpful to them.

⁶⁶ The developer did not submit any summative documentation of on-site support provided to teachers in Year 5.

Over time, the majority of teachers reported the training should have been provided over time or the amount was insufficient in preparing them to teach the program.

Fewer teachers, in contrast, reported satisfaction with the amount of professional development they received. All but two teachers who responded to the survey and/or interview questions regarding professional development explained that they were not satisfied with the amount of training provided and felt more training was necessary. These teachers expressed concern that: (1) the initial training was insufficient to prepare them to teach the program or (2) no training was provided beyond the initial year. This latter group of teachers explained that they would have benefited from a refresher training where developers would demonstrate how to implement various parts of an Xtreme lesson in the classroom, review how to implement specific aspects of the curriculum, and/or illustrate how the various reading strategies could be integrated throughout the year.

Receipt of and Satisfaction with Materials

Four of the five Xtreme Reading teachers received adequate ratings for provision of materials in Year 5 (e.g., teacher materials, student binders and materials, books for the classroom library, and Xtreme Reading posters). The remaining teacher reported via survey that not enough Xtreme Reading posters and classroom library materials (i.e., Bluford books) were provided, though this teacher also reported receipt of all materials necessary to implement the classroom model with fidelity. Over time, the majority of teachers reported via survey that they had received all of the required materials. Several teachers noted that they did not have enough books in the classroom library; however, these perceptions could be due to opinions that the book selection should be expanded to include a greater range of titles (both within the Bluford series and beyond it).

Over time, teachers reported widely divergent views on the quality and usefulness of materials such as lesson plans, teaching tools, and student handouts.

According to survey and interview data, most teachers who taught in Year 1 and/or Year 2 indicated that the materials provided were “among the worst” they had encountered and were riddled with errors, disorganized, dense, and confusing. In response to teacher feedback, SIM-CERT developers reorganized the teacher and student binders in Year 2 and again in Year 3. Teachers who taught Xtreme Reading in later years had more favorable opinions of the materials, with some indicating that these materials were “among the best.” Teachers also had mixed opinions on the quality and usefulness of the assessment tools provided, often changing their minds over time. Teaching experience and/or exposure to other curricular materials for use in reading and ELA instruction may have been a mediating factor in teacher levels of satisfaction with materials.

Over time, a majority of teachers reported the classroom library sparked high interest from their students and stimulated motivation to read.

Multiple teachers across the years reported that the Bluford series was the component of Xtreme Reading that students liked the best and was the most engaging part of the program. Two teachers were critical of the Bluford series, but they noted little student enthusiasm for the books due to low reading levels and personal dislike of the subject matter and characters featured in the series. Teachers, as a group, had more varied opinions on the interest levels and appropriateness of student materials in general, with some teachers specifically reporting that the reading passages in the student binder were neither high interest nor appropriate for use in their lessons.

Over time, all teachers except one had mixed opinions about their ability to cover all material within the specified amount of time in the school schedule.

Teachers were asked via the survey whether the teaching materials and overall structure of the program, including daily lesson plans and pacing guides, were feasible to implement given the scheduled amount of instructional time. The majority of teachers reporting either trouble covering all material or that they absolutely could not cover it all. The four teachers who were most critical of the feasibility of covering all material taught in Springfield. These

teachers were responsible for providing both ELA and Xtreme Reading instruction within the same 90-minute block of time and received inadequate scores for pacing across the years.

Xtreme Reading Classroom Model

Instructional Practices: Adaptations to the curriculum, lesson plans, and teaching

In Year 5, according to observations and self-report data from teachers via interviews and surveys, four of the five Xtreme Reading teachers followed the curriculum and lesson plans with fidelity. These teachers reported and were observed to be using Xtreme Reading program materials (such as Bluford books, student learning sheets, and reading passages) to cover one or more of the six reading strategies in the current curriculum.

Minimal adaptations to the curriculum and lesson plan structure were reported by this group of teachers. Of the small changes observed and reported, the most common were expanded writing assignments, usually the focus of the activator at the beginning of class, and additional vocabulary instruction, an instructional unit that was removed from the Xtreme Reading curriculum in Year 3.⁶⁷ Apart from the small changes described above, one of the four on-model teachers mentioned adding MCAS material and school and district requirements to Xtreme Reading instruction. The other three on-model teachers explained in interviews that they decided not to incorporate additional material beyond what was outlined in the curriculum and program materials. According to one teacher, these additions to the curriculum “take away so much” and “you have to give up something somewhere, there's just not enough time.” As occurred in prior years, one of the five teachers was observed to be making significant changes to the curriculum and content, reportedly implementing self-created lessons on strategies not included in the core curriculum.⁶⁸

⁶⁷ In fact, three of the Xtreme Reading teachers reported in surveys that they spent between four to five days per week, on average, on vocabulary instruction or word study.

⁶⁸ This teacher explained, “The changes I make to the curriculum are all probably very major,” but that “the seminal aspects of the program are still intact and faithfully there, they are things good teachers do anyway, so there's no real trick to it.” When asked how frequently the Xtreme Reading lesson plans were used, this teacher said, “I don't use the lesson plans. They probably exist somewhere but I don't use them.” However, data from observations, interviews, and surveys

Over time, individual teacher interpretations of how to implement the Xtreme Reading program curricular components and instructional strategies varied, with many supplementing rather than supplanting program content.

Individual differences were self-reported but were also apparent in observations conducted by the evaluator, the developer, and the district. Teachers reported and were observed to be making a range of adaptations, mostly as supplements rather than substitutions, to the core instructional focus of the units they were teaching. As in Year 5, teachers across years reported providing additional vocabulary and writing activities to support either reading comprehension for Bluford books or coverage of the reading strategies in the curriculum. Some teachers also reported either shortening or elongating the amount of time spent on certain units of study. In surveys and interviews, teachers reported making these changes based on areas where they felt students needed the most help, what they personally liked best in the curriculum, or a combination of the two.

In Springfield, a common adaptation among most but not all teachers was the blending of ELA and Xtreme Reading instruction. For example, some teachers encouraged students to apply Xtreme Reading strategies such as visual imagery or self-questioning to district-required ELA texts. While blending may be arguably sound pedagogical practice (and was recommended and encouraged by SIM-CERT mentors over time), plans for consistent implementation across districts precluded it. As described previously regarding the blending of READ 180 and ELA, districts agreed to a planned implementation for add-on interventions to regular ELA instruction. In Springfield, block scheduling provided for an opportunity to blend that Chicopee did not benefit from, so blending was not to occur.

illustrate that the lessons taught by this teacher included arguably major modifications to the program such as choice of texts (none of which were Bluford books and one of which was a Scholastic title), instructional focus and content taught (additions and substitutions to the six core reading strategies in the curriculum), and the sequence of instruction over the course of a year (changed from pacing guides).

Dosage: Pacing and amount of instructional time

As mentioned previously, two of the five Xtreme Reading teachers received inadequate ratings for pacing in Year 5.⁶⁹ Both of these teachers were assigned to Springfield schools and were responsible for teaching both ELA and Xtreme Reading within a 90-minute instructional block. In both cases, these teachers were observed to be significantly behind schedule, with one teacher lagging by more than three months. In interviews, these two teachers expressed concern that they would not be able to reach Inferencing, the last reading strategy in the curriculum, before the end of the year. It is important to note that Inferencing is a reading strategy that helps students to activate higher order thinking skills to draw conclusions about larger chunks of text, a skill necessary to access grade-level texts across content areas and meet state graduation requirements by passing the MCAS.

Over time, teachers in Springfield responsible for providing instruction in both ELA and Xtreme Reading were consistently behind schedule, due to ELA teaching or testing requirements. In later years this difference was reduced due to developer changes made to pacing schedules.

Observations from the developer, district, and evaluator, as well as self-report data via interviews and surveys, illustrate that teachers in Springfield lagged behind their counterparts in Chicopee in scheduled intervention delivery. Some of the Springfield teachers abbreviated units of study, shortened lessons, and eliminated certain components in the program. The reasons provided included: (1) ELA curriculum requirements required shorter than the planned 45-minute period of instruction, and (2) testing requirements as per the school or district ELA department. However, scores for pacing in the Xtreme program increased over time, largely due to developer-made modifications to the pacing calendar and curriculum in Years 2 and 3 of the grant. Specifically, the elimination of the socio-behavioral units and the vocabulary unit, and refinements made to the final two units (Paraphrasing and

⁶⁹ One teacher could not be assigned a score for pacing as this teacher was not observed to be implementing a lesson from the Xtreme Reading program.

Summarizing/Inferencing) made it more feasible and likely that teachers would cover all of the curricular units of study.

Over time, data illustrate that patterns related to the amount of instructional time devoted to Xtreme Reading (i.e., dosage) differed by individual teacher.

The three key mediating factors that appeared to determine whether students received Xtreme Reading instruction in the amount specified by the model (i.e., 45 minutes) were: (1) teacher buy-in and satisfaction with the program; (2) teacher ability to manage student behavior and elicit student engagement with material; and (3) prevalence of reported barriers such as ELA and/or district or school assessment requirements as well as low rates of student attendance, which interfered with the timing and delivery of curricular content

Use of Assessments

In Year 5, all five teachers received high scores for use of assessments, pre- and post-unit tests as well as the GRADE. These teachers reported in surveys that they had implemented assessments at least once or twice during the year, the minimum. One Year 5 teacher was highly critical of Xtreme Reading assessment tools and reported relying more on self-developed tests and quizzes on vocabulary and literary content to measure student progress. The other four teachers reported using pre- and post-tests for units in the curriculum to gauge knowledge and understanding of the strategies. Three of the five teachers mentioned using the leveled comprehension quizzes embedded in the Xtreme Reading program to inform instruction and determine whether to review or to move on to the next lesson.

Over time, patterns of Xtreme Reading assessment use varied by individual teacher, with a shift in teacher opinion and actual administration of assessments in the later years.

Over the study years, the developer modified, added, and eliminated assessment requirements and recommendations for use, including the actual assessments teachers were expected to administer and to use to inform instructional planning. These changes over time likely contributed to the difference in opinion regarding assessments among teachers trained in the

beginning of the study to those trained later in the study or implementing in later years. Of all the assessment tools in the Xtreme Reading program, the most individual variation was observed for fluency assessments, with teachers reportedly administering them once per month to every day.

Over time, teachers provided limited information about how they assessed student growth in reading comprehension and other indicators of literacy development to inform instruction.

Only one teacher mentioned the implementation of progress monitoring or involving students in understanding their reading progress over time. Developers documented their concern with “variability” in assessment use in Years 3 and 4, noting that “overlooking this area of instruction results in lack of feedback to students regarding their performance and guidance in how to effectively put the strategies into practice.”

Cross-Targeted Intervention Barriers

Finally, the factors influencing implementation across interventions (i.e., districts and schools) were driven more by the context of the interventions than the interventions themselves.⁷⁰

Although these points were made within each intervention and related to inputs and classroom model, they are listed here because they relate more specifically to contextual circumstances within the districts than to the intervention specifications or requirements themselves.

Previously noted barriers across interventions included: (1) requirements to teach ELA and blend requirements with the interventions which were supposed to have been add-on; (2) low attendance and smaller class sizes interfering with on-model delivery (e.g., timing, rotations); and (3) requirements for assessments, both internal to the programs and external (e.g., MCAS, formative).

ELA requirements and blending. Intervention teachers were only expected to teach intervention courses; any requirement to teach ELA was a district one, likely for convenience

⁷⁰ Data sources providing information triangulated here include focus groups and interviews with administrators across districts as well as the teacher and developer interviews, over time.

especially when 45 minutes of Xtreme was to be taught within a 90-minute block scheduling framework. These concerns also appeared more specific to Springfield because they had the 90-minute block scheduling.⁷¹ Teaching a back-to-back ELA and intervention courses classes was a convenience that may have led to unplanned blending. Developers encouraged blending in the case of both interventions, and though it may make sound sense to blend intervention content with teaching ELA, it did not in the larger picture of consistent implementation plans across districts as required. That is, districts agreed the interventions would be implemented as supplements to regular ELA instruction and therefore these courses were not scheduled back-to-back with ELA for the same group of students across all schools.

Low attendance and smaller class sizes. Although maximum class sizes were dictated by the intervention developers, minimum class sizes were not. As explained in the previous sections as well as in the following section describing final sample sizes for the impact analyses, class sizes were not unduly small overall, despite the fact that districts and/or schools did not combine classes when final enrollment was settled in the fall as anticipated. However, attendance rates appear to have been varied over time, and fewer students than anticipated did impact classroom implementation in terms of rotations and pacing. Requirements were made less stringent by developers over time in response to these concerns, so no further impacts should have been noted.

Assessment requirements. Smaller class size was reported as a concern by teachers and as a barrier to implementation as were district assessments and the addition of intervention assessments to track ongoing progress.

Teacher buy-in and satisfaction with the program was a previously noted barrier for both interventions. However, although rates of buy-in and satisfaction were related to the program

⁷¹ As reported, Springfield originally intended to blend READ 180 and ELA. They later agreed to the implementation of READ 180 as an add-on to remain consistent with Chicopee, as required by ED. [Administrators and teachers in Chicopee conducted a review of READ 180 following the award of the grant and indicated their opinion it could not meet curricular standards for ELA. Whether it made pedagogical sense to blend the two was not of issue, the districts agreed to implementing the intervention consistently as an add-on following the award.] Therefore, as planned, READ 180 was intended to be an add-on to the regular ELA instruction that all students received.

itself, they were also likely related to contextual circumstances within the districts or the study assignment process itself. That is, as originally planned, districts were to hire new teachers to deliver the add-on interventions. In practice, many teachers were laid off prior to the beginning of the study due to budget cuts and some were rehired as the “new” teachers to be assigned. Buy-in and satisfaction problems may have been inherent in this approach; those hired back in this new role, though perhaps happy to have a job, were not likely happy that the job was a substantively different one.⁷²

Finally, recommendations made by select teachers and administrators included: (1) placing more students in the intervention classes and making it easier to schedule students into these classes; (2) making it easier to implement interventions in 45-minute blocks and not require ELA be replaced; and, (3) making the interventions more available to SPED and non-ELA students.

Unfortunately, these recommendations reflect a lack of understanding of district implementation plans, developer specifications, and the requirements of the rigorous study design. These requirements include the screening level within which a student must be assessed to be eligible for placement, the verification processes for placement and scheduling, and the inclusion/exclusion requirements for SPED and ELA students (these students were only excluded if deemed to be functioning at levels developers indicated should exclude them). Such recommendations were reported throughout the grant period but less so over time given the SR district implementation team efforts to inform each and every person involved of the plans and requirements. The short start-up phase for the grant hindered the team’s efforts to engage buy-in, inform staff district-wide, and work with administrators on outlining requirements for the receipt of grant funds and the accountability plans in the event requirements were not met.⁷³

However, these concerns/perceptions also reflect authentic concerns regarding how best to serve the needs of struggling readers and were genuinely representative of the challenges the districts

⁷² Those assigned to business-as-usual were most likely to be teaching as they had in the past if they were teaching ELA. Students receiving normally provided services were “accounted for” in ELA classes with control teachers, where they received whatever additional services for reading were normally provided.

⁷³ ED later revised their phase-in schedule of the second cohort of Striving Readers grantees based on the first cohort of grantee and evaluator recommendations.

encountered and sought to overcome while implementing the interventions (targeted and whole school).

VI. Evaluation of the Impacts of the Targeted Interventions

The Springfield and Chicopee School Districts implemented two targeted interventions for Striving Readers, READ 180 and Xtreme Reading, in five high schools across the two districts.⁷⁴ The primary research question addressed by this study as required by the grant is: *Does participation in a reading intervention increase reading achievement?*

To assess the effectiveness of the interventions, a randomized controlled trial (RCT) was employed. Eligible incoming ninth-grade students were assigned to one of three conditions: Control, READ 180, or Xtreme Reading.⁷⁵ Each of the treatment group impact estimates—for READ 180 and Xtreme Reading—was assessed in comparison to the control group. Because students were randomly assigned to intervention groups, students are the primary unit of analysis.⁷⁶ To answer the primary research question regarding the effectiveness of the interventions and to provide estimates of their “true” effects on reading achievement, average reading achievement scores of students in each of the two interventions were compared to the scores of students in control group classrooms, pooled across sites and study years.⁷⁷ Power estimates based on the numbers of students in the ninth-grade cohorts are included below.

⁷⁴ One additional high school in Springfield is not included in the grant and is not part of the study sample.

⁷⁵ Although these interventions were also implemented in the upper grades (10th, 11th, and 12th) as per the districts’ request a control group was included only in ninth grade. Therefore, only ninth grade students were included in the impact analysis.

⁷⁶ Randomization of teachers was also conducted, which was possible because new teachers were hired with the agreement they would be placed at random in one of three positions: READ 180, Xtreme Reading, or Control (business as usual). Refer to Appendix A for more information regarding teacher assignment.

⁷⁷ Note that cohort in this instance is equivalent to year (e.g., Cohort 1 was treated in Year 1). Because students were randomly assigned to intervention groups, they are the primary unit of analysis.

Measures, Screening, and Random Assignment

The primary outcome for the analysis of student impacts is the Stanford Diagnostic Reading Test, Edition 4 (SDRT-4).⁷⁸ The SDRT-4 score comprises four key indicators of reading achievement: decoding (phonetic analysis), vocabulary, comprehension, and scanning.⁷⁹ This assessment was administered to all students school-wide, including struggling readers, by the districts in the spring of each year.

The Scholastic Reading Inventory (SRI) was used as the districts' screening tool as this assessment was already in use in some of their schools. The Massachusetts Comprehensive Assessment System (MCAS) English Language Arts test was used as the covariate in the analytic models to control for prior reading achievement level. The rationale for the inclusion of the MCAS as a covariate rather than the Scholastic Reading Inventory (SRI) is described in more detail in Appendix D.⁸⁰ This appendix also includes a summary of the data collection process and psychometric properties of the measures used for the estimation of student impacts.

Screening as Planned

All incoming ninth-grade students identified as struggling readers based on the screening process were included in the pool for random assignment to interventions. The SRI has overlapping Lexile levels and, as a result, the range for identifying eligible incoming ninth-grade struggling students had to be established (therefore, the 50th Normal Curve Equivalency or NCE was used as the benchmark). Refer to Exhibit 25 below for the established screening range.

⁷⁸ The SDRT-4 was also administered to participating struggling readers in the fall of the first two school years (2006–07, 2007–08) to further assess placement via the district screening process but later eliminated due to the burden on students and teachers. Data collected by the districts in the 2007–08 school year were not available for analysis in Year 2, but were provided following the Year 2 reporting period.

⁷⁹ The SDRT-4 serves as both the outcome measure for the impact analysis as well as the screening measure for identifying struggling readers in grades 10–12 (students not included in the RCT).

⁸⁰ The preliminary impact analyses conducted in the first year included the MCAS for seventh and eighth grade ELA separately to assess any potential impact use of the seventh grade MCAS would have. The correlation in the combined sample between the seventh and eighth grade MCAS scores remained $r = .56$. (Refer to the Year 2 report.)

Exhibit 25. SRI ranges from norms file: Unpublished data provided by Scholastic ⁸¹

Student enrolled grade level (spring)	Reading level	Minimum SRI-Lexile score (50 th NCE for 4 th grade)	Maximum SRI-Lexile score (50 th NCE for two grades below)
8 th	6 th – 4 th grade	680	855

Districts established testing schedules and assessment protocols for the administration of screening. The SR district team worked with the middle schools to screen the incoming ninth-grade students in their final months of eighth grade to ensure they could be assessed for eligibility and scheduled as appropriate prior to the fall. The SR district team worked with Scholastic to implement the SRI online so that it could be used for both assessing students at baseline and, subsequently, for monitoring progress in READ 180 over time. The districts provided the student test data, which evaluators then used to randomly assign students.

Several steps were taken to review the accuracy of the SRI assessment scores. Once randomized, district and school staff members reviewed the assignments and discussed any concerns with evaluators as well as potential exclusions.⁸² Refer to Appendix D for information provided to district staff regarding this process. A careful review of the eligibility of each student was conducted school-by-school and the SR district team, based on criteria established for exclusion (including prior grade history and MCAS performance) to avoid solely basing the decision on the SRI score in the event individual performance differed from actual eligibility. Students were excluded from the study if they met any of the following criteria: (1) their Individual Education Plans (IEPs) explicitly specified a different form of reading support; (2) they lacked the necessary English language or comprehension

⁸¹ Scholastic provided secondary data used to establishment this range or threshold.

⁸² School and district responsibilities are the same but referred to here as “school” responsibilities. FTP is the file-transfer protocol site established by the evaluator to maintain data confidentiality as per data sharing agreements. Research protocols and requirements were established whenever possible in collaboration with the SR district team. The district maintained responsibility for communicating with their staff regarding all Striving Readers activities. However, the SR district team worked with evaluators to distribute information about the research study, schedule information sessions at staff meetings, and hold question-and-answer sessions about the study at each of the schools.

skills; (3) their parents formally refused participation in the interventions;⁸³ (4) they were enrolled off-campus in a “twilight school,” an evening program without a Striving Readers Program, or in an “early college high school,” a college preparation program;⁸⁴ (5) they had high grade histories and MCAS scores that were at least proficient; or (6) they were deemed “inactive” by the districts, meaning that the district was not able to determine whether they were enrolled in any of the schools.

Randomization Process as Planned

Approximately equal numbers of students were assigned to one of the three conditions. Randomization was conducted by the evaluator. Pre-randomization blocking of students (by special education and ELL status) was employed where numbers permitted, to ensure the similarity of students across groups on observable characteristics relevant to the outcome and to increase the precision of impact estimates.⁸⁵ Sample size estimates did not exceed the districts’ ability to serve; therefore, all those students screened and eligible were to be included in the pool to be randomly assigned.⁸⁶

The exhibit below represents the random assignment process as planned.

