



An Evaluation of the McGraw-Hill Education Reading Wonders Program

Roisin P. Corcoran, Ph.D.

Jane M. Eisinger, M.S.

Elizabeth Kim, Ed. M

Steven M. Ross, Ph.D.

November, 2015

Contents

Executive Summary	3
An Evaluation of the McGraw-Hill Education Reading Wonders Program.....	6
Program Background	6
Methodology	8
Design	8
Measures	9
Data Analysis	12
Results.....	13
Descriptive Statistics of Continuous Research Variables	13
Program Effects on Vocabulary.....	14
Program Effects on Comprehension	14
Program Effects on GRADE Reading Achievement	14
Teacher Survey Findings	15
Classroom Observation	25
Principal Interviews	28
Conclusions.....	42
Appendix.....	47
References.....	50

Executive Summary

The purpose of the present study was to evaluate the impact of the McGraw-Hill Reading Wonders (RW) program being implemented in 17 elementary schools in geographically diverse school districts in California, Kansas, and Illinois. The program was adopted by the districts due to its alignment with Common Core State Standards (CCSS) and judged potential for improving instructional practices. Its purposes were to address the following research questions:

1. What is the impact of the RW program on fourth grade students' reading achievement after one year of exposure compared to students in non-RW business-as-usual schools?
2. Are there measurable differences in instructional practices, professional development, and beliefs between RW teachers and comparison teachers?
3. How do principals and teachers perceive the effectiveness of RW and the professional development?

Methodology

A quasi-experimental design was used involving 17 elementary schools (13 RW and 4 comparison) in public school districts in California, Kansas, and Illinois. Principals and teachers in the treatment group received all aspects of the intervention, while comparison schools did not receive any RW services or exposure to RW materials prior to completion of the study.

Comparison schools received the program at the end of the study, so this was a delayed treatment group.

Results and Conclusions

Academic achievement findings. Overall, the RW program students were similar to the non-participating students on three types of reading achievement outcomes in Fall 2014 and

Spring 2015. In addition, the changes on various reading achievement outcomes between Fall 2014 and Spring 2015 fall were small for both the RW program participants and non-participants. There was no statistically significant difference in 2015 GRADE outcomes between the RW program students and non-RW program students after controlling for 2014 GRADE and free lunch status, $t(13.22) = -.34, p > .05, \text{pseudo-}R^2 = .077$.

Teacher survey findings. Results from the *Teacher Reaction Survey* indicated that the majority of Reading Wonders teachers agreed that they enjoyed using the program, that the program enhanced their knowledge of reading content, and that they supported the goals of the program. Positive response rates were particularly high for statements regarding student engagement with over 94% of participants reporting that their students used thinking/reasoning skills and strategies with the program. All participants reported that RW was aligned with the CCSS in the areas of reading fluency and understanding key ideas and details of texts. Elements of the program that were rated most frequently as being successful included; technology, lesson plans, grammar lessons, and stories. Participants indicated that a lack of time to make full use of the program, as well as a lack of training were the program's greatest challenges. The majority of respondents (over 85%) would choose to implement Reading Wonders again.

Principal interview findings. In all, analysis of the interview suggests a number of observations regarding the Reading Wonders program.

- Ten out of 10 RW principals interviewed agree that RW is a beneficial program, especially in its alignment with the Common Core and the way it meets the needs of most of its students.
- RW principals overwhelmingly find that students enjoy the program and are challenged cognitively by the activities and content.

- Six out of 10 principals adopted RW as part of a district mandate, but most principals agreed that they would recommend the program to another educator regardless of their initial reason for adopting the program.
- Suggestions for improving the program include bolstering writing components, clarifying the pacing and timing of instruction, differentiating the curriculum for a wider range of student abilities, and providing more comprehensive and on-going professional development for teachers.

An Evaluation of the McGraw-Hill Education Reading Wonders Program

McGraw-Hill Education partnered in 2014 with Johns Hopkins University's School of Education to conduct a third-party evaluation of its K–6 core reading program, Reading Wonders (RW), to gauge the efficacy of the program. Its purposes were to address the following research questions:

1. What is the impact of the RW program on fourth grade students' reading achievement after one year of exposure compared to students in non-RW business-as-usual schools?
2. Are there measurable differences in instructional practices, professional development, and beliefs between RW teachers and comparison teachers?
3. How do principals and teachers perceive the effectiveness of RW and the professional development?

Program Background

McGraw-Hill Education launched Reading Wonders, a K–6 core reading program in January 2013. Reading Wonders, built on the Common Core State Standards for English Language Arts, was designed to promote advanced skills development for learners at all levels. As per the McGraw-Hill website, “The program provides the flexibility of digital implementation and the content for grades K-6 includes Reading/Writing Workshops, Literature Anthologies, Classroom Libraries, and a variety of assessments to measure student performance and mastery.”