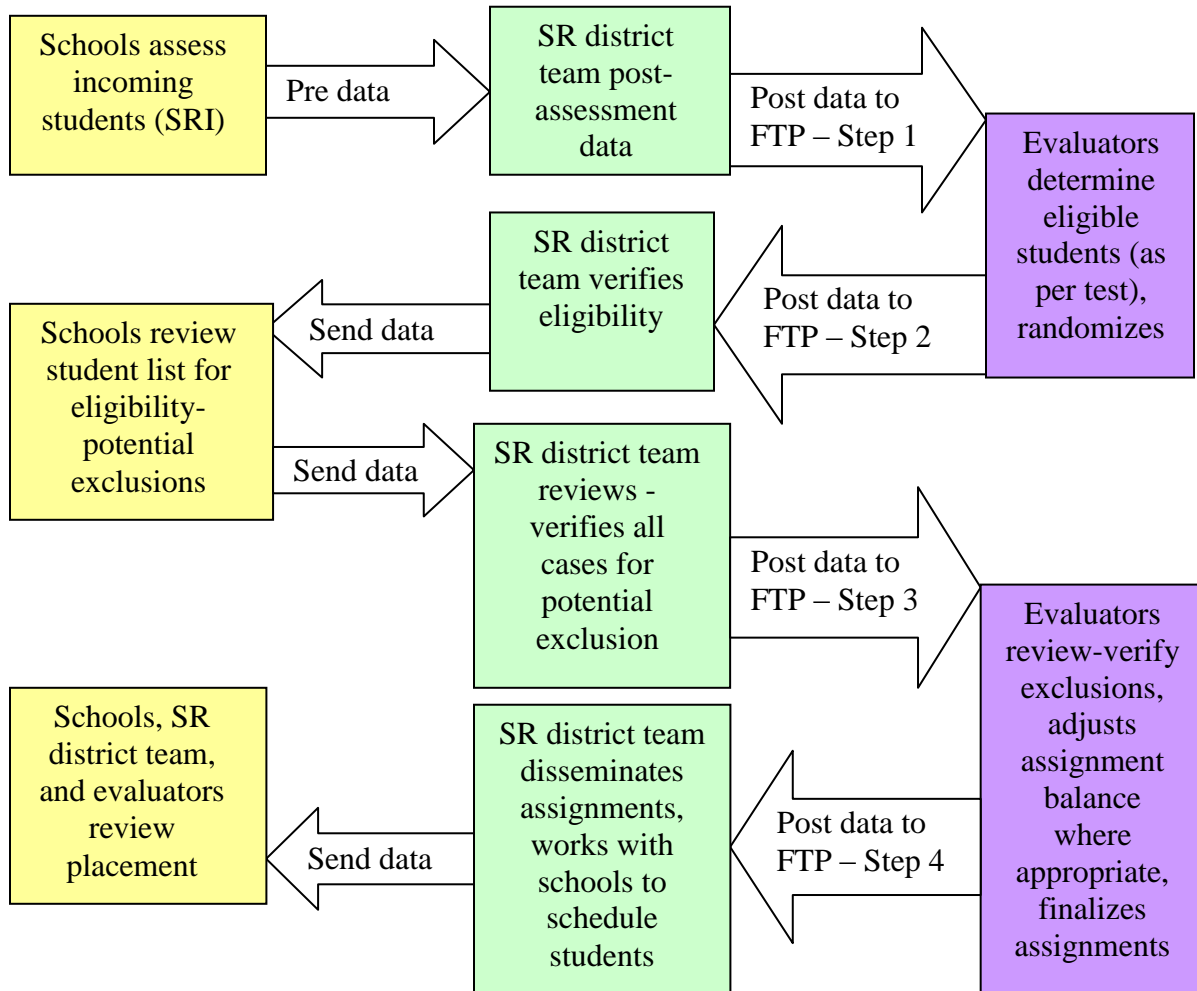
⁸³ Parents with questions about student placement spoke to the coordinators in either district, and then discussed concerns with the vice principals or principals. If, after an explanation of the study and placement parents still requested the student be removed, they were asked to provide a letter stating their request to not have their child participate and the student was removed from the intervention class. No parent refused to have their son or daughter participate in ninth grade.

⁸⁴ Off-campus enrollment was the case only in SPS.

⁸⁵ The constraint placed on the range of struggling readers to be identified left little opportunity to block on levels of screening status (Xtreme Reading serves only those students reading at a fourth-grade level or higher).

⁸⁶ Students who were reading below a fourth-grade reading level would not participate in the study but would receive the supports and interventions normally provided by the district (i.e., business as usual). Special education students whose Individual Education Plans (IEPs) stipulate that they receive services different from the interventions were excluded from the study. Students enrolling in schools after the fall verification period (mid-October) would not participate in the study that school year.

Exhibit 26. Processes for the final randomization (Ninth-grade screening test)



Following the receipt of SRI scores, evaluators randomly assigned students to one of the targeted interventions or the control group. This process occurred over approximately a one-week period, given that complete data were provided including grade, school, state identification number, and other data used for assignment within strata.

Final Sample

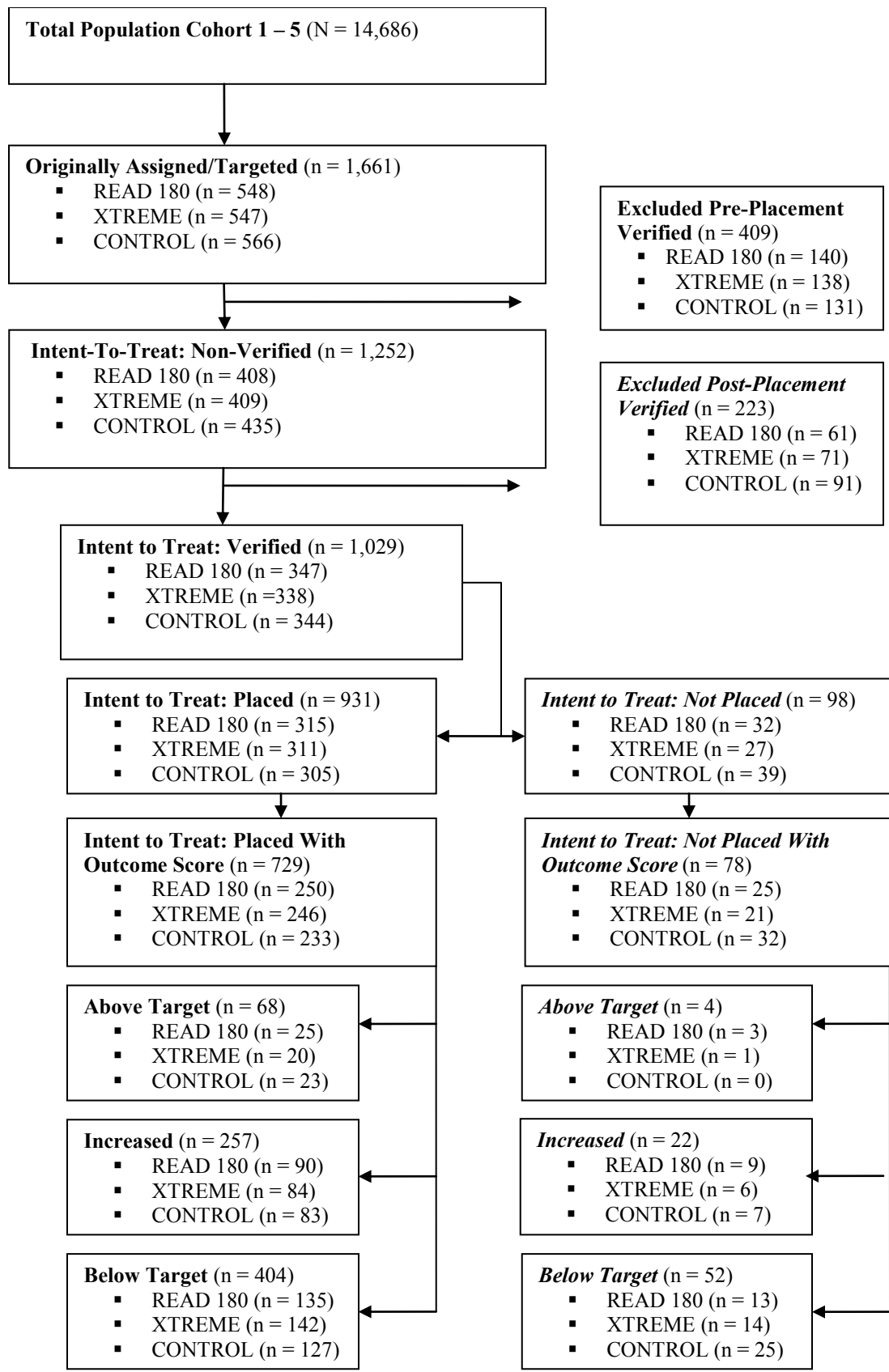
Student Screening and Random Assignment

Five cohorts of ninth-grade students from the 2006–07 through the 2010–11 school years have participated in the RCT.⁸⁷ All cohorts have been combined for the final analysis of targeted intervention impacts.

Exhibit 27 illustrates the size of the sample at each stage of the study. Post-placement exclusions took place prior to or at the onset of the school year as incoming student schedules were adjusted in conjunction with normal school year start-up operations. Verification was also required at this time because assignment took place in the summer when test data for the fall assignment were provided and when most staff had already completed the school year. The same valid exclusion criteria were applied during post-placement as for pre-placement. District-provided reasons for the numbers of students assigned but not placed related to difficulties with enrollment, scheduling, and verification in general and did not systematically differ across the three assigned groups. Refer to Appendix D for more information regarding exclusions.

⁸⁷ Refer to the following section for a description of sample power and for more information regarding the number of cohorts.

Exhibit 27. Screening and assignment and sample



Intent-to-Treat

Exhibit 28 presents the final number of students in the Intent-to-Treat (or ITT) condition for Years 1-5. The ITT group forms the basis for the analytic sample as it is comprised of all those students originally assigned at random.

Exhibit 28. Final numbers of the Intent-to-Treat randomly assigned students by school

Assignment	Cohorts 1–5 Total					TOTAL
	CCHS	CHS	Commerce	Putnam	SciTech	
Control	59	44	66	86	76	331
READ 180	49	44	72	89	79	333
Xtreme Reading	54	42	59	87	76	318
Total	162	130	197	262	231	982

Approximately 10% of the ITT group over time (98 students) had initially been reported inactive by the SR district team but were actually in attendance at least 75% of the time, based on both rosters and district attendance records. Of these students, 20 did not have outcome scores. Given attrition following the eligibility assessment conducted in the spring prior to the fall placement, the overall ITT sample is reduced to those with outcome scores. A total of 807 students had outcome scores of those in the full ITT group ($n = 982$), and 684 of these students had both pretest (MCAS) and post-test (SDRT-4) scores; this matched group is considered to be the final analytic sample.

Power to Detect Effects

Minimum detectable effect size (MDES) estimates have been computed to determine whether the study design provides sufficient power to detect an impact if one exists for either intervention. The MDES indicates how small an effect the intervention can have on students' reading

achievement and still be detected (Orr, 1999).⁸⁸ Current MDES calculations were calculated for a single-level trial as developed under Optimal Design (Raudenbush & Liu, 2001; Raudenbush, Spybrook, Liu, & Congdon, 2004).⁸⁹ Specifications for the power estimates in Year 5 met the desired 80% power to detect an effect with two-tailed tests of significance (at the .05 significance level).⁹⁰ The following exhibit presents the power estimates for the pooled cohort samples, including the MDES with the pretest covariate. Each of the two intervention groups of students (Xtreme Reading and READ 180) were compared to the control group of students in the same model.

Exhibit 29. MDES for pair-wise comparisons: By N of students and covariate

Number of Students		Minimum Detectable Effect Size (σ) By Covariate Correlation	
		<i>No covariate</i>	<i>r = .47*</i>
3 Cohorts	<i>N = 406 per contrast</i>	.28	.25
4 Cohorts (estimate)	<i>N = 500 per contrast</i>	.25	.22
5 Cohorts (estimate)	<i>N = 600 per contrast</i>	.23	.20

Note. Covariate *r*, .80 power, 5% significance level, two-tailed test. In Year 2, combined three-cohort estimates were *n* = 376 with .29 for the MDES estimate, .25 for the MDES estimate with the inclusion of a covariate ($r^2 = .27$). Current estimates were almost identical.

⁸⁸ Effect sizes are reported on a scale of 0 to 1, and the higher the score, the greater the magnitude of the treatment effect (Cohen, 1998; Lipsey, 1990). The framework used to assess the magnitude of effect sizes was Cohen’s (1988): .20 as small, .50 as moderate, and .80 or above as large (as cited in Bloom et al., 2005). “This interpretation is supported by Lipsey and Wilson’s (1993) review of meta-analyses across psychological, educational, and behavioral outcomes, which concluded that effect sizes of 0.10 to 0.20 should not be seen as trivial” (Vernez, & Zimmer, 2007). More recent research provides other empirical benchmarks for evaluating effect sizes related to education-focused interventions (Bloom, Hill, Rebeck Black, & Lipsey, 2007; Vernez & Zimmer, 2007). Vernez and Zimmer (2007) recommend interpreting effect sizes from data related to educational interventions aimed at positively impacting student achievement levels as follows: 0.05-0.10 as small, 0.15 as medium or moderate, and 0.25 as large.

⁸⁹ Initial power estimates were based on a two-level framework and the planned assignment of teachers/classes. However, the number of teachers was fewer than anticipated and resulted in only one teacher per condition, per school—effectively rendering teacher equal to school in these analyses (which is insufficient for multilevel modeling using classroom as the cluster).

⁹⁰ In Year 1, estimates of the correlation coefficients between pretest scores, or prior achievement scores, and post-test scores at various levels were made in the absence of the availability of actual data (Raudenbush, et al., 2004; Bloom, 2004). In Year 3, there was a relatively weak, statistically significant relationship between the SRI and MCAS ($r = 0.21$, $p < .01$) and the SDRT-4 ($r = 0.22$, $p < .01$). There was a moderate, statistically significant relationship between the MCAS and SDRT-4 ($r = 0.47$, $p < .01$); this correlation was used in the current power estimates presented.*

The MDES estimate was .23 for the five-cohort study. Including the MCAS ELA prior achievement score as a covariate $r^2 = .27$ lowers the MDES estimate to .20 for the five-cohort study.⁹¹ Blocking was conducted for student assignment by school and grade but also by disability and ELL status, which should increase the precision of estimates (Raudenbush, Martinez, & Spybrook, 2005).⁹²

Statistical Analyses

The analysis is designed to estimate the impact of the two interventions separately by comparing the achievement scores of each treatment group on average to that of the control group. Using reading scores from standardized assessments taken in the spring of the ninth-grade year, student performance in reading for each of the two treatment groups will be compared with the control group.⁹³ Cohorts of ninth-grade students, five in total, were combined for analysis. As described previously, given projected and actual power estimates, a third (2008–09 school year), fourth (2009–10 school year), and fifth (2010–11 school year) cohort were added with control groups, which yielded a larger than originally planned sample included for final impact analyses.

Analyses were designed to answer the research question *Does participation in READ 180 improve ninth graders' reading achievement relative to that of a control group?* using students as the primary unit of analysis. A fixed-effects approach using OLS regression was used and is presented here as in the past for ease of interpretation. Four indicator variables were entered for the five high schools in the final model. In addition, random effects were also assessed despite the limited number of schools.⁹⁴ Multilevel models were fit to determine the amount of variance

⁹¹ Results approximate those presented in research scenarios estimating sample size for randomized trials, though many of the estimates presented in past research included higher pre-test covariate correlations (refer to Bloom et al., 2006).

⁹² Although blocking by screening level was initially proposed, it was not ultimately pursued due to the restricted reading-level threshold (two levels below grade down to a fourth-grade level) imposed by the Xtreme Reading developers. This threshold yielded a smaller pool of striving readers than originally anticipated. Data for blocking were provided by the districts each year at the time of assignment.

⁹³ As per district request, after one year, students in the ninth-grade control groups are randomly assigned to one of the two interventions for 10th grade if they are not yet reading at or above grade level.

⁹⁴ Recall that students are the primary unit of analysis. Although there was random assignment of students (and teachers), students remained clustered within schools and, if clustering was not accounted for, the standard errors could be miss-specified and overestimate treatment effects (Raudenbush & Bryk, 2002). Despite the small number of schools or the “n” for the cluster level, multilevel models were also fit using SAS. Refer to results included in Appendix D.

in reading scores to be predicted (92% at the individual student level; 8% between schools) and to assess the percentage of this original variance explained by the final model.⁹⁵ Results from these analyses are presented in more detail in Appendix D.

Analytic Model and Specifications

Treatment effect size estimates and average achievement across schools were calculated using ANCOVA models. Effects of participation in the interventions were separately assessed in the same model. The final model for this cross-sectional analysis of the impact of the targeted intervention presented below was specified with fixed effects for schools. In other words, the overall impact of each targeted intervention is estimated as a treatment effect averaged across schools. However, as previously described, the treatment effect was also estimated between schools.

The dependent variable (outcome) used to estimate the impact of the targeted intervention on students' reading achievement is the Stanford Diagnostic Reading Test version 4 (SDRT-4). The outcome, reading achievement, was measured on a continuous scale (using SDRT-4 scaled scores) and normal curve equivalency scores (NCEs) were calculated to present final model results on an equal-interval scale for ease of interpretation (these scores can be averaged and have a mean of 50).

The final model includes the baseline/pretest score as a covariate (MCAS ELA scores from grade 8). Model covariates assessed for inclusion in the final model were student-level characteristics coded as dummy variables: race/ethnicity, free and reduced lunch status, special education status, ELL status, minority status, and gender. Cohort and school differences were also assessed. Refer to Appendix D for a detailed description of the variables included in the analytic model and their coding specifications; decisions regarding the handling of missing data and information regarding the decision rule for the inclusion of covariates.

⁹⁵ The multilevel model yielded an intraclass correlation of .08; that is, the amount of variance in the reading scores to be predicted between groups (i.e., schools) is 8%, while the variance to be predicted at the individual level is 92%. This intraclass correlation is consistent with similar research on school effects and the predominance in cross-sectional data of the individual characteristics (Bloom et al., 2006; Raudenbush & Bryk, 2002).

Analytic Sample

The following exhibits present descriptive information about the sample by district and treatment group. Characteristics are presented for the combined cohorts and for the ITT analytic sample, which includes all cases with post-test scores (807 of the 982) and all cases of post-test scores with pre-test scores (684 of 807). Of these cases, 679 had complete demographic information (679 of 684) for inclusion in final adjusted models. Patterns observed in the descriptive variable percentages between districts and among the treatment groups in the analytic sample remained similar to those observed for the complete ITT sample (refer to Appendix D for additional presentations of data by district and cohort).

As illustrated in the exhibits below, aggregate student characteristics differ between districts for select variables.

Exhibit 30. Student sample characteristics by district: Pre-and post-test sample

Characteristics	District		Total (average)
	Chicopee	Springfield	
Minority (%)	58	84	71
Female Gender (%)	51	61	56
Special Education Status (%)	22	19	21
English Language Learner Status (%)	2	5	4
Free and Reduced Lunch Status (%)	51	87	69
Attendance (% of total possible days)	93	90	92
MCAS Score (mean)	230.7	229.5	230.0
Sample size (<i>n</i>)	259	420	679

Note. Other includes a combination of White, Black, Asian, American Indian, Native Hawaiian, and Hispanic.

Students in both districts scored similarly on the SRI reading achievement assessment screen and the MCAS, as would be expected if the same group of targeted students were being identified. Chicopee students in this sample scored only slightly higher on average on the MCAS as

compared to Springfield. Note that the sample sizes between the districts differed (the balance is 38% Chicopee versus 62% Springfield), which may influence the significance of the differences observed for MCAS scores; however, the relative differences were still large. Differences between districts within the sample were not unexpected given the population characteristic differences (refer to Section II and district context).

Across all students included in the preliminary analysis sample and assessed at baseline, more than half were non-minority students with the majority in Springfield as compared to Chicopee (84% and 58%, respectively). In addition, Springfield had significantly higher ($p < .05$) numbers of females than Chicopee (61% versus 51%, respectively). There were significant differences among Common Core Data (CCD) collected and provided by the district including those classified as ELLs and those with free and reduced lunch status with the exception of SPED status. In this student sample, 87% in Springfield as compared to 51% in Chicopee qualify for free or reduced-price lunch, a proxy used to represent student socio-economic status. MCAS scores between districts differed ($p < .15$) and attendance rates also differed significantly between the two districts.

The following exhibit presents the data for the pre-post ITT analytic sample by treatment group.

Exhibit 31. Student sample characteristics by treatment: Pre- and post-test sample

Characteristics	Group			Total (Average)
	Control	READ 180	Xtreme Reading	
Minority (%)	71	74	78	74
Female Gender (%)	53	61	57	57
Special Education Status (%)	19	18	24	20
English Language Learner Status (%)	4	3	4	4
Free and Reduced Lunch Status (%)	74	69	76	73
Attendance (% of total possible days)	91	90	91	91
MCAS Score (mean)	230.3	229.4	230.2	230.0
Sample size (<i>n</i>)	225	231	223	679

Note. Other includes a combination of White, Black, Asian, American Indian, Native Hawaiian, and Hispanic.

Patterns for the final combined sample Years 1 through 5 in general remain the same as in the past years. No difference at the $p < .05$ level among groups was observed for any of the demographic covariates including percentages of Special Education Status (SPED) and English Language Learner Status (ELL); in fact, no difference was observed at the $p < .15$ level. In prior years, analysis results indicated that, on average, the random assignment process was generally effective in creating equivalent groups based on the variables measured and those used in stratification (SPED and ELL percentages did not differ across groups). Significant differences were not observed pretreatment (e.g., pretest MCAS scores) nor for attendance rates which could be predictive of treatment or influenced by treatment as an outcome.

Using criteria outlined by What Works Clearinghouse (WWC) for assessing the rigor of designs and analysis, baseline or pre-test scores were assessed to identify pre-treatment differences among the groups. No significant differences were observed among the groups. Pretest scores were not observed for the three groups (two treatments and one control) to be over a .05 standard deviation difference. Students' screening and baseline covariate scores (SRI and MCAS) were similar across groups, although the student SRI scores were three and four points higher in the combined cohorts (Years 1-4) for the control and Xtreme Reading groups respectively, in comparison to the READ 180 group.

In addition, the numbers of "actual" exclusions were examined to identify differential attrition between groups (i.e., these exclusions would have been noted at the time of screening and assignment review but were not available to evaluators until late fall). No differences in attrition estimates among treatment groups were greater than 20%.⁹⁶

Impacts on Students

Unadjusted means represent the true difference between groups in a random assignment study. The mean reading outcome scores are presented by treatment group in the table

⁹⁶ Refer to What Works Clearinghouse (WWC) standards.

below. However, adjusted means were calculated in the event random assignment did not yield equivalent groups due to the smaller sample sizes.⁹⁷

Exhibit 32. Mean student reading achievement scores by group (SDRT-4 Scaled Scores)

	Unadjusted Means		
	Control	READ	Xtreme Reading
<i>Number of Schools = 5</i>		180	
Reading Achievement Mean	667.91	670.83	667.56
Reading Achievement Standard Deviation	27.53	27.51	27.52
Number of Students ^b	225	231	223

^b Sample for the regression-adjusted model was dictated by the numbers with both pre- and post-tests (n = 684 of those with post-tests n = 807 of the ITT sample n = 982) with covariate data (n = 679).

As the table above illustrates, there were mean differences between the treatment group and the control group. However, these differences were not significant without covariates in the model to adjust for pretest reading levels, etc. The final and covariate adjusted models are included in Exhibit 33 below. The final model presented included only covariates significant in this complete model below the $p < .20$ level (ELL status, SPED status, and gender). School and cohort year were both included in the models, effect coded.

⁹⁷ As stated in a technical assistance provider memo: In the ideal (i.e., when random assignment works perfectly), the difference between these two means would be the unbiased estimate of program impact. However, all sites are planning to use covariates to adjust the model to help guard against bias that may have been introduced because random assignment did not work perfectly. The regression adjusted means and impact estimate will reflect these adjustments.

Exhibit 33. Impact of intervention on student reading achievement by group (SDRT-4 NCE Scores)

	Unadjusted Means			ANCOVA-adjusted Means		
	Control	Treatment		Control	Treatment	
		READ 180	Xtreme Reading		READ 180	Xtreme Reading
<i>Number of Schools = 5</i>						
NCE Mean	32.70	34.20	32.59	21.75	24.14	21.95
NCE Standard Deviation	13.38	13.37	13.38	38.21	39.18	38.20
NCE Standard Error	.89	.88	.90	2.55	2.58	2.56
Estimated Impact	--	1.5	0	--	2.39	.2
Effect Size ^a	--	.11	0	--	.06	0
P-value	--	.23	.93	--	.03	.85
Number of Students ^b	225	231	223	225	231	223

^a Effect sizes were calculated (Glasses) for unadjusted means using the control group standard deviation.
^b Sample for the regression-adjusted model was dictated by the numbers with both pre- and post-tests (n = 684 of those with post-tests n = 807 of the ITT sample n = 982) with covariate data (n = 679).

Observed and significant effects of one of the interventions as compared to the control group resulted from the final analyses. READ 180 students scored significantly higher as compared to control students (1.5 and 2.39 unadjusted and adjusted NCE scores, respectively). The significant READ 180 intervention effect was observed for the combined sample of cohorts assigned each year, over the five-year grant period, and was consistent with combined sample results from prior years (Years 3 and 4). Glasses' Δ effect size estimates were calculated (Abt communication; Rosenthal, 1994). Refer to Appendix D for more information regarding effect sizes and additional model results.

The final multilevel analytic model fit to assess the percentage of original variance in reading scores between individuals accounted for 23% of the original variance (92%). This model also accounted for 49% of the original variance remaining to be predicted between schools (8%). As anticipated, the multilevel model results mirrored those already presented.

Finally, the mean scores at post-test, though higher than at pre-test, represent less than grade-level performance. As current research indicates, when achievement gains are assessed across grade level, effect sizes decrease in the upper grades (Bloom, Hill, Rebeck Black, & Lipsey,

2007). Therefore, Striving Readers in the high schools would generally be expected to gain less than those in the lower grades simply as a result of the trajectory of student growth or development of reading skills. Although a treatment effect for one of the two interventions was observed, the relative cost of the investment to yield such an effect was beyond the scope of work for this study. However, future study should include such an assessment to help teachers, schools, and districts to further evaluate interventions considered to be effective and determine what is best for their students given severely limited resources.⁹⁸

⁹⁸ Further study of this nature should include a specification of minimum implementation levels and requirements for optimal results when considering the relative cost of any intervention.

VII. Targeted Intervention Impacts and Implementation

The goal of the targeted implementation study was to inform the interpretation of impact findings by *describing* the context in which the interventions were implemented. More specifically, implementation levels were established to characterize the context and its complexity and, as a result, to provide a gauge by which to judge any observed effects relative to the context. Therefore, the following analysis describing the relationship between classroom-level implementation and impact scores was purely exploratory and not intended to predict the impact of the interventions.⁹⁹ The true “cause” of an effect cannot be identified without an experimental design. In the case of the analysis described here, a study randomly assigning levels of implementation would have to be conducted to identify which level would be responsible, or “cause,” an observed effect.¹⁰⁰

Describing the implementation context in relationship to observed impact involved several steps. The first step was to combine classroom implementation ratings across two years in order for this information to more accurately represent the context of the combined cohort data assessed in the impact study.¹⁰¹ Overall ratings were calculated by adding ratings across years and dividing by the total number of possible items to be rated, thereby weighting the scores (refer to Appendix A for more information).¹⁰² The second step involved summarizing the implementation levels to represent both study years combined, as had been done for each individual year with the following four levels: No evidence (0–24%), Low (25–49%), Moderate (50–74%), and Adequate or High (75–100%). The third step involved examining

⁹⁹ The hypothesis that higher levels of implementation would be related to higher levels of observed impact was not empirically tested; analyses were purely illustrative. As described in the Enhanced Reading Opportunities Study, such analyses: “...are not able to establish causal links between these aspects of implementation and variation in program impacts across sites, because school characteristics and other implementation factors may confound the association between...impacts and the implementation factors included in the exploratory analysis” (Corrin, et al., 2008).

¹⁰⁰ Refer to Shadish, Cook, & Campbell (2002).

¹⁰¹ Classroom implementation was used to describe context for this purpose. Input levels were previously discussed as influences on classroom implementation context in concert with other non-intervention factors (e.g., school). As previously described, classroom is equivalent to school as only one class was constituted for each intervention within the schools rather than several as planned.

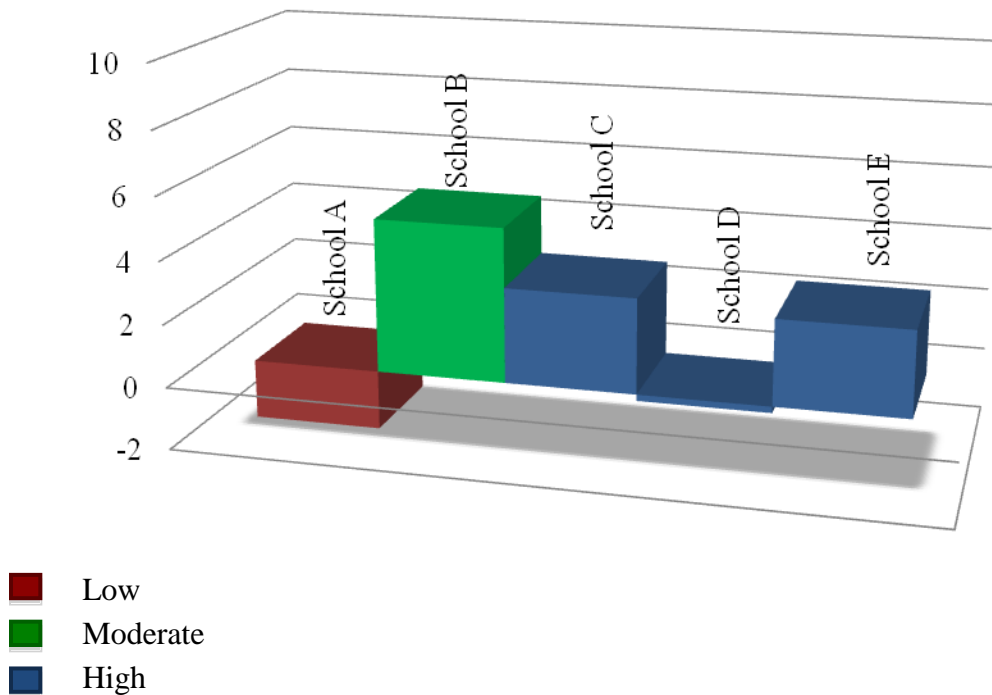
¹⁰² It is important to remember these data were collected in snapshots and by definition represent only a picture of implementation at that precise point-in-time.

the implementation and impact results together for each intervention to identify emergent patterns. This examination was also conducted across interventions to illuminate any overall patterns that may have emerged across both interventions. A discussion of this analysis is provided at the conclusion of this section.

READ 180 Classroom Implementation and Impact

The comparison of classroom implementation and impact results for READ 180 is included in Exhibit 34 below. This exhibit illustrates that in schools where classroom implementation levels were observed to be moderate and high (as coded by color), the average reading scores of READ 180 students were higher relative to students in the control group (the difference represented on the Y axis in reading achievement scores or SDRT-4 NCEs).

Exhibit 34. Impact of READ 180 by level of classroom implementation (Years 1-5)



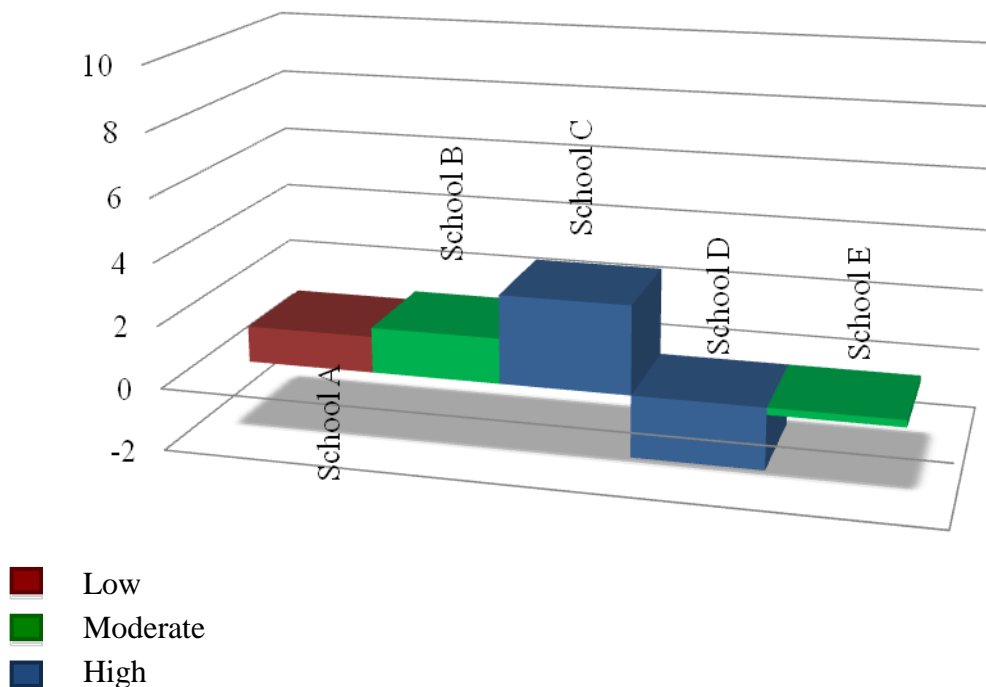
Note. Averages were calculated weighted by the total number of items across years. Implementation levels: No evidence (0–24%), Low (25–49%), Moderate (50–74%), and Adequate or High (75–100%).

READ 180 implementation levels were assessed in relationship to outcome scores for READ 180 students, and this relationship visually represented in the exhibit was significant. That is, higher levels of READ 180 implementation were associated with higher reading scores. Four of the five teachers with the highest classroom ratings had taught this intervention the longest, three for three years and one for four years. Results were more consistent over time for the majority of teachers especially those implementing at high levels over the entire study period. On average, READ 180 student scores were higher at post-test, controlling for pre-test scores and other student characteristics than control group student scores, and this difference was statistically significant.

Xtreme Reading Classroom Implementation and Impact

The comparison of classroom implementation and impact results for the Xtreme Reading intervention is included in Exhibit 35 below.

Exhibit 35. Impact of Xtreme Reading by level of classroom implementation (Years 1-5)



Note. Averages were calculated weighted by the total number of items across years. Implementation levels: No evidence (0–24%), Low (25–49%), Moderate (50–74%), and Adequate or High (75–100%).

This exhibit illustrates that in schools where classroom implementation levels were observed to be moderate and high (as coded by color), the average reading scores of Xtreme Reading students were higher relative to students in the control group in only two of four schools (the difference represented on the Y axis in reading scores or SDRT-4 NCEs).

The pattern of prior teaching was not as easy to discern for Xtreme Reading; as noted in the prior scoring section, one of the two teachers with the lowest overall ratings had been implementing since the initial grant year.

Xtreme Reading implementation levels were assessed in relationship to outcome scores for Xtreme Reading students, and this relationship visually represented in the exhibit was not significant. That is, higher levels of Xtreme Reading implementation were not associated with higher reading achievement scores. On average, the Xtreme Reading student scores were approximately the same at post-test; controlling for pre-test scores and other student characteristics than control group student score, there was not a statistically significant difference observed between the two groups.

Implementation Patterns as Predictor

Despite the many complications related to implementation, particularly in Year 1 of the study, a pattern of medium (i.e., moderate) and high (i.e., adequate) targeted implementation levels and higher overall student reading scores was observed. This pattern was more pronounced for READ 180 and was significant when assessed in relationship to reading scores.

Over time, the targeted teachers had more experience, and the control classroom teachers had higher levels of education. As a result of teacher turnover, the backgrounds as compared to control classroom teachers changed. Background and experience, in addition to overall teaching quality (not directly measured), among other unmeasured factors could influence and moderate any observed results.

Although impact estimates were established across years, implementation levels and impact results varied by year, which itself has implications and at a minimum requires caution when interpreting any of these findings. It is important to note that these cautions should be exercised for both interventions, as there were differences in implementation between years for both Xtreme Reading and READ 180, including teacher turnover in earlier years.

VIII. Evaluation of the Implementation of the Whole-School Intervention

The goals for the whole-school implementation study were the same as those for the targeted implementation study: to present a broad picture of the overall level of implementation in context and to provide a sense of the variability that may have occurred.

Whole-School Research Questions and Methods

Similar to the approach used for examining implementation of the targeted interventions, implementation research questions were developed for the SIM-CERT whole-school intervention.

1. What was the level of implementation and variability of professional development and support for teachers/administrators/literacy coaches?
2. What was the level of implementation and variability of classroom instruction?
3. What was the context of implementation (e.g., potential influences on implementation)?¹⁰³

Refer to Appendix B for exhibits including specific implementation research questions within each primary question listed above based on the program model and their intended activities, methods, objectives, and ultimate outcome goals. The implementation data collected via each method is also described in Appendix B with measures included in Appendix C. Scoring and implementation levels are described in more detail in the following section.

¹⁰³ This question has been implicit in the evaluation of implementation across years, and data have been collected, analyzed, and reported regarding the general context of implementation but is now explicitly included in this section.

Whole-School Implementation Teachers

Selection of SIM-CERT Teachers

Prior to grant implementation, the districts developed explicit criteria for selecting and prioritizing teachers for inclusion in SIM-CERT cohorts, to observe developers' SIM-CERT training requirements, and to avoid potentially confounding study results.¹⁰⁴ Participation in SIM-CERT training was to be mandatory and determined in accordance with selection criteria (i.e., content area and grade level).¹⁰⁵ Participants were to be randomly selected from the priority groups, a more equitable process and one avoiding complications in the interpretation of outcomes given all teachers were eventually required to participate in SIM-CERT training over the period of the grant as per district implementation plans.¹⁰⁶

The majority of SIM-CERT-trained teachers in the initial grant years were from the three content areas (science, math, and social studies) specified as “least likely” to confound study findings from the targeted interventions while still meeting the intervention standards. [Refer to Section B2 in Appendix B for more information about the criteria and numbers of teachers included by content area.] However, adherence to the established criteria was not always consistent. In the initial grant years, Springfield faced difficulties in implementing the professional development as planned, resulting in lower than anticipated numbers of teachers trained. Beginning in Year 2, Springfield teachers were recruited for participation in SIM-CERT training on a voluntary basis to better meet their goals for training numbers as specified in the grant. In Years 3 and 4, Springfield added optional and paid training sessions

¹⁰⁴ From the start of the grant, efforts were to be made during the selection process to limit the exposure of READ 180 and Control students to SIM-CERT trained teachers to avoid complications related to the interpretation of impacts (SIM-CERT was not business-as-usual prior to this grant). Criteria were established in consultation with evaluators and detailed in the implementation and evaluation plans to ensure model fidelity would be maintained as well as the integrity of the evaluation/study within and across districts.

¹⁰⁵ If only teachers motivated to participate were included, observed outcomes could be the result of such motivation. This selection bias is a threat to the validity of the whole-school study, implemented over time. Selecting from the pool of all required participants, or those identified in groups first, is a method for avoiding selection bias and is often understood to be a more equitable way of including all teachers because all teachers were required to be trained by the conclusion of the grant.

¹⁰⁶ In addition, mandatory district professional development was congruent with business as usual practices for a whole-school initiative. Teachers in the upper grades (beyond ninth grade) were to be given priority in the selection process based on the established criteria for training in both the first and second years as planned.

and increased recruiting efforts to further increase their numbers of trained teachers over time. Therefore, only Chicopee adhered to the requirement that SIM-CERT teachers be trained on a mandatory basis.

Characteristics of SIM-CERT Teachers: Over time

According to district documents, across the five grant years a total of 623 teachers have received some form of SIM-CERT training.¹⁰⁷ A total of 400 of those trained were from Springfield, and the remaining 223 were from Chicopee. Surveys were conducted to gather information regarding participation and prevalence of SIM-CERT knowledge and use over time. The survey was the primary source of information regarding teacher characteristics.¹⁰⁸

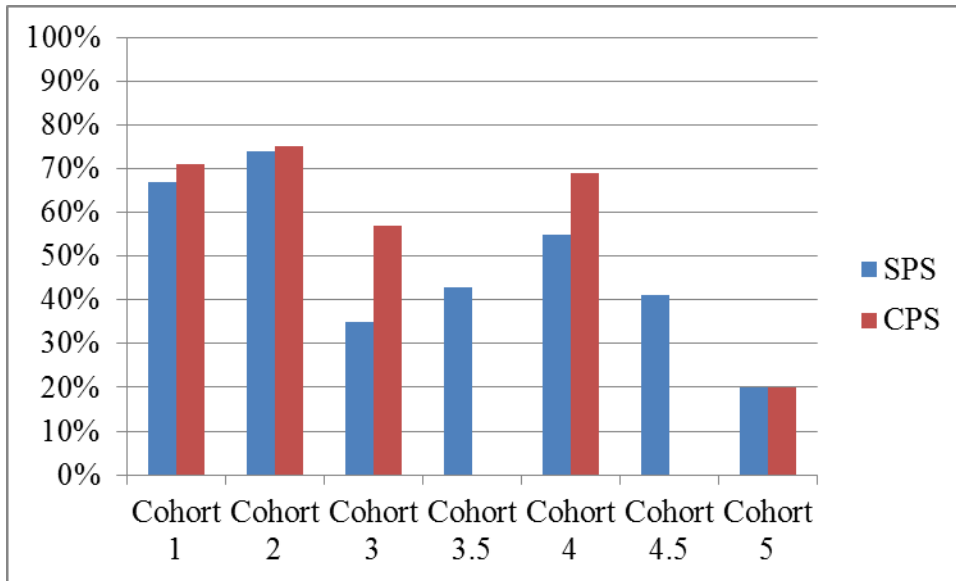
In Year 5, survey completion rates were the lowest to date at 66% of those reportedly trained by the district. In contrast, in Year 4 the highest percentage of those trained in SIM-CERT (79%) completed the survey. In Years 2 and 3, 67% and 73% of teachers reportedly trained in SIM-CERT responded, respectively.

In each cohort and in both districts, the Year 5 SIM-CERT-trained survey respondents indicated that they were certified at the professional level at varying rates. The following exhibit includes rates over time by cohort and district. The highest rates of certification were observed in the initial years and a reduction in the rates in subsequent years appears reflective of district training patterns; the lowest rates of certification were observed in the final year. [Cohorts 3.5 and 4.5 were added in Springfield and not in Chicopee for reasons described in Section IX below.]

¹⁰⁷ This number does not account for attrition and does not include literacy coaches.

¹⁰⁸ Initially, districts were to provide documentation regarding teacher characteristics but, after incomplete information was received in Year 1, this information was collected via surveys. The individual teachers who responded in any given year may differ; responses have been presented by cohort.