Johns Hopkins University, at the request of McGraw-Hill Education, conducted a quasi-experimental evaluation of the McGraw-Hill Reading Wonders (RW) program with 4th-grade students to provide a greater understanding of the program's efficacy. The study involved 17

elementary schools in California, Kansas, and Illinois (13 treatment and 4 comparison) that began the program in Fall 2014.

Significance

Students who experience difficulty reading are at risk for long-term struggles with academic achievement (Neuman & Dickinson, 2003). In fact, achieving reading proficiency by the end of third grade has been established as a key predictor of life success (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996). Still, more than 80% of elementary school children living in high-poverty neighborhoods across America have trouble reading (Slavin, 2014). Adding to the complexity of this issue is the recent transition by most American school systems to the Common Core State Standards (CCSS). Although research has demonstrated the strong potential of the CCSS to eventually improve the performance of U.S. students on international reading metrics such as the PISA exam (Organisation for Economic Co-operation and Development, 2012), the adoption of these standards has dramatically decreased the relevance of core reading programs aligned to less contemporary curriculums—many of which were the predominant method used for delivering reading instruction in many school districts. In turn, schools across the majority of the United States are now pressed to either locate or develop reading materials aligned with these new standards. A recent study conducted by the Center on Education Policy, which surveyed districts from across the US on the CCSS transition, found that almost half of the districts reported that this task poses a major challenge (Rentner & Kober, 2014).

In light of the combination of these factors, the identification of strategies and programs that can be used in the Common Core era to improve reading achievement is urgent. To this end, this study was designed to examine a program with the distinct potential to begin addressing this problem: McGraw-Hill Reading Wonders. McGraw-Hill Education partners globally with

students, educators, administrators, and other professionals to deliver engaging, adaptive, and personalized solutions that improve performance and results. They combine proven, research-based content with the best emerging digital technologies for guiding assessment, teaching, and learning to achieve the best possible outcome for students, instructors, and institutions. McGraw-Hill Education employs more than 6,000 people in 44 countries and publishes in more than 60 languages. McGraw-Hill, the largest publisher of education programs in the United States (*Publishers Weekly*, 2014), combines a variety of research-based instructional approaches and fully aligns with the CCSS.

Methodology

Design

A quasi-experimental design was used involving 17 elementary schools (13 RW and 4 comparison) in public school districts in California, Kansas, and Illinois.

Student study sample. Fourth-grade students attending the RW-treatment schools in fall-2014 comprised the student study sample. Each of the 17 participating schools had between two and four fourth grade classrooms, averaging approximately 24 students per classroom. This yielded a total of 43 classrooms, with 1,037 students in total (739 T and 298 C). These students were followed for one year, through the end of the 2015 school year, when they completed fourth grade. Students who were not enrolled in the study sample at the point of assignment were considered to be joiners and were not included in the analysis sample.

Demographics of student participants. Demographic information of the participants is listed in Table 1. Among the students with complete data, the number of female students was slightly more than male students in both the control and treatment groups. More than half of the participants self-identified themselves as Caucasian or Hispanic in each study group. In addition,

approximately half of them also reported receiving free lunch in both groups. The percentages of students as special education students or English language learners were about 4 and 17, respectively.

Table 1

Participant Demographics (N = 1153)

Variable	Control		Treatment		Total	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Gender						
Female	167	53.02	420	50.12	587	50.91
Male	142	45.08	416	49.64	558	48.40
Missing	6	1.90	2	.24	8	.69
Ethnicity						
Caucasian	128	40.64	257	30.67	385	33.39
African American	13	4.13	139	16.59	152	13.18
Hispanic	65	20.64	193	23.03	258	22.38
Asian	62	19.68	136	16.23	198	17.17
Other	31	9.84	79	9.43	110	9.54
Missing	16	5.08	34	4.06	50	4.34
Free lunch						
Yes	148	46.98	420	50.12	568	49.26
No	154	48.89	337	40.22	491	42.59
Missing	13	4.13	81	9.66	94	8.15
Special education						
Yes	16	5.08	33	3.94	49	4.25
No	282	89.52	739	88.19	1021	88.55
Missing	17	5.40	66	7.87	83	7.20
English language learner						
Yes	55	17.46	143	17.06	198	17.17
No	247	78.41	662	79.00	909	78.84

Missing	13	4.13	33	3.94	46	3.99
Total	315	27.32	838	72.68	1153	100.00

Measures

Academic achievement measure (AA). Group Reading Assessment and Diagnostic Evaluation (GRADE) (Williams, 2001). As described by publishers, “The GRADE is a norm-referenced research based reading assessment which can be group administered. Its wide age range and multiple levels facilitate use with pre-kindergarten children through young adult post-secondary students.” (Williams, 2001, p.1). As further described by Williams, “The GRADE is meant to be a diagnostic tool to see what pre-reading or reading skills individuals have and what skills they need to be taught.” In the outlined study this measure was administered to fourth grade students by their classroom teachers in the fall of 2014 and the spring of 2015.