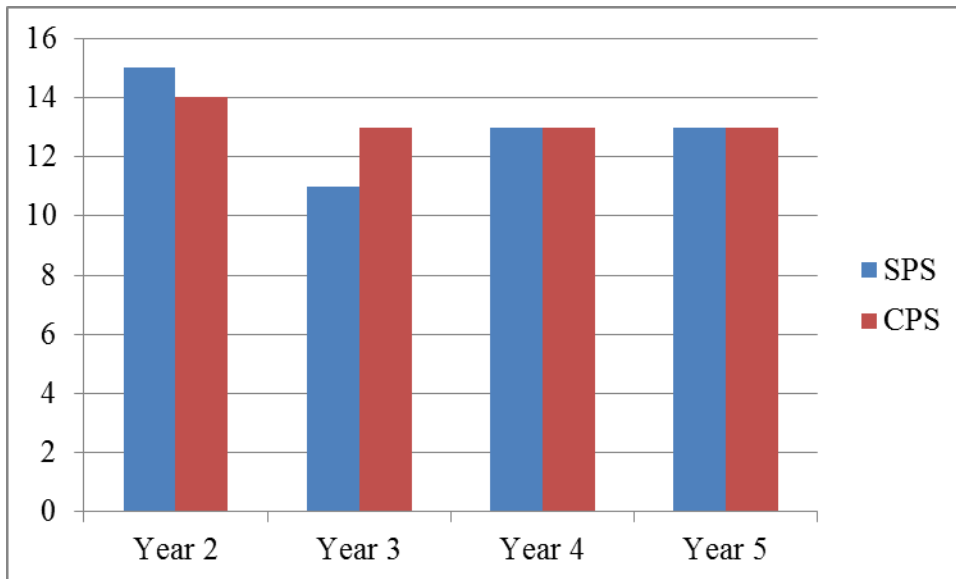
Exhibit 36. SIM-CERT teacher rates of certification at the professional level



Note. When survey and resume data conflicted, resume data were used for analysis and reporting.

In Year 5 and similar to Year 4, across cohorts, the average number of years of teaching experience reported by SIM-CERT teacher respondents was very similar. Refer to Exhibit 37 below.

Exhibit 37. SIM-CERT teacher average number of years of teaching experience



Teachers reportedly had received most of their teaching experience from their current positions. In Chicopee, over time, the average number of years of teaching experience in Chicopee within the current school was 8 years. In Springfield, this average varied across Years 2, 3, 4, and 5 (7, 6, 7, and 8 years, respectively).¹⁰⁹

Whole-School Implementation Coaches

A total of five literacy coaches, one per school, were hired as planned to support and promote implementation of SIM-CERT throughout the course of the grant. Coaches were to be certified by the developer to deliver training and support teachers in implementing SIM-CERT.

Characteristics of SIM-CERT Coaches: Over Time

In the final grant year, a substantial change was made by the districts which altered the capacity to provide the training and support for the interventions as planned. There were only three coaches rather than five available to deliver the intervention across schools; one in the Chicopee schools and two in the Springfield schools. In addition, one of the Springfield coaches was only part-time and the other was absent for a portion of the school year for medical reasons. The districts reported additional coaching support was provided by select teachers in the cadre of professional developers trained in prior years.

¹⁰⁹ Incomplete information was received in Year 1 from districts and later was obtained via surveys in subsequent years.

IX. Whole-School Intervention Implementation: Results and Implications

Whole-School Implementation Components

As with the two targeted interventions, ratings were created to establish the level of adequacy of implementation of the whole-school literacy intervention. Ratings were assigned for two components: (1) inputs consisting of the professional development and materials and (2) classroom model. Adequacy has been defined as the implementation of intervention components as specified by the developers and the districts, as depicted in the whole-school literacy intervention logic model (Exhibit 16 included in Section III of this report). Model components including the extent of training and use of SIM-CERT routines were assumed to be specified by the developers at the level necessary to promote change in content literacy. Additional contextual information related to the implementation of the professional development and classroom instruction models are also presented in this section of the report.

Professional Development

The district goal for the number of teachers to be trained in SIM-CERT was originally set at 125 per year and 25 teachers per school, but recent district documentation indicates a revised goal of 130 with a total of 650 teachers to be trained across Years 1–5, inclusive of literacy coaches. In terms of the number of teachers selected and trained, the districts did not meet the updated goal across the five years of the grant. According to district records of professional development attendance, across Years 1–5 a total of 623 teachers were selected for inclusion in SIM-CERT cohorts and received some portion of SIM-CERT training.¹¹⁰ In Year 1, recruitment numbers for both districts were below the expected amount, particularly in Springfield (48 of the targeted 80 in Springfield and 44 of the targeted 50 in Chicopee). In Year 2, recruitment numbers were

¹¹⁰ This number does not account for attrition and does not include literacy coaches or those who were also trained as targeted teachers (which occurred in the final study year). The total trained accounting for attrition (including those still in the district but no longer teaching) was 503; with 306 from Springfield and 197 from Chicopee.

closer to the target amount, but, across years, still below expected requirements. In Years 3 and 4, however, both districts exceeded the target amount of selected or recruited teachers who received any portion of SIM-CERT training. In Year 5, recruitment target levels were not reached in either district. Exhibit 38 below displays these results.

Exhibit 38. SIM-CERT training: Numbers of teachers attending any training that occurred

Cohort	Springfield	Chicopee	Total As Planned	Total Teachers Attending Any Training
Cohort 1	47 ¹¹¹	44	130	91
Cohort 2	80	46	130	126
Cohort 3	60	52	130	158
Cohort 3.5	46	-	Not originally planned	-
Cohort 4	79	55 ¹¹²	130	171
Cohort 4.5	37 ¹¹³	-	Not originally planned	-
Cohort 5	51	26	130	77
Total Years 1–5	400	223	650	623

Professional Development Context

The increase in training numbers was primarily attributable to the addition of Cohorts 3.5 and 4.5 in Springfield and was the goal of including these additional cohorts. As previously indicated, Springfield added optional and paid training sessions and increased recruiting efforts to include teachers voluntarily, thereby increasing the numbers of trained teachers to meet the expectations in Years 3 and 4. Chicopee was on model as per inclusion and recruitment plans for every year of the grant except Year 5. Refer to Appendix B1 for additional information regarding fidelity to the original selection, inclusion, and recruitment plan across the five years of the grant.

¹¹¹ Originally this number was 48 but included a teacher also trained in a targeted intervention.

¹¹² Originally this number was 54 but based on updated district records it is now 55.

¹¹³ Originally this number was 36 but based on updated district records it is now 37.

In Year 5, differences in training rates observed across schools appeared due to fewer coaches on staff (more the case in Springfield) and to fewer untrained teachers eligible for SIM-CERT training in the final year (more the case in Chicopee). In Springfield one school trained only two teachers in Year 5, while the remaining two schools trained 30 and 19 respectively. Only two coaches remained Springfield in the final year; one of the two was part-time and the other had been on longer term leave.¹¹⁴ In Chicopee one school trained only 4 teachers in Year 5 and, although only one coach remained to serve both schools, it appears the explanation for the lower training rates observed in this district was the limited pool of untrained teachers. A majority of the teachers in this district had already received training in SIM-CERT and the rates of overall turnover were lower as well so new teachers in need of training were not hired.

Professional Development Ratings

Starting in Year 3, fidelity to the professional development plan was assessed in two ways: (1) number of days in attendance at required professional development sessions and (2) amount of training content received as required. Refer to Section III for an explanation of changes made to the professional development model by the developer and/or district over time. As in Years 2, 3, and 4, professional development implementation ratings were based on district records of professional development attendance by individual teachers. Year 1 professional development scores were based on teacher self-report.

Number of Days in Attendance

According to the model, districts were to provide four six-hour or day-long training sessions within the first year of implementation and two day-long training sessions in the second year of implementation.¹¹⁵ To receive an adequate rating, teachers must have attended training either

¹¹⁴ Of the remaining two Springfield schools, one had no coach resulting from a promotion to an administrative promotion, and the other had a coach who was on medical leave for a good portion of the year. Specific numbers were not obtained regarding the number of teachers who received coaching in Year 5.

¹¹⁵ In Year 3, the developer determined that the second year of training is recommended, but not required. Developers did not specify the amount of time for a full training day; however, based on a review of agendas and other records, the evaluators determined that one training day equals 6 hours.

prior to or during the academic year (August–May), in which they were expected to apply what they had learned in the classroom with their students. Any training received after the end of the school year (i.e., in June and August) would be applicable in the following year. Separate scores were assigned for the first and second year of planned training for each SIM-CERT teacher identified by the SR district team. An adequate rating reflects full attendance at *all* required professional development sessions for each individual teacher.¹¹⁶ The percentage of adequate teacher ratings overall for Years 2–5 is presented by district and cohort in the exhibit below.

Exhibit 39. Professional development days required: Percent of teachers receiving adequate ratings by district and cohort¹¹⁷

District/ Cohort	Training for first year of implementation	Training for second year of implementation ¹¹⁸
	Four Days Required	Two Days Recommended
SPS Cohort 2	18% (n = 14/80)	74% (n = 55/74)
SPS Cohort 3	0% (n = 0/60)	39% (n = 22/57)
SPS Cohort 3.5	0% (n = 0/46)	40% (n = 18/45)
SPS Cohort 4	0% (n = 0/79)	7% (n = 5/72)
SPS Cohort 4.5	0% (n = 0/36)	88% (n = 30/34)
SPS Cohort 5	0% (n = 0/51)	N/A
All SPS	4% (n = 14/352)	46% (n = 130/282)
CPS Cohort 2	61% (n = 28/46)	76% (n = 34/45)
CPS Cohort 3	77% (n = 40/52)	84% (n = 41/49)
CPS Cohort 4	74% (n = 40/54)	74% (n = 37/50)
CPS Cohort 5	62% (n = 16/26)	N/A
All CPS	70% (n = 124/178)	78% (n = 112/144)
Total	26% (n = 138/530)	57% (n = 242/426)

¹¹⁶ Those who did not achieve an adequate rating either did not attend or only attended part of the training sessions. Refer to Exhibit 38 for partial training rates.

¹¹⁷ Attendance is reported according to updated model specifications outlined prior. For information regarding teacher attrition, refer to Appendix B.

¹¹⁸ Note that differences between denominators in the first and second columns were the result of attrition.

The ratings presented above illustrate extensive district variation in the implementation of the SIM-CERT training component of the model. Across Years 2–5 of the grant, an average of 70% of Chicopee teachers received adequate ratings for attending all required training sessions during their first year in the SIM-CERT program. In Springfield, an average of 4% of the teachers participating in these training sessions received adequate ratings across Years 2–5. That is, 4% of all those trained reached the threshold of training to receive an adequate rating. In Year 2 in Springfield, 18% of the teachers attended the required number of training days. In Years 3, 4, and 5, none of the Springfield teachers received adequate ratings, indicating that they had not participated in the required four days of training within their first year of inclusion in the program.

Criteria for assessing implementation in Year 1 were not provided by the developers as plans for implementation were still being formulated in this year. Therefore a separate framework was used to evaluate fidelity to the professional development model, which was aligned to original developer specifications. Although district variation was still apparent in the first year of the grant, the percentage of teachers receiving adequate ratings for initial training in Springfield was much higher (87%) as compared to later years; percentages for ongoing training were lower (1%).¹¹⁹ The percentage of teachers receiving adequate ratings was also higher in Chicopee (98%) as compared to later years: percentages for ongoing training were only slightly lower (71%).¹²⁰

District variation in the implementation of the professional development model was apparent for second-year training rates. In Chicopee, the majority of teachers (78%) received adequate ratings for attending the required two days of training during their second year of inclusion in the

¹¹⁹ Although the majority of teachers in Springfield received the first two days of training prior to the school year, only one teacher (1%) received the requisite remaining two days of training in the first year because these days were delivered post school year rather than as in-service days as originally planned. This one teacher was originally part of Cohort 1 in Chicopee and received the first year of training in that district prior to transferring to Springfield in Year 2.

¹²⁰ In Year 1, initial training was defined as two full days (or the equivalent) of training prior to the first year of classroom implementation. Ongoing training was defined as two full days (or the equivalent) of training before the end of the first year of implementation.

program across program Years 2-5. In Springfield, across Years 2-5, less than half of the teachers (46%) received the required two-day follow-up training during the second year of program implementation.

Professional Development Training Ratings Context

In Springfield, the timing and structure of the professional development schedule accounted for the low percentage of adequate ratings for implementation of the professional development model in that district. At the start-up of the grant the professional development model had to be modified to accommodate issues involving buy-in, communication, in-service scheduling, contract concerns, etc.¹²¹ In Years 1 and 2, the in-service training was eliminated, preventing the professional development model from being implemented with fidelity to the original plans as proposed. Because in-service professional development days were not available, teachers received only two days of training rather than four during the first year of classroom implementation; teachers received the additional two days of training in the second year of implementation. In other words, the professional development delivery schedule in Springfield did not offer the required training days within the initial year as planned and therefore teachers could not receive adequate ratings for attendance at training as planned.

In Year 2, 18% of Springfield teachers were able to receive adequate ratings due to the addition of a one-day mid-year training session. This training session was not offered in Years 1, 3, 4, and 5 but additional cohorts were included in subsequent years to further increase training numbers. As mentioned earlier, the district strategized to increase the numbers of teachers trained in SIM-CERT through the creation of Cohorts 3.5 and 4.5. These cohorts began training in the second semester of the first year of implementation (e.g., Cohort 3.5 began training in January of Year 3). Subsequently, the district succeeded in meeting, and exceeding, target numbers for teacher inclusion in the initial SIM-CERT training. However, the professional development schedule and structure for these additional cohorts consisted of less than the total

¹²¹ According to district documents, interviews with the Striving Readers district team, and as reported by other administrative and teaching staff.

required days of ongoing training. Therefore, as was the case with the other cohorts in Springfield, the professional development schedule for Cohorts 3.5 and 4.5 did not meet the criteria for fidelity to the professional development model, resulting in 0% ratings of adequacy across the district.

In Chicopee, training occurred as planned; that is the scope and sequence of training occurred as specified by the model. In contrast with Springfield, Chicopee was able to use already scheduled in-service days as planned to provide SIM-CERT training during the school year. However, there was a reduction in the percentage of Chicopee teachers receiving adequate training ratings in Year 5 relative to the previous two years. This finding may be due to the absence of a coach in one of the schools and perhaps due to already high rates of trained subject matter teachers (given the smaller teaching staff in Chicopee, few content teachers remained to be trained).

Exhibit 40 displays the professional development model, as planned, and the professional development delivery schedule as actually implemented in Springfield.

Exhibit 40. Springfield SIM-CERT training: Delivery of professional development

	2006–07 (Year 1)	2007–08 (Year 2)	2008–09 (Year 3)	2009–10 (Year 4)	2010–11 (Year 5)	Total
Cohort 1	2/4	2/2				4 of 6
Cohort 2		4/4	2/2			6 of 6
Cohort 3			3/4	2/2		5 of 6
Cohort 3.5			3/4	2/2		5 of 6
Cohort 4				3/4		3 of 4
Cohort 4.5				1+ /4	2/2	3+ of 4
Cohort 5					1+ /4 ¹²²	1+ of 4

¹²² The 1+ days of training time applies to two schools. District documentation indicates that one school received training for 5.5 hours. The developer’s training agenda and the evaluator’s observation record however, both indicate the training at this one school took place over 2.5 hours inclusive of two breaks.

In Springfield, Cohort 1 received only four of the six planned days of training over two years but received an additional day after the two-year period (five of six in total). In subsequent years, the six days were completed. Cohort 4.5 received a total of eight hours of training instead of the requisite 24 hours of training in their first year of implementation (provided either via two hours after-school on four weekdays or four hours on two Saturdays, January or February). Cohort 5 received approximately 10 hours of training over two days. The 1+ day(s) of training time included in Exhibit 40 was observed for two of the Springfield schools. District documentation indicated that one school received training for 5.5 hours, almost meeting the threshold of 6 hours for a full day. However, the developer’s training agenda and the evaluator’s observation record both indicate the training at this school took place over 2.5 hours inclusive of two breaks.

Exhibits 41 below displays the professional development model, as planned, and the professional development delivery schedule as actually implemented in Chicopee.

Exhibit 41. Chicopee SIM-CERT training: Delivery of professional development

	2006–07 (Year 1)	2007–08 (Year 2)	2008–09 (Year 3)	2009–10 (Year 4)	2010–11 (Year 5)	Total
Cohort 1	4/4	2/2				6 of 6
Cohort 2		4/4	2/2			6 of 6
Cohort 3			4/4	2/2		6 of 6
Cohort 4				4/4	2/2	6 of 6
Cohort 5					4/4	4 of 4

The majority of teachers across districts received the majority of training necessary for implementation of the classroom model according to the developer (with the exception of Cohort 3.5 in Springfield).

Receipt of Training in Specific SIM-CERT Routines

In response to a low number of adequate ratings for teachers attending the required number of professional development days in Springfield, the district worked in collaboration with the developer and evaluator in to create an alternative framework for assessing fidelity to the professional development model. This alternative framework established in Year 3 evaluates the extent to which individual teachers across districts received training in the required SIM-CERT topics or content. Scoring related to the receipt of SIM-CERT content presents a different view of teacher professional development than that obtained by examining the number of training days completed, which evaluators are required to report. Although a teacher may not have attended all training days, as defined by the original model, they may have been trained in all of the required content or SIM-CERT routines.

In Year 3, the developer had confirmed that teachers would have the knowledge or inputs necessary to achieve fidelity to the classroom model if they received training in the required topics, regardless of how many days it took to cover the material. Specifications regarding what content was required were not available prior to Year 3. Particularly during the initial years of the grant, developer specifications regarding the required content to be delivered in training sessions remained intentionally vague in order to allow district tailoring.¹²³ The exhibit below depicts required and recommended training content.¹²⁴

¹²³ The SR district team reported that developers stressed the importance of meeting the needs of the individual schools and districts, which has led to fluctuations in the model as planned. Developers report that they modified the program based on their continuous-development philosophy but also tailored the program to district needs.

¹²⁴ This information was provided during a developer, district, and evaluator call in July of 2009. The specifications for training provided following the first year were specified previously but were also reportedly individually determined based on teacher needs and requests.

Exhibit 42. Required and recommended content for SIM-CERT trainings

Year 1 (Required)	Year 2 (Recommended)
Unit Organizer	Course Organizer
Framing	Concept Comparison
LINCing	Integrated Units ¹²⁵
Concept Mastery	

Only required, not recommended, fidelity components were assessed as part of the implementation study. Furthermore, only Cohorts 3, 4, and 5, inclusive of Cohorts 3.5 and 4.5, were given ratings for the receipt of required content since this alternative framework for assessing fidelity to the professional development model was not confirmed by the developer until Year 3. Exhibit 43 displays the percentage of teachers who received adequate ratings for training in required SIM-CERT routines (i.e., content) for their first year.

¹²⁵The training in Integrated Units covers ways to integrate and connect two or more SIM-CERT routines for classroom instruction.

Exhibit 43. Percentage of teachers who received adequate levels of training in the required routines for the first year of implementation

	Receipt of all four core required routines (Unit Organizer, Framing, LINCing, Concept Mastery)
SPS Cohort 3 (n = 60)	93% (n = 56)
SPS Cohort 3.5 (n = 46)	54% (n = 25)
SPS Cohort 4 (n = 79)	86% (n = 68)
SPS Cohort 4.5 (n = 36)	89% (n = 32)
SPS Cohort 5 (n = 51)	41% (n = 21)
All SPS (n = 272)	74% (n = 202)
CPS Cohort 3 (n = 52) ¹²⁶	87% (n = 45)
CPS Cohort 4 (n = 54)	91% (n = 49)
CPS Cohort 5 (n = 26)	73% (n = 19)
All CPS (n = 132)	86% (n = 113)
Total (n = 404)	78% (n = 315)

Rates of training in required SIM-CERT routines for first-year SIM-CERT teachers were high overall. The majority of SIM-CERT teachers trained in Years 3 and 4 received the required training in the four core routines during their first year of implementation: Unit Organizer, Frame, LINCing, and Concept Mastery. Minimal district variation was observed in the percentage of teachers who received adequate ratings for training in core content. The one exception to this pattern is Cohort 3.5 in Springfield, where approximately half of the teachers received training in all four core routines, and half did not. Lower scores for this group of teachers in terms of the number of days training, as presented previously, and the receipt of content may be attributed to the difficulties unique to the initial implementation of this mid-year cohort strategy to increase target numbers.

¹²⁶ Two teachers did not attend professional development sessions, but received training in required content (four core routines) from literacy coaches. These two teachers were recorded as receiving full SIM-CERT content.

Professional Development Ratings Context

In Springfield, 100% of Cohort 5 teachers received training in the Unit Organizer and Framing routines, and all teachers except those from one school received training in the LINCing and Concept Mastery routines as well. This school was responsible for the lower rates of training in content observed in Cohort 5 as compared to prior cohorts; teachers received only two of the four core SIM-CERT routines as outlined in Exhibit 42. In Chicopee, the lower rates of teachers receiving training in all four routines may be due in part to a decline in overall enthusiasm for an initiative whose funding stream was ending¹²⁷ and in part due to the reduction in coach support,¹²⁸ which in the past had enabled those who missed formal training sessions to make up these sessions with coaches.

Taken together, the scores for number of training days attended and number of routines learned indicated that the majority of teachers across districts, and over time, received the minimum training necessary for implementation of the classroom model according to the developer, with the exception of Cohort 5 in Springfield. This was the first year since content scoring was implemented that a majority of Springfield teachers (41%) did not receive an adequate rating for having received training in all developer-required routines. Thus, in Year 5, Springfield teachers did not receive adequate ratings for attendance at the required number of professional development days nor did they receive training in the developer required routines for implementing the SIM-CERT intervention. In Chicopee, ratings within both frameworks (training hours and routines) were relatively high. However, the percentage of Chicopee teachers receiving an adequate rating for content was the lowest it had been in the three years since content scoring was assessed.

Similar to previous years, a majority of teachers in Chicopee received adequate ratings for the number of training days and adequate ratings for the receipt of training in all required SIM-CERT routines. A majority of teachers in Springfield did not receive adequate ratings for the

¹²⁷ For example, in Chicopee focus groups teachers indicated that in Year 5 previously clear expectations around SIM-CERT had become “loose” with “not much focus on collecting” devices.

¹²⁸ One coach in Chicopee was promoted to an administrative position.

number of training days, but did receive adequate ratings for the receipt of training in all required SIM-CERT routines.¹²⁹

In Springfield, the adequate ratings for training days were not achieved in general despite a reduction in the number of training days required, as set by developers and the district over time. More information was covered in a condensed amount of time, partially in response to Springfield's challenges in providing training given barriers related to initial start-up issues and a professional development delivery schedule that did not fit the original plans for in-service training as proposed. Specifically, developers confirmed that the following training sessions were equivalent in terms of content covered: June 2008 (3 days) = August 2008 (2 days) = January/March 2009 (1.5 days).

In the later years of the grant, evaluator observations in Springfield revealed that the developer, and later the school-based trainers, reduced or eliminated collaborative work time for teachers to include SIM-CERT routines to their lesson plans, and instead provided training in all required content in a shortened amount of time. Originally, training sessions presented one SIM-CERT routine and give teachers time to apply that routine to their course content with colleagues from their department. In Year 5 training time was cut even shorter, which was not sanctioned by developers. The final reduction in training time rendered it impossible in most cases for teachers to receive training in all of the required routines. In Chicopee, the professional development plan including the number of days, the content taught, and content delivery remained consistent across Years 1–5 and was implemented as originally proposed.

Classroom Implementation Ratings

Classroom-level implementation was the second component of the overall implementation ratings of SIM-CERT. The following minimum classroom model specifications¹³⁰ were used for

¹²⁹ In a few instances, teachers attended the majority of the training day but were released early by SIM-CERT trainers or received instruction in the missed content from a literacy coach at a later date.

¹³⁰ The first two specifications were mandatory, and the third specification was optional. Classroom model specifications were not provided to assign ratings in Year 1; therefore, only ratings across Years 2–5 are reported.

scoring in Years 2–5. Teachers trained in SIM-CERT were required to: (1) utilize at least one Unit Organizer in one course during the academic year; (2) implement at least one additional routine during the academic year (e.g., LINCing, Framing, Concept Mastery, Concept Comparison, Course Organizer); and (3) implement other routines as appropriate. Refer to the SIM-CERT logic model presented in Exhibit 16 for additional information regarding requirements.¹³¹

Ratings were assigned based on survey responses (i.e., self-report data) regarding the use of SIM-CERT routines during Years 2, 3, 4, and 5.¹³² Respondents who met the minimum developer-defined requirements as described above received a rating of adequate, and those who did not received a rating of inadequate. Respondents who received a rating of adequate reported meeting minimum requirements: that is, use of the Unit Organizer routine plus one additional routine.¹³³ Respondents who did not receive a rating of adequate for usage either used only the Unit Organizer routine or indicated that they had not used the Unit Organizer routine during the current school year.

A similar rating framework to that used for minimum usage requirements was also applied to determine which respondents exceeded developer-defined classroom model requirements. Thus, teachers who indicated they had used the Unit Organizer routine plus two or more additional routines received a rating of adequate.¹³⁴ Separate ratings were assigned to individual teachers for classroom-level implementation for Years 2, 3, 4, and 5 based on survey responses for each respective year of implementation.

¹³¹ According to district communications, the expectations or criteria provided by the developer for the classroom model has not been comprehensive (i.e., much of classroom implementation was left to individual teacher discretion). Thus, the criteria used for scoring the implementation of the classroom model include only the minimum developer-defined requirements.

¹³² Scores for classroom usage of SIM-CERT routines were assigned according to teacher self-reports regarding the implementation of each routine at some point during the 2010–11 school year. Scores did not take into consideration the frequency or the quality with which teachers implemented each routine in the classroom (i.e., whether teachers used a Unit Organizer for every unit taught or do so appropriately) due to minimal information received from the developers on classroom model specifications during all four years of the intervention.

¹³³ Ratings were not assigned to respondents with missing information regarding the Unit Organizer.

¹³⁴ Percentages for exceeding minimum usage requirements are derived from the total number of teachers indicating they have used the Unit Organizer plus two or more additional routines from the total number of teachers who reported meeting minimum classroom usage requirements. Percentages are NOT based on the total number of SIM-CERT trained teachers; these data are self-reported.

Ratings for the implementation of the classroom model across Years 2–5 are presented in Exhibit 44.

Exhibit 44. Classroom model ratings by district across Years 2, 3, 4, and 5

Year	District	Met Minimum Usage Requirements	Exceeded Minimum Usage Requirements
		Unit Organizer + 1 additional routine	Unit Organizer + 2 or more additional routines
Year 2 2007–08	CPS (n = 64)	89% (n = 57)	86% (n = 49)
	SPS (n = 77)	71% (n = 55)	65% (n = 36)
	Total (n = 141)	79% (n = 112)	76% (n = 85)
Year 3 2008-09	CPS (n = 94)	96% (n = 90)	80% (n = 72)
	SPS (n = 132)	71% (n = 94)	68% (n = 64)
	Total (n = 226)	81% (n = 184)	74% (n = 136)
Year 4 2009–10	CPS (n = 140)	86% (n = 120)	80% (n = 96)
	SPS (n = 218)	65% (n = 142)	64% (n = 91)
	Total (n = 358)	73% (n = 262)	71% (n = 187)
Year 5 2010–11	CPS (n = 124)	73% (n = 91)	78% (n = 71)
	SPS (n = 172)	53% (n = 92)	63% (n = 58)
	Total (n = 296)	62% (n = 183)	70% (n = 129)

Exhibit 45 presents classroom usage scores for Year 5, disaggregated by district and cohort.

Exhibit 45. Year 5 classroom model ratings by district and cohort

Cohort	District	Met Minimum Usage Requirements	Exceeded Minimum Usage Requirements
		Unit Organizer + 1 additional routine	Unit Organizer + 2 or more additional routines
1	CPS (n = 24)	88% (n = 21)	81% (n = 17)
	SPS (n = 10)	60% (n = 6)	50% (n = 3)
2	CPS (n = 24)	79% (n = 19)	79% (n = 15)
	SPS (n = 31)	45% (n = 14)	57% (n = 8)
3	CPS (n = 35)	71% (n = 25)	72% (n = 18)
	SPS (n = 32)	56% (n = 18)	78% (n = 14)
3.5	SPS (n = 27)	63% (n = 17)	59% (n = 10)
4	CPS (n = 31)	65% (n = 20)	90% (n = 18)
	SPS (n = 38)	45% (n = 17)	59% (n = 10)
4.5	SPS (n = 21)	62% (n = 13)	69% (n = 9)
5	CPS (n = 10)	60% (n = 6)	50% (n = 3)
	SPS (n = 13)	54% (n = 7)	57% (n = 4)
All	CPS (n = 124)	73% (n = 91)	64% (n = 58)
All	SPS (n = 172)	53% (n = 92)	77% (n = 71)
Total	(n = 296)	62% (n = 183)	70% (n = 129)

Classroom Implementation Rating Context

Across Years 2, 3, and 4 of the grant, roughly three-fourths or more of SIM-CERT-trained teachers met minimum requirements for implementation of SIM-CERT in the classroom, whereas in Year 5 this overall rating decreased to 62%. Despite this reduction in Year 5 percentages as compared to prior years, the majority of teachers across years who responded to the survey reported using the Unit Organizer once or more during each school year along with another SIM-CERT routine of their choice. Across districts, 70% of the group of teachers who received adequate scores for classroom model fidelity exceeded minimum requirements. Across Years 2–5, there was a minimal but steady decline in the percentage of teachers who reported exceeding classroom model requirements, the reasons for which remain unclear but may have

been related to administrative changes in program oversight and accountability in the final grant years.

Across all years and cohorts, evidence of district variation was observed. As shown in Exhibits 44 and 45 above, a greater percentage of Chicopee teachers met and exceeded classroom model specifications than Springfield teachers in Years 2, 3, 4, and 5. District variation is most apparent in the percentage of teachers meeting minimum requirements, with 73% of responding teachers in Chicopee meeting minimum requirements and 53% in Springfield. However, of those implementing the minimum requirements, a higher percentage of Springfield respondents reported exceeding minimum requirements than Chicopee respondents (77% and 64% respectively). Unlike in Year 4 where school-level variation was observed only in Springfield for teachers who met minimum implementation requirements, in Year 5 both districts exhibited school-level variation. In Year 5, both districts also continued the general pattern of decreasing percentages of teachers meeting minimum requirements that was first observed from Year 3 to Year 4.

Frequency of Classroom Use-Implementation

Literacy coach and administrator interviews, district- and developer-provided documentation, and teacher self-report data (survey and focus group) provided more nuanced and detailed information regarding how often and in which situations SIM-CERT routines were implemented in the classroom. Exhibits 46, 47, and 48 below show the percentage of survey respondents reporting classroom use of individual SIM-CERT routines across Years 3, 4, and 5.¹³⁵

¹³⁵ In the Year 2 survey, teachers could select “yes”, “no”, or “don’t know” in response to whether they have used each of the six SIM-CERT routines in the classroom. Year 2 results could not be combined with Year 3, Year 4, and Year 5 survey results due to the existence of the “don’t know” response option.

Exhibit 46. Classroom usage of SIM-CERT routines: Year 3

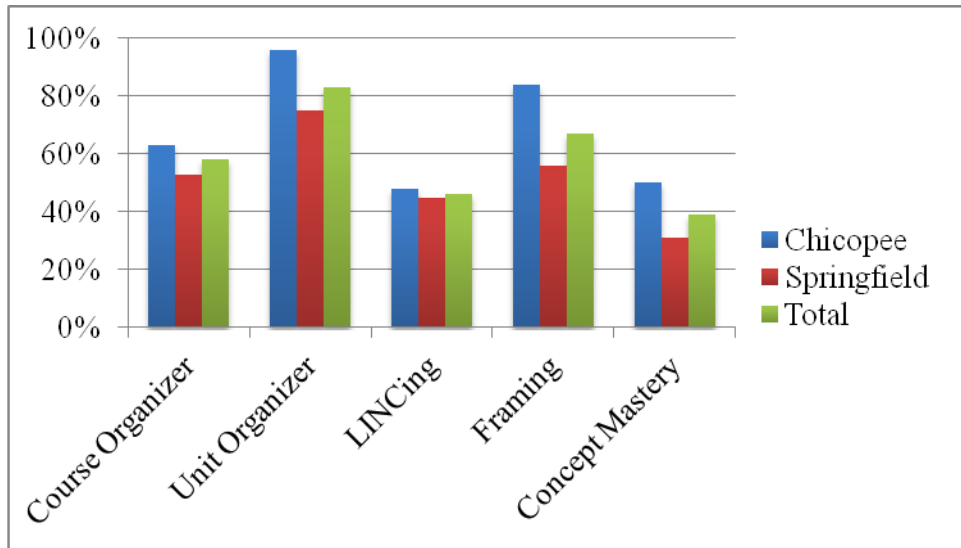


Exhibit 47. Classroom usage of SIM-CERT routines: Year 4

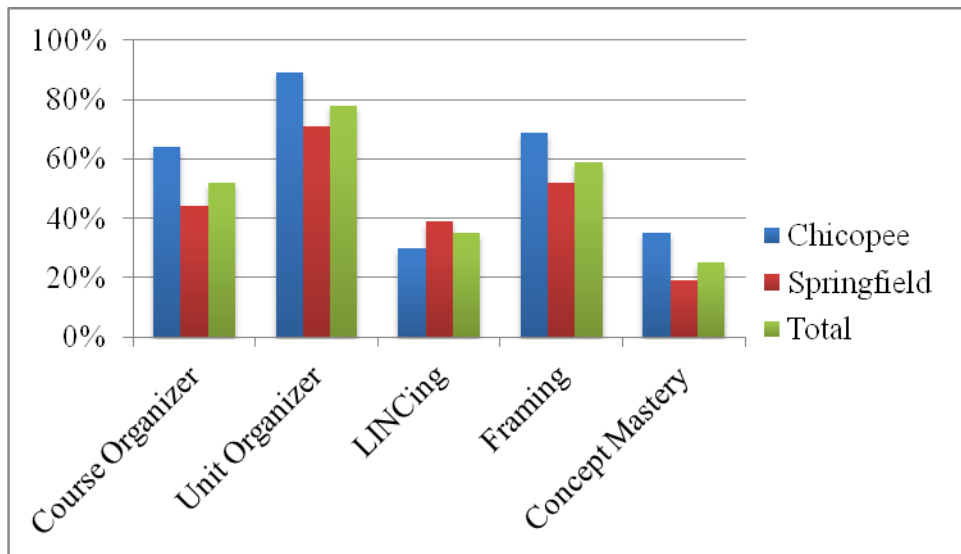
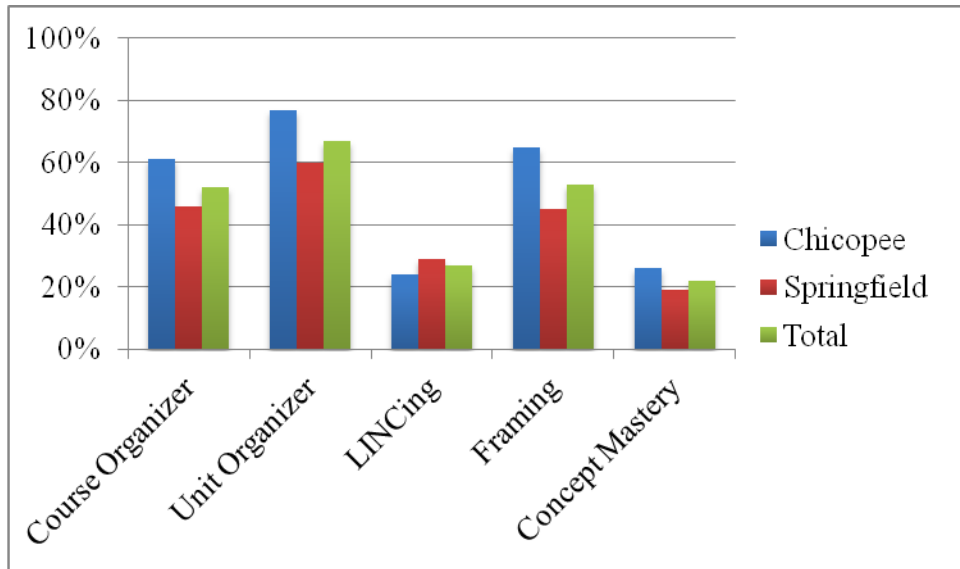


Exhibit 48. Classroom usage of SIM-CERT routines: Year 5



Of the six routines, the Unit Organizer (the foundational routine) was reported to be used most often by teachers in the classroom, according to Years 3–5 survey responses. Eighty-three percent of teachers in Year 3, 78% of teachers in Year 4, and 67% of teachers in Year 5 reported using the Unit Organizer one or more times during the 2010–11 school year. Framing and, to a lesser extent, Course Organizer, were reported to be used by over half of teacher respondents in Years 3, 4, and 5. The other three routines, one of which was covered in the second year of training, were reported to be used by less than half of the teachers during the 2008–09, 2009–10 and 2010–11 school years. In both districts, reported classroom usage of nearly all routines declined from Year 3 to Year 5. Some district variation in teacher-reported use of specific SIM-CERT routines in the classroom emerged from survey findings, with Springfield reporting a lower percentage of use for all routines relative to Chicopee with the exception of LINCing.

Focus Group data across Years 2–5 mirror the survey results in terms of which routines teachers tended to implement more than others. Of the SIM-CERT routines presented in professional development workshops, teachers reported implementing the Unit Organizer and Framing routines most often. Teachers provided mixed opinions regarding the applicability of LINCing, varying by subject area and type of audience (e.g., better for ELL and SPED students), and

offered only minimal comments about Concept Mastery. In Year 3, several teachers across districts had positive feedback about the Concept Comparison routine—a change from Year 2 where Concept Comparison had not been mentioned. Similar to Year 4, in Year 5 teachers across districts reported that they did not know enough about the Concept Mastery or Concept Comparison routines to be able to implement them in the classroom (see Appendix B for additional focus group findings regarding how and why teachers used particular SIM-CERT routines in the classroom).

On the survey, teachers also reported the frequency with which they implemented specific SIM-CERT routines in the classroom. In Year 5, teachers across both districts reported implementing the Unit Organizer most frequently of the six routines, consistent with prior years. The Framing routine was reported as the second most frequently implemented, followed by the Course Organizer. However, over half of the teachers who reported using the Course Organizer indicated that they had rarely (once or twice during the academic year) implemented this routine. A similar pattern of infrequent use was also noted on the Year 5 survey for the Concept Mastery and Concept Comparison routines. Even though a large percentage of respondents did not report using the LINCing routine (27% overall), approximately 60% of those who reported using it planned more than two units using LINCing. The following exhibit displays the reported number of units teachers planned using the Unit Organizer during the 2010–11 school year, according to survey results.

Exhibit 49. Frequency of classroom implementation: Unit organizer

	CPS	SPS	Total
	(n = 95)	(n = 104)	(n = 199)
1–2 units	33 (35%)	40 (38%)	73 (37%)
3–4 units	28 (29%)	34 (33%)	62 (31%)
5 or more units	34 (36%)	30 (29%)	64 (32%)

Note. Percentages were based on the total number of teachers who reported that they have used the Unit Organizer routine (i.e., valid percentage).

In Year 5, there was little district variation in the reported frequency with which teachers chose to implement the Unit Organizer. This was a marked difference from previous years when a larger percentage of Chicopee teachers reported using the Unit Organizer routine to plan a greater number of units than Springfield teachers. There were, however, intra-district differences in use of the Unit Organizer routine. In Springfield, the number of units planned using the Unit Organizer was very similar across schools with the percentages of respondents within each school distributed fairly evenly across the categories of units planned.

District variation was still observed in the reported frequency of implementation of the other SIM-CERT routines. Similar to Years 3 and 4, in Year 5 a higher percentage of teachers from Springfield reported more frequent use of LINCing and Framing than their Chicopee counterparts.¹³⁶ As in Years 3 and 4, of those teachers who reported using the Framing routine, Springfield teachers implemented this routine in the classroom more frequently than Chicopee teachers. This same trend was observed for all routines with the exception of the Unit Organizer. This pattern suggests that while a greater percentage of Chicopee than Springfield teachers overall reported having tried or used each routine at least once (with the exception of LINCing), a greater percentage of Springfield teachers tended to use the routines more frequently to plan their lessons than Chicopee teachers (again, with the exception of the Unit Organizer).

Whole-School Intervention Implications: What Ratings Do Not Illuminate

The whole-school implementation study presents a broad picture of the level of SIM-CERT implementation but also provides contextual information to facilitate the interpretation of these implementation findings relative to overall study results. For SIM-CERT implementation, the district (inclusive of schools, personnel, resources, and students), the developer, and their interactions comprise the context.

¹³⁶ Sixty-three percent implemented the LINCing routine to plan three or more units in Springfield compared with 53% in Chicopee and 80% implemented the Framing routine to plan three or more units in Springfield compared with 58% in Chicopee.

Over time, contextual factors have consistently affected implementation plans and fidelity, both in the classroom and in the provision of professional development, across all four years of SIM-CERT implementation. These factors have operated interdependently to influence the way in which the whole-school intervention has been implemented in each district, within schools, and over time. Three key factors have shaped the context in which the intervention took place over the past five years: (1) intervention and implementation specifications, from both the developer and from district staff; (2) professional development scheduling and participant recruitment efforts; and (3) support and accountability for program implementation related to literacy coaches and school administrators. Finally, general teacher satisfaction with the professional development and support they have received influenced these key factors in context.

Intervention and Implementation Specifications

Developer requirements for implementation. In the first three years, district staff including teachers reported a lack of clarity as well as ongoing revisions regarding expectations for the delivery of professional development and the implementation of the classroom model. Beginning in the initial year of the study and continuing throughout the grant period, the developer indicated that all plans for implementation were to be determined in collaboration with the district. The model, as per the developer, was flexible to allow administrators and teachers to tailor plans to align with the unique contexts of the districts and schools.

Expectations regarding the content to be covered in professional development sessions were not clearly defined until the third year of implementation; efforts were made by the districts to gain clarity beginning in the first year. Minimum requirements for teacher attendance (in terms of number of days) were adjusted and refined repeatedly over time. Expectations for implementation in the classroom also shifted over time and were subject to teacher discretion. At the classroom-teacher level, providing a wide range of implementation options was intended to allow the teacher the choice of selecting which components of the intervention would best fit his or her subject area, the material or content being covered in each lesson, and the

characteristics of the students in the class. Self-reported data collected throughout Years 1–5 from multiple stakeholder groups, including SIM-CERT-trained teachers,¹³⁷ indicated widespread confusion about the requirements for classroom implementation as well as uncertainty about how individual teacher’s implementation should be measured or monitored. In most cases, coaches reportedly developed the implementation specifications over time for the classroom model.¹³⁸ This lack of definition for what constituted classroom fidelity led to district variation in implementation requirements, as well as variation across years.

District requirements for implementation. Over time, multiple tools for monitoring classroom fidelity were developed and provided to the districts and schools, but a common core of expectations was not implemented. In Springfield, the effects of the ambiguity in the intervention plans and expectations were even more pronounced as this district was less successful in developing a common framework for implementation than Chicopee.