This measure consists of a battery of 5 subtests: Pre-reading, Reading Readiness, Vocabulary, Comprehension, and Oral Language. In this evaluation, fourth grade students were administered Level 4, Form A, consisting of the following subtests: Vocabulary, Sentence Comprehension Passage Comprehension, and Listening Comprehension.

The measure was standardized and normed during the spring and fall of 2000 using a nationwide sample of over 33,000 pre-school through 12th grade students across 122 sites. Students’ raw scores on the measure can be converted into a variety of standard scores including stanines, percentiles, normal curve equivalents, and grade equivalents. The raw scores can also be converted into growth scale values which “provide a means for tracking reading growth when the student is given different GRADE levels over the years.”

Teacher online survey. In the present study, teachers utilizing the Reading Wonders program were administered a brief online survey via Qualtrics in spring 2015. The survey

consisted of both Likert-type rating items, using a 5-point scale from 1 (*strongly disagree*) to 5 (*strongly agree*), and open-ended questions. Specifically, the survey consisted of items involving fidelity of implementation and intervention, teacher program perceptions, and student engagement. The survey was developed by JHU for the purpose of this study and was designed to require about 20 minutes to complete. The link to the online survey was emailed to teachers at participating schools. Regular reminders were sent by JHU to these participants in order to solicit participation. The response format varied, but typically participants rated their level of agreement.

Classroom organization rubric (COR). Observations occurred in spring, 2015. Each observation consisted of three consecutive 15-minute intervals, followed by a summary section. The checklist, as well as the summary, is divided into four key areas: Instructional Orientation, Instructional Strategies, Behavior Management, and Student Engagement. Two raters independently observed classrooms, and all observations were conducted during the students' language arts class. The observation measure was developed for the purpose of this study and was loosely based on the Classroom Assessment Scoring System (CLASS; Pianta, Karen, Paro, & Hamre, 2008). Observer training was planned and conducted by JHU. The training consisted of a webinar facilitated by JHU via WebEx on April, 1, 2015. This webinar introduced the observers to the COR and covered the contents of the COR Observers' Manual (2014). Training activities included viewing a sample classroom observation video from the METX¹ database, discussion of practice ratings in order to establish a shared understanding of the COR items, scoring guidelines and IRR among the observers, as a follow-up assignment, the observers conducted independent observations of two additional sample videos and submitted their scores for IRR analysis.

¹ <https://tle.soe.umich.edu/MetxInfo>

Observation data was collected by two observers at each of the 17 schools between April 7 and May 29, 2015 - - a total of 34 classrooms. One of these classrooms was observed for 30 minutes (two 15-minute intervals). The rest were observed for 45 minutes (three 15-minute intervals). The two observers conferred after each observation to hand calculate percent agreement and resolve any large discrepancies in scoring. Data was entered by each observer into Excel for statistical analysis.

Inter-rater reliability. Raters indicated the frequency with which each item was seen on a scale of 0 (Absent/Not Observed), 2 (Occasionally), 3 (Frequently), and 4 (Extensively.) This Likert scale was treated as categorical for inter-rater reliability analyses and as a continuous variable for the two-sample t-tests. The inter-rater reliability for the observer training data was calculated by hand using the ratings from a master coder as anchor scores. Items that were scored within 1 point of the anchor score were considered as agreement. Percent agreement for the total number of items (36) was calculated for each paired observation using a minimum threshold of 80 percent. Overall percent agreement on the training ratings was 82. The inter-rater reliability was calculated for the complete data set using STATA. The results of this analysis indicate that the target of 80 percent agreement was met for each pooled² item-type across all observations. The kappa coefficient for the items ranges from 0.43 to 1.00.

Principal interviews. Interviews with principals were conducted during spring, 2015. The interview protocol focused on three main areas: the principal's perception of the RW program, the school's use of the program, and the student response to the program. Interviews were approximately 20–30 minutes in duration.

Data Analysis

² By pooled item-type, we mean that the items were pooled for each classroom across the 15-minute intervals.