Initially, the districts worked to develop monitoring tools to share in the absence of any developer tools. In Springfield, coaches either used the evolving SIM-CERT checklists of expectations or worked on their own to develop expectations, but there were differences in approach across schools. In Year 3, SIM-CERT developed a monitoring tool for measuring implementation levels but, according to staff interviews, this tool was not seen as practical and was not used at the school level. In Chicopee, school and district staff worked collaboratively to develop a common framework for implementation (both in the areas of professional development and the classroom model), which helped to provide intervention consistency across schools and over time in this district. This set of expectations was separate from those developed by SIM-CERT (but was later approved). Literacy coaches distributed documentation of their expectations to teachers and administrators in Year 2. However, these adjustments did not completely compensate for ongoing developer revisions and modifications to minimum requirements for training and classroom implementation. According to coach interviews, there was little monitoring of implementation in Year 5 in comparison to prior years. As a result,

¹³⁷ Note that in Year 5 no focus group was conducted with teachers from School A.

¹³⁸ No additional clarification was provided to districts by the developer in Year 5 regarding implementation requirements.

coaches were able only to estimate how well teachers were implementing, and indicated they believed the “majority” of teachers were implementing “frequently/occasionally.” Coaches added that a lack of communication with school leadership also hindered coaches’ ability to keep track of implementation.

Professional Development Scheduling and Recruitment

Professional development scheduling. A review of district documents and professional development records, along with self-reported data from teachers, literacy coaches, and administrators, consistently shows that Springfield did not provide the professional development structure necessary to implement the original model specifications as per the initial logic model; that is, the proposed in-service training did not occur as planned. In fact, throughout the years of the grant and especially in Year 4, the developer approved the restructuring of SIM-CERT training workshops to cover more material in less time. Teachers in Springfield did not have the option of attending ongoing workshops, as planned, to provide support and reinforcement for using the routines in the classroom during the academic year. Rather than participating in the four full days of training in the first year of implementation as Chicopee teachers did, Springfield teachers were only able to attend three or fewer days of training to prepare them for SIM-CERT classroom implementation. Over time, the difference in the availability of training and professional development between districts may have contributed to generally lower rates of classroom usage in Springfield as compared to Chicopee. Teachers may not have had enough preparation or practice to incorporate what they had learned in their classroom.

District variation in the amount of time allotted for collaborative work time during training for teachers to create devices with their peers may also be a factor in the variation in classroom usage rates between districts and in the decline in classroom usage and satisfaction with professional development that began in Year 4 and continued in Year 5. In Chicopee, where classroom usage rates remained relatively high (between 73% and 96% of surveyed teachers reporting meeting minimum requirements), the time designated for professional development workshops remained the same across years (a total of four days or 24 hours of training in the first year of implementation, and a total of two days or 12 hours of training in the second year of

implementation). In Springfield, teachers in later cohorts, especially those in Cohorts 4.5 and 5, were given less time to learn about SIM-CERT routines and to apply them to the content taught in collaborative work sessions with their peers. The majority of Springfield teachers in Years 3 and 4 received training in the required routines, but this information was covered in a condensed period of time.¹³⁹

Pronounced differences were observed in Year 5 in comparison to prior years related to training in required content and satisfaction levels. For the first time in the history of the grant, less than half of Springfield teachers received training in all four core routines, a finding that perhaps was the result of a significantly reduced training schedule in one school in particular. Moreover, the monthly, after-school, training workshops in Springfield that were discontinued in Year 4 were not reinstated in Year 5, which potentially further contributed to lower teacher satisfaction levels in Springfield. Although the developer concurred with Springfield in Year 3 that a specified and defined amount of time devoted to training teachers in SIM-CERT was not critical to the overall implementation of the intervention (i.e., shorter sessions were equivalent to longer sessions), Springfield data on classroom usage and self-reports of satisfaction with professional development contradict this assertion. In addition, in Chicopee there were also lower levels of teacher satisfaction with professional development and lower percentages of teachers being trained in all four core routines than had been observed in prior years. This shift may have resulted from several factors. First, both Chicopee schools had to share a single coach in Year 5. Second, teacher and administrator interviews confirm that there was less emphasis on the intervention given that this was the final grant year. Finally, many of the remaining teachers trained in Chicopee were from non-academic content areas, and administrator interviews noted that this population of teachers did not perceive the intervention to be pertinent to the content they taught.

¹³⁹ As stated previously, the reasons why training time was condensed over time were unclear but may have been related to administrative changes in oversight and accountability.

Participant recruitment. The manner in which teachers were informed about their participation in SIM-CERT was not conducive to the creation of widespread buy-in among school staff.¹⁴⁰ For example, at the start-up of the grant, teachers were given minimal notice that they could not attend other professional development sessions in August 2006 in a well-intentioned effort to keep proposed plans for training on track despite a later-than-anticipated start. Teachers were instead required to attend training in SIM-CERT. Across years, coaches explained that a large part of their work included building teacher buy-in for the intervention so as to increase levels of implementation in the classroom.

Over time, the decline in satisfaction (observed in Year 4) may have been influenced by a lack of communication and understanding about original training requirements, as per the model and grant stipulations. These were years when the leadership was to take an even more active role in CERT implementation given less of a role in the past, especially within schools. Although problems with communication of the SIM-CERT implementation plan were reported in both districts among multiple stakeholders, this concern was voiced more frequently in Springfield. Across years, teachers and administrators in Springfield reportedly had misconceptions about the roll-out of the intervention, including a misunderstanding that only some, rather than all, teachers were to be trained. High levels of administrative turnover in Springfield (discussed below) resulted in diminishing numbers of administrators trained in SIM-CERT over time. Although district team efforts were reportedly made, newly hired administrators possessed limited knowledge of the intervention or the grant stipulations for implementation.

Support and Accountability

Coach support. Results from focus groups and interviews with teachers indicated that the relationship between the coach and the teacher as well as the coach's association with accountability efforts and support from administrators collectively influenced the coach's efficacy. Teachers, administrators, and literacy coaches stated that the relationship established

¹⁴⁰ This conclusion was based on a triangulation of data gathered from teacher focus groups and surveys, interviews with literacy coaches, administrators, the developer, and the SR district team including district documents. Difficulties in the initial year have been described in detail in prior reports.

between individual teachers and the coach determined, in large part, how much impact the coach could have on teachers' instructional practice. Over time, multiple data sources suggested that the school-based coach was an essential component to supporting and increasing levels of classroom implementation. These data highlighted the importance of coaches assuming a supportive, rather than an evaluative role in the implementation of the whole-school intervention. In Years 2 and 3, literacy coaches stated that teachers were more likely to seek help with implementation if they perceived the coach to be accessible, approachable, non-judgmental, and generally supportive. Coaches and teachers reported that willingness among coaches to answer questions, trouble shoot, and individualize feedback contributed to a successful coach/teacher relationship. Furthermore, coaches' willingness to assist teachers with issues not directly related to SIM-CERT, such as classroom management and procuring teaching materials, helped build the necessary trust for engaging in other discussions pertinent to SIM-CERT implementation.

Over time, teachers, administrators, the SR district team, and coaches reported a consistent positive rapport between teachers and coaches in Chicopee but a more mixed rapport in Springfield (in Years 4 and 5 in particular). Levels of satisfaction with the coaching received declined markedly in Springfield during Year 5, as did the number of teachers working directly with the coaches, which may correspond with lower rates of classroom usage in this district. An analysis of interview and focus group data revealed that the reasons for teachers' dissatisfaction with literacy coaches varied by school and individual teacher but was generally related to the coach's availability, the degree to which the coach was seen as a "watchdog" for administrators, and the extent to which teachers felt the coach provided practical support. Additionally, it is important to note that in Year 5 there were fewer coaches across all schools in Springfield than in any prior year of the grant. In Chicopee the general consensus by teachers in Year 5 was that even though the coach was spread too thin because she was shared between two schools, she was still responsive to teacher needs.

From the coaches' perspective, the following components enabled them to initiate and follow-through on their responsibilities to support classroom implementation: teacher willingness to engage in conversations about changes in teaching practice; school culture and expectations

regarding open classrooms; and administrator support and union stipulations to allow teacher observations and feedback (teachers explained that when they perceived the coach as evaluative and critical, they were less likely to open their classrooms for observations and invite the coach to help them incorporate SIM-CERT into their instruction). In Year 5, coaches indicated in interviews that they primarily provided the following supports to teachers: observing teacher lessons, leading workshops and trainings, and modeling lessons.

Over time, an analysis of interview, focus group, and document data suggested that the ability of coaches to maintain this supportive role depended in part on the support the coaches themselves received from administrators. More specifically, coaches reported that it was of paramount importance to their efficacy that administrators: (1) preserve direct work with teachers as coaches' primary responsibility (i.e., support for classroom implementation via classroom visits and planning/reflective meetings with individual teachers on instructional practice); (2) limit coaching responsibilities not directly related to supporting teaching practice and building rapport with teachers; and (3) assume direct responsibility for accountability in communications with teachers. In Year 4, Springfield coaches were involved in administrator "learning walks" and collaborated with administrators to collect SIM-CERT portfolios. Although the "learning walks" did not continue in Year 5, focus groups data showed that Springfield teachers in particular continued to reflect on their negative experiences when they perceived coaches to be in an evaluative role. In general, focus group data from Springfield in Years 4 and 5 indicated perceptions of the literacy coach's helpfulness diminished when the coach was seen as affiliated with these SIM-CERT accountability efforts, particularly related to directives with a bearing on teacher performance evaluations.

Administrator support and promotion of accountability. Administrator support and interest in SIM-CERT was reported to be minimal or non-existent by coaches and teachers in Year 5 interviews and focus groups. Across the years, one of the most frequently cited barriers to implementation among teachers, literacy coaches, and the developer and district team was the lack of accountability for implementation from school-level administrators. Although the developer noted this challenge across districts, other reports indicated it was a more significant

issue in Springfield where lower rates of classroom usage were observed as compared to Chicopee.

In Year 5, coaches and ELA ILS department chairs specifically cited a diminishing level of accountability and monitoring as one of the barriers to the implementation of SIM-CERT. Particularly in Years 2 and 3, Springfield teachers and literacy coaches explained that administrators did not require teachers to either attend trainings or to use SIM-CERT routines with their students (despite the efforts of the district team to hold school leadership accountable for implementation in their schools). Rather, inclusion in trainings and use of SIM-CERT was “recommended” and predominantly left to individual teacher discretion. In some cases, administrators and literacy coaches in Springfield reported that the teacher-contracts or bargaining agreement prevented administrators from establishing requirements for SIM-CERT implementation. Furthermore, it was reported that the union prohibited mandatory classroom visits, allowing administrators (and literacy coaches) entry into only those classrooms where they were invited, thus restricting the ability of administrators and coaches to monitor implementation levels across classrooms.

The loss of interest in and support for the intervention was noted in both districts and may have contributed to the low levels of respondent satisfaction with SIM-CERT training (as well as the observed reduction in reported use of CERT). One coach noted that there was a lack of top-down accountability for the program, and another commented that the administration did not have the “chutzpa” to hold people accountable for implementing the intervention despite their plans to do so. One teacher noted that the new principal at her school was “busy putting out fires” so SIM-CERT had become less of a focus. Another teacher indicated that SIM-CERT was completely dropped and “never mentioned.”

High administrator turnover in Springfield may be another factor related to lower rates of classroom use and satisfaction in this district as compared to Chicopee (see Appendix B for details on administrator attrition in Years 1–5). For example, in Chicopee, one school retained the same principal and assistant principal (responsible for SIM-CERT) all five grant years, whereas the other school had two principals during this period. In Springfield, one school had

five principals, one school had three principals, and one school two principals across the five grant years. According to the original implementation plan, which had assumed low attrition rates of administrators across the five years of the grant, administrators were to be trained in SIM-CERT in Year 1 to promote implementation over time. However, with high administrator turnover in Springfield, new administrators did not receive the same training despite district team efforts and new administrators generally lacked knowledge of the intervention and implementation requirements. Springfield coaches reportedly provided information to administrators in Years 3 and 4, but indicated it was another task added to their workloads. In Year 5, Springfield coaches reportedly did not continue to provide information to administrators as they appeared to have limited time and little to no interest in the program during this final grant year.

In Year 4, the SR district team collaborated with the school-based literacy coaches to: (1) transfer more accountability for implementation to the schools and (2) to provide school administrators with the tools to follow-up on implementation levels with their teachers. A review of district and developer documents and interviews with literacy coaches shows that learning walks were conducted during the fall semester of Year 4, and that attempts to collect SIM-CERT portfolios (examples of SIM-CERT devices or graphic organizers as a lesson planning tool) were also made by administrators, as planned. These efforts were not sustained in the spring semester or in Year 5 of the grant.

In general, teachers in Springfield indicated in focus groups that there was no longer any administrator support for or mention of SIM-CERT at all. Analysis of available data indicates that accountability efforts in Chicopee, though structured differently with department chairs responsible for implementation, remained consistent over time. Chicopee teachers in Year 5 however, indicated that the accountability requirements had become looser in the final year of the grant and that there was no longer a focus on collecting devices and other implementation data from teachers (when leadership and accountability was the responsibility of the school rather than the district). Although Springfield teachers recounted their previous negative experiences with school administrators and SIM-CERT accountability, Chicopee teachers at one

school indicated in the Year 5 focus group that administrators held somewhat of a neutral role, neither supporting nor impeding the implementation of SIM-CERT. At the other Chicopee school, however, teachers stated that administrators were highly supportive of their efforts to implement SIM-CERT. Although teachers differed by district in their perceptions of the supportiveness of administrators, coaches across both districts indicated that they did not feel supported by administrators. In turn, school administrators contended that communication and support from the district to the school was lacking and in need of improvement.¹⁴¹

Satisfaction with Professional Development

In addition to district-supplied documentation regarding teachers' receipt of SIM-CERT training, teachers also provided information via surveys and focus groups about their professional development experience.¹⁴²

Satisfaction with Formal Training

Across Years 2–5 and across all cohorts, teachers were asked in the survey whether SIM-CERT training prepared them to implement the classroom model and whether they were pleased with the amount and quality of training received.

When looking at a cross-section of the SIM-CERT teachers who received training in Years 2, 3, and 4, levels of satisfaction generally rose from Year 2 to Year 3 and then fell in both Years 4 and 5 (Refer to Appendix B for more detail regarding response rate and survey respondent characteristics).¹⁴³ Similar patterns of teacher responses were observed for the levels of satisfaction with the amount and quality of training received. In contrast to previous years, less than half of the SIM-CERT-trained teachers who responded to the survey indicated that they

¹⁴¹ This comment was primarily directed at district leadership rather than the district implementation team; the latter had authority from the former and acted only with administrative and leadership support.

¹⁴² In Year 2, 67% of SIM-CERT-trained teachers responded to the survey. In Years 3 and 4, 73% and 79% responded, respectively, and in Year 5 65% responded. Percentages refer to the proportion of SIM-CERT-trained teachers (of the total possible as reported by the district) who completed the survey in Year 2, 3, 4 and 5. The individual teachers who responded in any given year may differ; responses have been presented by grant year.

¹⁴³ Categories of “agree” and “strongly agree” were collapsed across three items related to teacher satisfaction levels as reported above.

were satisfied with SIM-CERT training sessions. Refer to Exhibit 50 below for teacher responses on average to these questions.

Exhibit 50. Teacher satisfaction levels with SIM-CERT training workshops

Year	District	Survey Item		
		Training sessions prepared me to effectively use these routines in the classroom	I am pleased with the amount of SIM-CERT training	I am pleased with the quality of SIM-CERT training
Year 2 survey	SPS (n = 78)	64%	67%	74%
	CPS (n = 67)	70%	67%	60%
	Total (n = 145)	67%	67%	68%
Year 3 survey	SPS (n = 135)	72%	77%	85%
	CPS (n = 73)	84%	96%	99%
	Total (n = 208)	76%	84%	90%
Year 4 survey	SPS (n = 156)	59%	54%	56%
	CPS (n = 79)	76%	78%	89%
	Total (n = 235)	67%	63%	67%
Year 5 survey	SPS (n = 173)	39%	29%	31%
	CPS (n = 124)	59%	65%	66%
	Total (n = 297)	47%	44%	45%

Similar to Year 4, the survey results above also illustrate district variation in satisfaction levels with professional development provided in Year 5 (see Appendix B for figures depicting district variation in teacher perceptions of SIM-CERT training sessions across Years 2, 3, 4, and 5).

Variation in satisfaction with the amount of training between districts became apparent in Year 3 (77% in Springfield versus 96% in Chicopee) and more pronounced in Year 4 (54% in Springfield compared with 78% in Chicopee). This gap between the districts grew even further in Year 5 (29% in Springfield and 65% in Chicopee), although satisfaction levels in both districts dropped appreciably from that in Year 4. Levels of satisfaction with the quality of training sessions increased dramatically from Year 2 to Year 3 in Chicopee (60% to 99%,

respectively), with levels of satisfaction in Year 4 (89%) remaining higher than in Year 2. In Year 5 in Chicopee the level of satisfaction with training quality (66%) did not drop to its low point of Year 2 but was markedly lower than in Year 4. In Springfield, the percentage of respondents satisfied with the quality of training increased from Year 2 to Year 3 (74% to 85%) but was lower in Year 4 (56%) and lower still in Year 5 (31%, the lowest level overall).

Survey Context

A number of key points emerged from the Year 5 survey data. Some of these findings were congruent with those from past surveys, and some were not. It is important then to note the distinction between the sample size of the Year 5 survey respondents and that of previous grant years. More specifically, in Year 5 the overall response rate from both districts combined was the lowest it had been in previous years at 66%. The districts' individual survey response rates were almost identical, with Springfield schools having a 66% percent average response rate and Chicopee schools having a 67% average response rates. Response rates by school ranged from a high of 71% in one Springfield school to a low of 59% in another Springfield school. Lower response rates not only make it more difficult to generalize findings to the larger group of SIM-CERT-trained teachers in the districts but may also suggest a waning interest by teachers in participating in SIM-CERT-related activities. Dwindling interest levels by teachers in SIM-CERT may be related to the relative absence of coaching support in Year 5 across all schools.

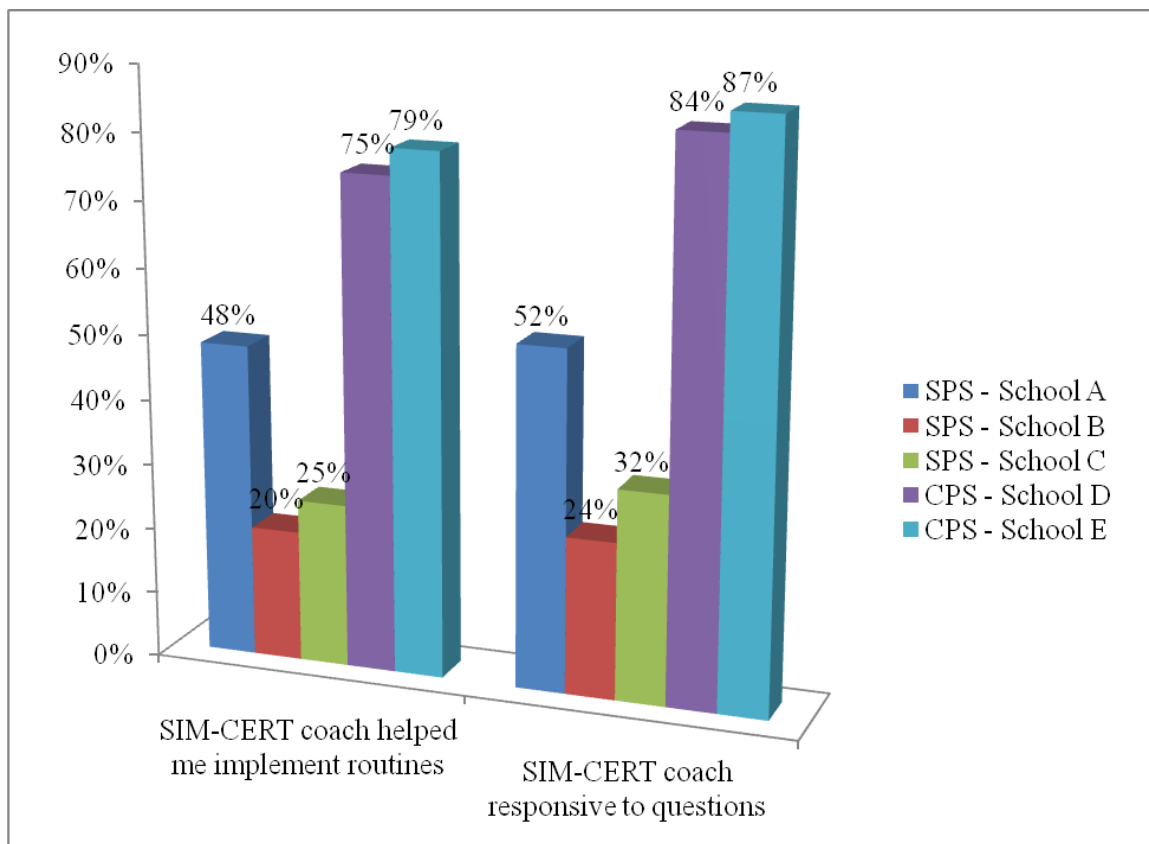
Satisfaction with Coaching Support and Training

In Years 2 and 3, the consensus among teachers and administrators was that the support provided by the literacy coaches had been instrumental in the classroom-level implementation of SIM-CERT. In fact, focus group participants in Years 2 and 3 cited school-based literacy coaches as the most critical factor in determining their implementation of SIM-CERT. Similar to Year 4, in Year 5 focus group participants, survey respondents, and administrators had mixed comments regarding the support of the literacy coach, with the majority in Springfield expressing negative perceptions and the majority in Chicopee expressing positive opinions. In general, survey and qualitative results indicated that coach support varied by district and by school, depending on the

rapport between teachers and the coach, the manner in which coaches communicated feedback to teachers on SIM-CERT implementation, and whether teachers perceived coaches as serving an evaluative function in their classrooms.

In Years 2–5, teachers were asked on the survey to indicate their satisfaction with the support and mentoring received from their school-based SIM-CERT coach. The exhibit below displays survey results across years.

Exhibit 51. Teacher perceptions of SIM-CERT coach supportiveness



Prior to the current year, the majority of teachers in both districts agreed or strongly agreed that coaches were “responsive to their needs” and supported their implementation of SIM-CERT. The percentage of respondents reporting support and responsiveness from their school-based

coach was generally higher in Years 2 and 3, falling slightly in Year 4 and falling appreciably in Year 5, particularly in Springfield.

There was also a decrease from Year 4 to Year 5 in the percentage of teachers within each district who agreed or strongly agreed that coaches provided support for implementing routines and were responsive to questions. Specifically, in Year 4 nearly all respondents from Chicopee agreed or strongly agreed that their coach supported implementation and was responsive to their questions (95% and 96% respectively), whereas in Year 5 these numbers decreased to 77% and 85% respectively. In Springfield, in Year 4, 58% and 65% of teachers respectively indicated that they agreed or strongly agreed that the coach helped them implement routines and was responsive to their questions, whereas in Year 5 these percentages fell to 30% and 34%, a much steeper drop than was observed in Chicopee. In both Years 4 and 5, Chicopee teacher ratings regarding assistance from the SIM-CERT coach were higher than those of Springfield teachers. A cross-sectional analysis of responses from groups of teachers who responded to the survey in Years 2, 3, 4, or 5 show that levels of satisfaction with coaching support were generally high in both districts in Years 2 and 3, but that as the percentage of teachers satisfied with their coaching experience in Years 4 and 5 decreased, especially in Springfield, this shift affected overall satisfaction levels.

Similar to Year 4, in Year 5 there was also much greater variation in satisfaction levels among Springfield schools than between Chicopee schools. As described earlier, this variability among Springfield schools was likely related to the following factors: (1) a coach at one school was part-time, (2) a coach at another school was on medical leave for a number of months, and (3) there was no coach at the third school as this person was promoted to an administrator position. Additionally, in Springfield the coaches were responsible for supporting teachers as well as enforcing accountability. Data from Year 5 focus groups indicated that teachers felt coaches were more effective when they were “easy going” and did not “jam SIM-CERT down our throats.” Several Springfield focus group teachers also noted their discomfort with literacy coach and administrator walk-throughs and indicated that these were not helpful, nor was negative feedback from the literacy coach regarding the devices they submitted for review.

X. Whole-School Intervention Impacts

The impact of the whole-school intervention (SIM-CERT) on student achievement, specifically achievement in English language arts (ELA) inclusive of reading, was estimated over time.¹⁴⁴ A quasi-experimental rigorous assessment of the impact utilized a short interrupted time-series analysis (SITS) inclusive of a comparison group.¹⁴⁵ Student achievement trends at the Striving Readers high schools were compared to trends at other high schools in Massachusetts serving similar student populations (see Exhibit 52). Aggregate student achievement scores as measured by the state ELA assessment (MCAS ELA, inclusive of reading) were obtained from both treatment and comparison schools. Aggregate scores were included for each cohort of 10th grade students from each of the five years pre-treatment (2001–02 through 2005–06) and from each of the first four years during the treatment period (2006–07 through 2009–10).

Analytic Sample

The analytic sample was comprised of the five treatment schools within the two participating districts and six comparison schools within four identified comparison districts. These districts—and the schools within them—were identified based on aggregate information including state assessment performance and demographic information publicly available at the time.¹⁴⁶ Given preexisting differences between the two treatment districts (and among schools within these districts), an appropriate match would include variability in outcomes and demographic characteristics (refer to district context, Section II). Initially, districts of similar sizes were identified; then performance and other characteristics were examined to provide an adequate comparison group, the appropriateness of the match to be assessed in later analyses.

¹⁴⁴ Outcomes for teachers were not proposed as there were no secondary data available to assess teacher-level outcomes.

¹⁴⁵ Refer to Bloom (2001). Source: <http://www.mdrc.org/>

¹⁴⁶ A data-sharing agreement was executed with the Massachusetts Department of Education (MA-ED) later in the study to obtain more complete data and associated common core data for comparison schools in the state. For the reported analyses, evaluators were given access to limited data for the comparison group, including the mean MCAS ELA scores by school within districts already identified as “matches” based on publicly available test score and common core data related to the student population. The District Analysis Review Tool (DART), launched in 2011, provides a method by which districts and schools can be matched for comparison; however, it was not available for prior sample selection. Source: <http://www.doe.mass.edu/apa/dart/>

The following exhibit presents descriptive information about the analytic sample inclusive of MCAS ELA scores by district and treatment group (i.e., SIM-CERT or matched comparison schools). As noted, variation was observed among aggregate demographic characteristics, although ELA performance levels were fairly consistent among the districts. Tables presenting this information at the school level have been included in Appendix E.

Exhibit 52. Sample characteristics for treatment and comparison groups by district

	Treatment			Comparison		
	SPS District 1	CPS District 2	District 3	District 4	District 5	District 6
School n	3	2	2	2	1	1
Race/Ethnicity (%)						
White	14.7	65.5	68.3	18.8	6.1	37.8
Black	22.3	3.1	7.0	3.3	1.9	6.6
Asian	2.2	1.6	0.8		2.4	28.4
American Indian	0.1	0.2	0.3	0	0.1	0.2
Other	4.1	2.3	3.1	0.1	0.2	1.7
Female Gender (%)	48.2	48.1	48.2	47.9	47.2	48.1
Special Education Status (%)	23.9	16.5	18.3	25.2	19.8	15.8
First Language Not English (%)	24.1	13.5	25.2	50.9	79.1	43.7
Limited English Proficiency (%)	13.1	4.5	5.2	23.3	23.1	32.4
Free-Reduced Lunch Status (%)	81.4	60.7	74.6	74.3	86.7	69.7
Attendance (mean)	164.3	168.2	165.9	164.7	168.2	168.7
MCAS ELA Performance Level (%)						
Advanced	3	7	5	3	3	6
Proficient	34	48	40	29	38	40
Needs Improvement	40	34	38	37	40	36
Failing	22	10	17	31	19	17
MCAS ELA Scaled Score (mean)	234.5	243.1				
Enrollment (mean)	25,141	7,845	9,886	5,901	12,284	13,331

Note. Data were obtained from the Massachusetts Department of Education and presented for the 2010 school year. Other includes a combination of White, Black, Asian, American Indian, Native Hawaiian, and Hispanic. The maximum number of days of attendance is 180.

Finally, as originally proposed, the complete analysis of the whole-school intervention (SIM-CERT) outcomes on student achievement was to be conducted and presented at the conclusion of

the Striving Readers grant, when complete data were available for all five years of implementation. The timeline for state data sharing was changed to the end of the calendar year, so the final study year of state assessment data (MCAS ELA) were unavailable. Therefore, the SITS analysis does not include scores from the final study year (i.e., 2010–11). The aggregate, school-level ELA scores of four cohorts of 10th grade students were combined for analysis.

Statistical Analyses

Analyses were conducted to answer the research question, “*Does school participation in SIM-CERT improve 10th graders’ ELA achievement relative to that of a comparison group?*” using schools as the primary unit of analysis (student scores aggregated at the school level). The analytic process included: (1) determination of a baseline projection model (Two types of projection models were considered: a baseline mean projection model and a linear projection model); (2) fitting preliminary impact models to the data to determine whether the models needed adjustments for potential autocorrelation; and (3) fitting a final short interrupted time series model to estimate the impact of SIM-CERT on post-treatment, school-level outcomes.¹⁴⁷

Analytic Model and Specifications

The dependent variable (outcome) used to estimate the impact of the targeted intervention on students’ ELA achievement, as previously noted, was the state English language arts assessment (MCAS ELA). These scores were measured on a continuous scale, using the scaled scores provided by the state, aggregated at the school level.

To assess whether a baseline mean projection model would be appropriate, or whether a slightly more complex linear projection model would be required, separate models were fit to the data for treatment and comparison schools for the years *prior* to the treatment period (2001–02 through 2005–06), and tested whether the pre-treatment time slope was

¹⁴⁷ SAS proc mixed was used to fit these models. The TA provider, in particular Cris Price, provided a critical review of our final models and provided the graph included.

significantly different than zero in each group. In both treatment and comparison schools, there was a significant, positive slope in the pre-treatment years (Models 1 and 2 included in the exhibit below). These results indicated that a baseline linear projection model was more appropriate than the less complex baseline mean projection model.

Exhibit 53. Multilevel models estimating slope of MCAS ELA scores in pre-treatment years for treatment, comparison, and combined schools

	Model 1 (Treatment Schools)	Model 2 (Comparison Schools)	Model 3 (All Schools Combined)	Model 4 (All Schools Combined)
Fixed Effects				
Intercept	234.32*** (2.66)	234.42*** (2.40)	234.36*** (2.47)	234.42*** (2.53)
Time	0.79* (0.32)	0.84*** (0.18)	0.81*** (0.18)	0.84** (0.26)
Treatment Group			0.03 (3.40)	-0.11 (3.58)
Treatment Group *Time				-0.04 (0.37)
Random Effects				
σ_{μ}^2	29.61~ (21.69)	26.88~ (19.24)	28.27* (14.47)	28.25* (14.47)
σ_{ϵ}^2	5.25** (1.70)	1.64** (0.53)	3.36*** (0.76)	3.44*** (0.79)
Goodness-of-Fit -2LL	124.0	101.5	231.5	231.6

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Note. Standard errors are in parentheses. Time is coded -5, -4, -3, -2, -1. Treatment school is coded =1; Comparison school is coded = 0.

Additional results shown in Exhibit 53 indicate that there was no significant difference between the treatment and comparison groups' intercept levels, as shown by the non-significant coefficient for the treatment group indicator in Model 3. There was also no significant difference in pre-treatment slopes, as indicated by the non-significant treatment group by time interaction coefficient in Model 4. Based on these results, the whole-school or SIM-CERT impact models assumed a baseline linear trend prediction model.

Preliminary whole-school impact models with adaptations to account for potential autocorrelation were also fit to the data. No evidence of autocorrelation among the repeated measures within schools was found. Therefore, the final models did not need adjustments for potential autocorrelation among the repeated observations within schools over time.

Whole-School Impact

The short-interrupted time series with comparison group model was constructed to estimate the impact of SIM-CERT on change over time in aggregate ELA achievement scores. The final model was used to estimate the impact of SIM-CERT on school means at the end of the first year of treatment and the impact of SIM-CERT on the growth in achievement scores during the treatment years. The amount of variance to be predicted between schools in aggregate ELA outcome scores, over time, was 57%.

Results from these analyses indicated there were no significant impacts of treatment on mean tests scores at the end of the first year of treatment or on the growth in achievement scores during the treatment years.

The final model is summarized in Exhibit 54. In this exhibit, the coefficient for “*Time*” is the model-predicted time slope in absence of treatment. The coefficient for “*Treatment School*” is the difference between treatment and control school intercept levels (i.e., the predicted mean at time = 0 if there were no treatment). The coefficient for “*Spline*” is the difference between the control schools’ projected mean and observed mean at the end of the first treatment year (time = 0). The coefficient for “*Treatment school * Spline*” is the treatment effect at the end of the first year of intervention. It is the difference in differences between the projected and observed means in treatment and comparison schools at the end of the first year of intervention. The effect was not significantly different than zero. The coefficient for “*Treatment School * Spline * Time*” is the treatment effect on the post-treatment slope. The effect was also not significantly different than zero, indicating that treatment and comparison schools had similar growth in mean test scores during the treatment years.

Exhibit 54. Multilevel model describing the relationship between MCAS ELA scores and SIM-CERT, across five pre-treatment and four study years

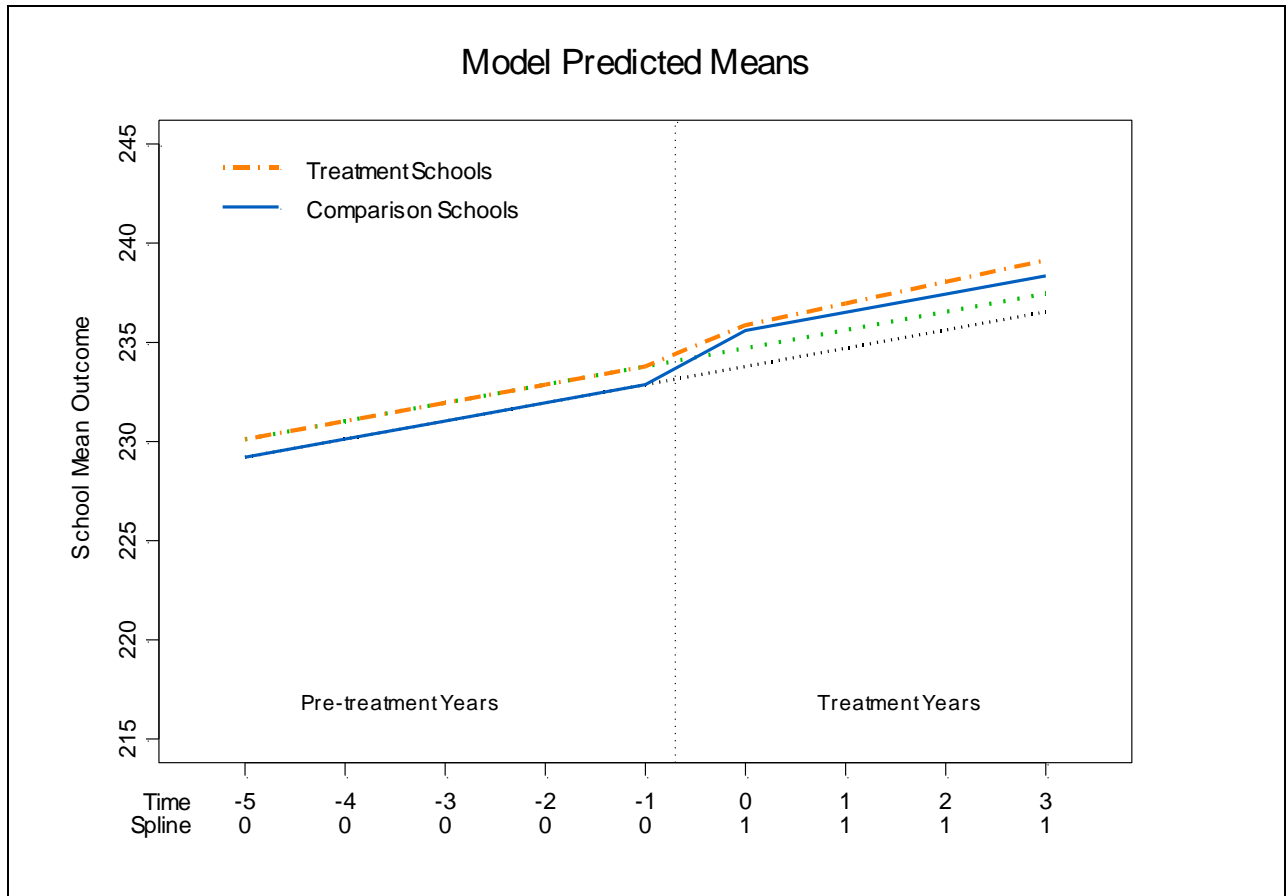
	Model
Fixed Effects	
Intercept	233.79*** (2.16)
Time	0.92** (0.18)
Spline	1.81~ (0.98)
Treatment School	0.91 (3.11)
Treatment School * Spline	-0.65 (1.05)
Treatment School * Spline * Time	0.17 (0.43)
Random Effects	
σ_{μ}^2	25.30* (12.20)
σ_{ϵ}^2	3.86*** (0.62)
Goodness-of-Fit -2LL	416.3

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Note. Time is coded -5, -4, -3, -2, -1, 0, 1, 2, 3. Treatment school is coded = 1; Comparison school is coded = 0. Spline is coded = 0 if pre-treatment year (i.e., time is -5,-4,-3,-2, or -1) and = 1 if treatment year (i.e., time is 0, 1, 2, or 3). The random intercept for schools is σ_{μ}^2 and σ_{ϵ}^2 representing residual variation of scores within schools over time.

On average, students' ELA achievement scores have increased by approximately 1 point per grant year, lower than the 2.3 point increase observed prior for three years of implementation. However, results from the current SITS analysis indicated the five Striving Readers schools were performing similarly to comparable schools in the state—in districts not participating in the Striving Readers grant—on the ELA portion of the MCAS. The final model predictions are depicted graphically in Exhibit 55.

Exhibit 55. Model predicted means over time for treatment and comparison schools



Note. Results presented from Exhibit 53, Model 4.

Impact Results Summary

In summary, the results from the pre-treatment years (summarized in Exhibit 53) indicate the treatment and comparison schools were well-matched in that they had statistically equivalent means and slopes in the pre-treatment years.¹⁴⁸

The short interrupted time-series analysis summarized in Exhibits 53 and 54 reveal that scores increased over time for both the treatment and comparison schools before and after the treatment years. While both treatment and comparison schools did exhibit an increase or “jump” in scores from pre- to post-treatment, this increase was not significantly different between the two. Nor was there a significant treatment effect on the growth in scores in the treatment years. Treatment and comparison schools had similar growth during the treatment years. In conclusion, although the five Striving Readers schools implementing SIM-CERT increased their ELA achievement scores over time, there was no evidence that the increases were due to SIM-CERT as similar increases were observed for the comparison schools.

Any number of similar initiatives may have been implemented in the comparison group schools that could explain a lack of observed impact results (i.e., no significant differences between the Striving Readers and non-Striving Readers schools on overall aggregate ELA achievement scores).¹⁴⁹ Comparison schools may have been implementing an intervention or made curricular changes with equal intensity to affect outcomes. In addition, a lack of observed impact results may be a function of a less than ideal sample size combined with less than ideal fidelity of implementation across treatment schools (refer to SIM-CERT implementation). That is, even if implementation was perfectly executed in one or two of the

¹⁴⁸ All analyses presented here were conducted with an additional comparison group constructed based on the District Analysis Review Tool (DART). DART analytically matches districts and schools, based on performance and demographic characteristics, to others similar in the state for comparison purposes. The DART district comparison group consisted of the four in the originally constructed comparison group with two additional districts included. Results from these analyses were consistent with those reported here for the matched comparison group selected prior to 2011 and the DART system of matching, with data available at that time.

¹⁴⁹ Especially in the context of schools in need of improvement and restructuring, this is likely to be the case. However, data were not readily available to assess this assumption.

schools, the overall effect may not have been strong enough to illustrate differences in comparison to the other schools with a small sample size.

XI. Whole-school Intervention Impact and Implementation

A non-experimental assessment of the relationships between SIM-CERT training and implementation and school-level achievement scores over time was explored. Student achievement scores, as measured by the MCAS ELA, from each cohort of grade 10 students assessed in participating high schools were analyzed for the first four years of the treatment period (2006–07 through 2009–10).

Although the previously presented analysis of the impact of the whole school intervention was conducted to assess a causal relationship, if one was present, the following analyses do not attempt the same.¹⁵⁰ The previous analysis included a well-matched comparison group to address the counterfactual (i.e., what would happen in absence of treatment); the analyses presented here do not include a comparison group. Any observed association or relationship between the whole-school intervention and ELA achievement scores over time does not imply cause but is merely correlational and descriptive in nature; the true cause cannot be identified without an experimental or quasi-experimental design.¹⁵¹

Levels of Implementation

Levels of implementation were defined based on developer specifications, as described previously in the SIM-CERT implementation section. In the case of SIM-CERT, the level of training and implementation in the classroom were defined as adequate or not by the number of teachers receiving the minimum amount of expected training hours and the number of teachers self-reporting the minimum amount of expected classroom implementation activities. These two implementation levels or ratings were officially “scored” as required for reporting purposes. In addition to these required ratings of adequacy, the number of teachers reportedly receiving the

¹⁵⁰ It is important to note the limitations of the prior analyses, as already described in the SIM-CERT impact section.

However, even with a well-matched comparison group included, an assessment of aggregate school-level impacts like those reported here would not currently be considered for review by the What Works Clearinghouse (WWC).

¹⁵¹ Refer to Shadish, Cook, & Campbell (2002).

minimum amount of required content delivered in the training sessions (even if they did not receive the minimum hours) and the number of Unlike in Year 4 where there was school-level variation only in Springfield for teachers who met minimum implementation requirements, in Year 5 both districts exhibited school-level variation. In Year 5, both districts also continued the general pattern of decreasing percentages of teachers meeting minimum requirements that was first observed from Year 3 to Year 4 teachers self-reporting exceeding the minimum of expected classroom implementation activities were also examined in relationship to aggregate student ELA achievement over time.

Analytic Sample

The analytic sample was comprised of the five treatment schools within the two participating districts. The aggregate school level ELA achievement scores of four cohorts of 10th-grade students were combined for analysis.¹⁵² Exhibit 56 presents descriptive information about the analytic sample inclusive of ELA scores by treatment school (i.e., SIM-CERT).

¹⁵² Because the timeline for state data sharing was changed to the end of the calendar year, the final study year of state assessment data (MCAS ELA) were unavailable.

Exhibit 56. Sample characteristics for treatment group by district and school

School	Springfield			Chicopee	
	A	B	C	D	E
Race/Ethnicity (%)					
White	7.1	11.8	11.1	75.1	67.0
Black	29.1	27.1	24.1	3.0	4.1
Asian	1.4	1.5	2.2	1.5	0.8
American Indian	0.1	0.1	0.0	0.3	0.1
Other	62.3	59.5	62.6	20.1	75.2
Female Gender (%)					
Female Gender (%)	49.9	53.2	46.6	44.0	53.4
Special Education Status (%)					
Special Education Status (%)	28.6	23.3	29.5	15.6	14.4
First Language Not English (%)					
First Language Not English (%)	31.2	27.0	32.6	9.5	15.1
Limited English Proficiency (%)					
Limited English Proficiency (%)	15.5	8.6	16.4	1.5	2.2
Free & Reduced Lunch Status (%)					
Free & Reduced Lunch Status (%)	77.1	72.4	80.9	48.3	52.5
Attendance (mean)					
Attendance (mean)	150.5	159.5	148.5	167.5	166.1
MCAS ELA Performance Level (%)					
Advanced	3	2	5	7	14
Proficient	34	42	29	56	54
Needs Improvement	45	46	45	31	25
Failing	19	11	21	5	7
MCAS ELA Scaled Score (mean)					
MCAS ELA Scaled Score (mean)	234.3	235.7	233.5	242.0	244.2
Enrollment					
Enrollment	1,380	1,632	1,320	1,437	1,200

Note. Data were obtained from the Massachusetts Department of Education and presented for the 2010 school year. “Other” includes a combination of White, Black, Asian, American Indian, Native Hawaiian, and Hispanic. The maximum number of days of attendance is 180.

As noted previously, variation was observed among aggregate demographic characteristics and ELA performance between the two districts.

Statistical Analyses

Analyses were conducted to answer two research questions. The first question had been posed in prior years: (1) *Was SIM-CERT implementation associated with between-school differences in ELA achievement scores?* The second question was included in this final year of analysis and reporting: (2) *Controlling for schools’ average performance on ELA scores,*

was variation in implementation over time related to ELA scores, such that within schools, were ELA scores better during the years when implementation was better?

The model fit to the data to address the first research question does not include school-fixed effects, so implementation variation was estimated across schools, while the model fit to address the second research question does include school-fixed effects.

Schools were the primary unit of analysis, with student scores aggregated at the school level. The analytic process included: (1) fitting initial models to estimate increases in student achievement over time; (2) assessing covariates for inclusion in the final model, if significant, given degrees of freedom; and (3) assessing any difference in student achievement scores between schools (see Exhibit 2) or within schools over time (see Exhibit 3) as predicted by variation in SIM-CERT training and implementation scores.

Analytic Model and Specifications

The dependent variable (outcome) used to estimate the impact of the targeted intervention on students' ELA achievement, as previously noted, was the state English language arts assessment (MCAS ELA). These scores were measured on a continuous scale, using the scaled scores provided by the state, aggregated at the school level.

The amount of variance to be predicted between schools in aggregate ELA outcome scores, over time, was 79%. The subsequent models that were fit as precursors show that, on average, test scores improved over time (time coefficient is significantly greater than zero). However, no significant school-level covariates were identified to be included in final models at the $p < .05$ level.¹⁵³ This threshold for significance is lower than the $p < .20$ rule for the exclusion of covariates for other reported analyses (e.g., targeted impacts) due to limited

¹⁵³ Individually, school-level measures of percent Limited English Proficiency and percent Special Education Status were significant predictors as was the district indicator, but together in the model none were significant.

degrees of freedom to predict any remaining between-school variation. All of the demographic variables presented in the prior exhibit were assessed.

Impact and Implementation Results Summary

Between School Results

Was SIM-CERT implementation associated with between-school differences in ELA achievement scores?

The results for the first set of analyses included in Exhibit 38 appear to indicate SIM-CERT implementation measures were associated with between-school differences in ELA scores. The models presented in Exhibit 57 indicate two of the four measures of SIM-CERT training and implementation levels were predictive. Schools that met the *minimum* training requirements had higher average ELA scores than schools that did not meet the minimum training requirements. Similarly, schools that *exceeded* the required classroom implementation thresholds also had higher average ELA scores. These results were consistent with those reported in prior years.

Exhibit 57. Multilevel models describing the relationship *between* participating Striving Readers schools' MCAS ELA scores and SIM-CERT

	Model 1	Model 2	Model 3	Model 4
Fixed Effects				
Intercept	232.39*** (0.71)	238.99*** (4.95)	226.08*** (1.99)	238.47*** (3.11)
Time	1.43*** (0.26)	0.95~ (.49)	2.09*** (0.43)	1.06 (0.68)
SIM-CERT				
Minimum–Required Training	9.68*** (1.01)	---	---	---
Minimum–Required Classroom Implementation ^a	---	-1.94 (5.48)	---	---
Exceeded–Required Classroom Implementation ^a	---	---	17.99*** (3.09)	---
Minimum–Required Training Content ^b	---	---	---	0.20 (2.86)
Random Effects				
σ_{μ}^2	---	15.42 (13.55)	---	16.34~ (12.77)
σ_{ϵ}^2	3.31*** (1.00)	3.06** (1.25)	4.94** (1.70)	4.59* (2.29)
Goodness-of-Fit				
-2LL	97.1	83.4	81.1	66.6

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

^a The minimum and exceeded levels of classroom implementation were not assessed in the initial grant year when specifications were provided.

^b The adequate score for the content delivery for professional development was added in Year 3 at the request of the district and was approved by the developer.

Note. Standard errors are in parentheses.

The reported minimum and exceeded levels of classroom implementation were assessed in only three of the four years included in these analyses (2008, 2009, and 2010). The receipt of required training content was assessed in only two of the four years included in these analyses (2009 and 2010). Because only one of the SIM-CERT implementation measures was collected in every year, a potential association with the outcome may be underestimated.

However, the results do not imply that higher implementation caused higher ELA achievement scores. It is equally plausible that schools that were *already* higher performing in terms of ELA scores were more likely to meet the minimum training requirements, and also that that higher performing schools in terms of ELA scores were more likely to implement SIM-CERT at a higher level.

Within School Results

Controlling for schools' average performance on ELA scores, was variation in implementation over time related to ELA scores, such that within schools, were ELA scores better during the years when implementation was better?

The results for the second set of analyses included in Exhibit 58 appear to indicate SIM-CERT implementation measures were *not* associated with ELA scores within schools, over time. There was no evidence that when an individual school varied in implementation levels over time, ELA scores were better in the years when the implementation occurred at higher levels. No variation in scores was present to be predicted.

As noted previously, only one of the SIM-CERT implementation measures was collected in every year, potentially underestimating an association with the outcome. In addition, three of the five schools *never* met adequate levels of professional development at any point over time. These schools could never attain adequate training levels as planned because sessions were delivered after the school year ended. Delivering the complete training in the summer following the implementation school year meant that these schools were always attempting to “catch up.”

Exhibit 58. Multilevel models describing the relationship within participating Striving Readers schools' MCAS ELA scores and SIM-CERT across study years

	Model 1	Model 2	Model 3	Model 4
Fixed Effects				
Intercept	232.62 (0.92)	239.80 (4.78)	236.28 (5.02)	235.41 (2.13)
Time	1.38*** (0.24)	0.62 (0.51)	0.85 (0.54)	1.05 (0.68)
SIM-CERT				
Minimum–Required Training	1.39 (4.11)	---	---	---
Minimum–Required Classroom Implementation ^a	---	-6.59 (5.86)	---	---
Exceeded–Required Classroom Implementation ^a	---	---	-3.35 (6.94)	---
Minimum–Required Training Content ^b	---	---	---	-0.35 (2.95)
School				
School B	0.85 (1.07)	0.75 (1.32)	1.33 (4.88)	1.46 (1.95)
School C	-0.85 (1.07)	-1.26 (1.32)	-0.79 (4.89)	-0.05 (2.19)
School D	5.68~ (2.72)	7.80*** (1.48)	7.74 (5.22)	7.69** (1.92)
School E	6.84~ (3.50)	8.37*** (1.86)	7.92 (5.40)	8.51** (2.22)
Random Effects				
σ_{μ}^2	---	---	11.13** (4.37)	---
σ_{ϵ}^2	2.88** (0.96)	2.95** (1.16)	3.18** (1.25)	4.61* (2.30)
Goodness-of-Fit				
-2LL	80.3	58.7	59.4	42.1

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

^a The minimum and exceeded levels of classroom implementation were not assessed in the initial grant year when specifications were provided.

^b The adequate score for the content delivery for professional development was added in Year 3 at the request of the district and was approved by the developer.

Note. Standard errors are in parentheses.

Finally, implementation study results indicate that a number of other interventions began school-wide in the treatment schools in Springfield over the course of the Striving Readers grant. When the onset of these additional interventions was assessed to determine if there was a relationship between outcome scores and the “shock” of the introduction of these interventions as separate from SIM-CERT, none was observed, though results were considered to be “borderline” and not fully conclusive.

Whole-School Impact and Implementation Summary

In summary, the results of this descriptive analyses (not implying causation) indicated that two of the four measures of SIM-CERT training and implementation levels were predictive of ELA achievement *between* schools. Three of the four SIM-CERT implementation variables were not measured in every program year, and therefore a potential association with the outcome may be underestimated. However, the results do not imply that higher implementation levels caused higher ELA achievement scores. Additional explanations for observed results include the possibility that higher performing schools, in terms of ELA achievement scores, may be more likely to implement SIM-CERT at higher levels. That is, schools performing at higher levels could be doing so as a result of factors unrelated to SIM-CERT, such as less staff and administrative turnover, potentially resulting in more clearly defined leadership and stability as a result.

Additional results indicate implementation was not a significant predictor of the growth in ELA achievement scores in the treatment years, *within* schools. There was no evidence that when an individual school varies in implementation levels over time, ELA achievement scores were better in the years when implementation occurred at higher levels. However, three of the five schools *never* met adequate levels of professional development at any point over time. Delivering the complete training in the summer following the implementation school year meant that these schools were always attempting to “catch up,” and this could explain a lack of observed results. Finally, there were a number of other interventions implemented school-wide in the treatment schools in Springfield over the course of the

Striving Readers, making disentangling SIM-CERT results difficult. Although attempts to assess the impact of the onset of these interventions versus SIM-CERT did not yield clear results, such an outcome could have been the result of an inability to define the onset more clearly rather than the mark of no influence at all.

XII. Evaluation Summary

The evaluation of the Springfield-Chicopee's Striving Readers Program had the primary goal of rigorously assessing the effectiveness of the interventions as implemented on reading achievement. In addition, implementation studies were included to present a broad picture of the overall level of implementation in context and a sense of the variability that may have occurred. Differing institutional contexts or constraints influenced the ways in which intervention components were implemented. Districts and schools possessed their own unique complexities, which may have supported or hindered implementation and, in turn, affected outcomes. Finally, implementation analysis indicated barriers faced and addressed throughout the grant period.

Final results from the implementation of Striving Readers interventions to date in Springfield and Chicopee school districts indicated a positive and significant impact on student reading achievement of one of the two targeted interventions. The impact of the whole-school intervention was not established. Implementation studies also indicated alignment of contextual results with outcomes observed and provided a deeper perspective regarding observed results.

The Springfield and Chicopee school districts have overcome many obstacles in the development, planning, and implementation of their Striving Readers grant. In particular, two dissimilar districts have implemented two targeted interventions (all other SR grantees implemented only one) as well as one whole-school intervention. Implementation studies reporting barriers to implementation in Year 1 resulted from both contextual and contractual factors, which did not necessarily emerge from the intervention models but may have resulted from attempts to fit the models as required into this context. Some of the contextual factors included the urban setting, population, and student needs; the various policies of the schools and districts addressing scheduling and administrative issues; and general staffing and personnel matters. Contractual complexities specifically refer to the requirements for the grant implementation; the monitoring and oversight of the fidelity of implementation; and the observance of the rigorous research specifications.

Given the challenges inherent in creating a successful collaboration between two districts and implementing two interventions, it is not surprising that complexities arose which would not normally be encountered in a standard literacy program implementation.

An initial barrier related to the rigorous research requirements, for example, involved the cooperation, ability, and willingness of both districts to incorporate a “true” control group to address the counterfactual (i.e., what would happen in the absence of treatment). Additional challenges involved the need to standardize implementation across two very different district and school systems. Intervention plans necessitated consistent tailoring to accommodate rigorous research study requirements, and district staff and evaluators spent unanticipated time to ensure successful implementation. At the same time, districts faced turnover in lead program staff and administrators, challenges related to communication with stakeholders and participants, and complications in screening, placing, and tracking the population of students who were randomly assigned to participate in the targeted interventions.

These difficulties have had some lasting influence, but over time the districts have sought to address each one as presented in the evaluation reports. Progress was made in overcoming many of the barriers outlined, particularly in Year 2, but also throughout Year 3. Several of the barriers remained insurmountable, particularly in Springfield as they faced unique challenges related to scheduling planned targeted intervention time; a lack of available planned in-service training for the whole school intervention; and high rates of administrative and teacher turnover.

However, districts implemented each of the targeted interventions while maintaining the integrity of the randomized controlled trial design and assignment to the best of their ability and repeatedly demonstrated their commitment to ensuring the success of the grant. District staff collaborated fully with evaluators in all phases of the evaluation. Their serious consideration of any potential positive or negative influences on study outcomes as well as “full disclosure” has been commendable. Such diligence ensures that these final study results have produced information that can be used by policymakers, district administrators, and school staff to make confident choices regarding effective literacy interventions for their students.

References

- Bloom, H. S. (2001). *Measuring the impacts of whole-school reforms: Methodological lessons from an evaluation of accelerated schools*. MDRC Working Papers on Research Methodology. Manpower Demonstration Research Corporation New York, N.Y. Last retrieved December 1, 2005 from: www.mdrc.org
- Bloom, H. S. (2004, March). *Randomizing groups to evaluate place-based programs*. MDRC Working Papers on Research Methodology. Manpower Demonstration Research Corporation New York, N.Y. Last retrieved December 1, 2005 from: www.mdrc.org
- Bloom, H. S., Richburg-Hayes, L., & Rebeck Black, A. (2005, November). *Using covariates to improve precision: Empirical guidance for studies that randomize schools to measure the impacts of educational interventions*. MDRC Working Papers on Research Methodology. Manpower Demonstration Research Corporation New York, N.Y. Last retrieved December 1, 2005 from: www.mdrc.org
- Bloom, H.S., Hill, C.J., Rebeck Black, A., & Lipsey, M.W. (2007, July). *Empirical benchmarks for interpreting effect sizes in research*. MDRC Working Papers on Research Methodology. Manpower Demonstration Research Corporation New York, N.Y. Last retrieved December 1, 2011 from: www.mdrc.org
- Content Learning Center, Kansas University. (2007, February). *CLC program evaluation implementation phase tool kit*. Author.
- Corrin, W., Somers, M. A., Kemple, J., Nelson, E., & Sepanik, S. (2008). *The enhanced reading opportunities study: Findings from the second year of implementation* (NCEE 2009-4036). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Faddis, B. (personal communications, 2007, November). Evaluation materials developed based on information provided by Dr. Suzanne Robinson, University of Kansas, Center for Research on Learning, November, 2007. RMC Research, Portland, Oregon.
- Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).
- Goodson, B. (October, 2006). *Measuring implementation*. Presentation at the U.S. Department of Education Striving Readers Program Local Evaluators Conference, Washington, D.C. Abt Associates, Inc.

- Guskey, T. (2000). *Evaluating professional development*. Thousand Oaks, CA: Corwin Press.
- Hasselbring, T. S., & Goin, L. I. (2004). Literacy instruction for older struggling readers: What is the role of technology? *Reading and Writing Quarterly*, 20(2), 123–144.
- Karlsen, B., & Gardner, E. (1996). *Stanford diagnostic reading test: 1995 multilevel norms book and technical information* (4th ed.). San Antonio, TX: Harcourt Brace Educational Measurement.
- Learning Point Associates. (2008). *CSR practitioner's guide to SBR—What is scientifically based research?* The Center for Comprehensive School Reform, Learning Point Associates. Retrieved January, 2008 from:
<http://www.centerforcsri.org/pubs/pg/sbr.htm>
- Massachusetts Department of Education. (1999, October). *Massachusetts comprehensive assessment system*. 1998 MCAS Technical Report Summary.
http://www.doe.mass.edu/mcas/1998/techrpt_sum.pdf
- Meltzer, J. (2006, October). *Measuring treatment-control differences in literacy instruction: Issues, considerations, and ideas*. Presentation at the U.S. Department of Education Striving Readers Program Local Evaluators Conference, Washington, D.C. Center for Resource Management-Public Consulting Group, Inc. (CRM-PCG).
- Meltzer, J., & Bray, C. (October, 2006). *Core components and instructional practices of reading support in classrooms for middle and high school struggling readers: Summary chart for instrument development/design*. Presentation at the U.S. Department of Education Striving Readers Program Local Evaluators Conference, Washington, D.C. Center for Resource Management-Public Consulting Group, Inc. (CRM-PRG).
- Murphy, R., Penuel, W., Means, B., Korbak, C., & Whaley, A., & Allen, J. (2001). *E-desk: A review of recent evidence on the effectiveness of discrete educational software*. Menlo Park, CA: SRI International.
- No Child Left Behind Act of 2001, Pub. L. no. 107–110 (2002). Retrieved from:
<http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf>.
- Orr, L. (1999). *Social experiments: Evaluating public programs with experimental methods*. Thousand Oaks, CA: Sage.
- Raudenbush, S.W. & Unlike in Year 4 where there was school-level variation only in Springfield for teachers who met minimum implementation requirements, in Year 5 both districts exhibited school-level variation. In Year 5, both districts also continued the general pattern of decreasing percentages of teachers meeting minimum requirements

- that was first observed from Year 3 to Year 4. Bryk, A.S. (2002). *Hierarchical Linear Models* (Second Edition). Thousand Oaks, CA: Sage Publications.
- Raudenbush, S. W., Martinez, A., & Spybrook, J. (2005, November). *Strategies for improving precision in group-randomized experiments*. *W.T. Grant Foundation's Building Capacity for Evaluating Group-Level Interventions*. Retrieved December 1, 2005 from: www.wtgrant.org
- Raudenbush, S. W., Spybrook, J., Liu, X., & Congdon, R. (2004). *Optimal design for longitudinal and multilevel research: Documentation for the "Optimal Design" software*. Last retrieved January, 2006 from http://sitemaker.umich.edu/group-based/files/od_documentation/chapters1-8_vmarch09-05.pdf
- Redfield, D. (2004, January). *An evaluator's guide to scientifically based research*. Institute for the Advancement of Research in Education at the Appalachia Educational Laboratory (AEL). Retrieved December, 2007 from: <http://www.thefreelibrary.com/An+educator's+guide+to+scientifically+based+research:+documenting...-a0112313352>
- Schmidt, S., & Kingman, M. (2006, November 15). Message posted to Striving Readers electronic mailing list.
- Schochet, P. Z. (2005). *Statistical power for random assignment evaluations of education programs*. Princeton, NJ: Mathematica Policy Research, Inc.
- Scholastic, Inc. (n.d.). *Scholastic reading inventory*. New York, NY: Scholastic, Inc.
- Scholastic, Inc. (n.d.) *Instructional model*. Retrieved October 20, 2007, from Scholastic.com: <http://teacher.scholastic.com/products/READ180/overview/instrmodel.htm>.
- Scholastic Professional Paper. (2006, March). *Because you can't wait until spring: Using the SRI to improve reading performance*. Kimberly A. Knutson, School District of Palm Beach County. New York, NY: Scholastic, Inc.
- Scholastic, Inc. (2005a). *READ 180 America's premier reading intervention program for elementary through high school*. New York, NY: Scholastic, Inc.
- Scholastic, Inc. (2005b). *READ 180 leadership implementation guide: Supporting READ 180 in your district*. New York, NY: Scholastic, Inc.
- Scholastic, Inc. (2005c). *READ 180 participant's guide: Using the SRI and the Lexile Framework effectively with READ 180*. New York, NY: Scholastic, Inc.
- Scholastic, Inc. (2005d). *READ 180 placement, assessment, and reporting guide*. New York,

- NY: Scholastic, Inc.
- Scholastic, Inc. (2005e). *READ 180 teacher implementation guide*. New York, NY: Scholastic, Inc.
- Scholastic, Inc. (2004). *READ 180 research protocols and tools*. New York, NY: Scholastic, Inc.
- Scholastic, Inc. (2002). *Scholastic's READ 180: A heritage of research*. Davidson, J., & Miller, J. (Scholastic Eds.). New York, NY: Scholastic, Inc.
- Sergiovanni, T. (2000). *The lifeworld of leadership: Creating culture, community and personal meaning in our schools*. San Francisco, CA: Jossey-Bass.
- Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton-Mifflin.
- Stanford Diagnostic Reading Test, 4th Edition (SDRT-4) Technical Manual*. Original SDRT4 Author: Bjorn Karlsen and Eric Gardner, 1995. Harcourt Assessments, Inc.
- UMass Donahue Institute (2008, August). *An evaluation of the Commonwealth Pilot School Initiative. Year one report: Intermediate outcomes of the Commonwealth Pilot School Model*. Retrieved November 1, 2008 from: www.doe.mass.edu/research/reports/eval.html
- U.S. Department of Education. (2007a). *Striving Readers: Purpose*. Washington, DC: U.S. Department of Education. Retrieved from: <http://www.ed.gov/programs/strivingreaders/index.html>.
- U.S. Department of Education. (2007b). *Striving Readers: Funding status*. Washington, DC: U.S. Department of Education. Retrieved from: <http://www.ed.gov/programs/strivingreaders/funding.html>
- Vernez, G., & Zimmer, R. (2007, October). *Interpreting the Effects of Title I Supplemental Educational Services*. Santa Monica, CA: RAND. Retrieved January 5, 2006, from www2.ed.gov/rschstat/eval/choice/.../achievementanalysis-sizes.doc
- Waxman, H. C., Connel, M. L., & Gray, J. (2002, December). *A quantitative synthesis of recent research on the effects of teaching and learning with technology on student outcomes*. Naperville, IL: North Central Regional Education Laboratory. Retrieved January 5, 2006, from <http://www.ncrel.org/tech/effects/index.htm>