Multilevel modeling. Two-level multilevel models (Heck, Thomas, & Tabata, 2014; Norusis, 2012) with one of the three GRADE reading achievement scores in Spring 2015 (i.e., Vocabulary, Total Comprehension, & Total GRADE) as the dependent variable, the same type of GRADE reading achievement scores in Fall 2014 and free lunch status as the level-1 control variables, and Reading Wonders (RW) program status as the level-2 independent variable were fitted with the SPSS linear mixed models routine to estimate the effect of the RW program. The use of a multilevel model would result in more accurate significance test results (Heck et al.). More details of the multilevel model under study are presented in the Appendix.

Results

Descriptive Statistics of Continuous Research Variables

The descriptive statistics of continuous research variables are presented in Table 2. Overall, the RW program students were similar to the non-participating students on three types of reading achievement outcomes in fall, 2014 and spring, 2015. In addition, the changes on various reading achievement outcomes between fall, 2014 and spring, 2015 were small for both the RW program participants and non-participants.

Table 2

Descriptive Statistics of Continuous Research Variables

Variable	Control			Treatment			Total		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
Vocabulary 14	105.15	13.32	299	105.33	14.27	795	105.28	14.01	1094

Vocabulary 15	106.35	14.82	292	106.70	16.25	711	106.60	15.87	1003
Comprehension 14	102.01	14.93	298	103.23	14.52	793	102.89	14.63	1091
Comprehension 15	103.84	14.85	292	103.76	16.35	709	103.78	15.92	1001
GRADE 14	102.83	14.50	299	103.71	14.72	796	103.47	14.65	1095
GRADE 15	104.74	14.41	292	104.69	16.25	713	104.71	15.73	1005

Program Effects on Vocabulary

In the multilevel model, there was no statistically significant difference in 2015 Vocabulary outcomes between the RW program students and non-RW program students after controlling for 2014 Vocabulary and free lunch status, $t(12.96) = -1.04, p > .05$, pseudo- $R^2 = -.002$. The adjusted 2015 Vocabulary group mean scores were 106.92 for the RW students and 105.65.75 for the non-RW students, respectively. The changes in unadjusted Vocabulary means scores between Fall 2014 and Spring 2015 was similar for the RW students (i.e., 1.37) and the non-RW students (i.e., 1.20).

Program Effects on Comprehension

The results did not support a statistically significant difference in 2015 Comprehension outcomes between the RW program students and non-RW program students after controlling for 2014 Comprehension and free lunch status, $t(12.63) = -.04, p > .05$, pseudo- $R^2 = -.098$. The adjusted group mean 2015 Comprehension scores were 104.71 for the RW program students and 104.65.78 for the non-RW program students, respectively. From fall, 2014 to spring, 2015, the RW program students improved by .53 and the non-RW program students by 1.83.

Program Effects on GRADE Reading Achievement

There was no statistically significant difference in 2015 GRADE outcomes between the RW program students and non-RW program students after controlling for 2014 GRADE and free lunch status, $t(13.22) = -.34, p > .05$, pseudo- $R^2 = -.077$. The adjusted group mean 2015 GRADE scores were 105.28 for the RW program students and 104.77 for the non-RW program students, respectively. The RW program students improved by .98 and the non-RW program students by 1.91 between fall, 2014 and spring, 2015.

Teacher Survey Findings

Participants. A total of 26 teachers, 17 implementing RW and 8 implementing a comparison language arts curriculum completed the survey. One additional participant indicated that she was implementing both RW and a different program; since this individual did not specify which program she was rating in her responses her data is not included in this report. The following programs were being implemented by the comparison participants who responded to a question asking them to name their language arts program (N = 7); Houghton Mifflin (37.5%), a teacher-developed program (25%), trade books and NewsELA (12.5%). Of the 26 respondents, 80.8% were female, and 19.2% were male. They represented 11 RW elementary schools and four comparison elementary schools. In the first section of the survey, participants were asked questions regarding frequency of use of a language arts program in their classroom. First they were asked to indicate how often they used the program in general with reported frequency of use. The majority of teachers indicated that they used their perspective program at least three times a week, with the full results being illustrated in Table 3.

Table 3:

In the past month, how often did you implement Reading Wonders in your classroom?

Feature	Every day, for more than 50% of instructional time	Every day, for less than 50% of instructional time	3-4 times a week	1-2 times a week	Less than once a week	N	M	SD
RW	35.3%	52.9%	11.8%	0.0%	0.0%	17	4.24	0.66
Comparison	50.0%	25.0%	12.5%	12.5%	0.0%	8	4.13	1.13

Next, RW participants were asked how often they utilized specific features of RW in their classroom. Over 75% of respondents reported using the Reading/Writing Workshop at least three days a week. Response data can be seen in Table 4.

Table 4:

In the past month, how often did you use the following features of Reading Wonders?

Statement	Never	1-2 days a week	3-4 days a week	Everyday	N	M	SD
Reading/Writing workshop	0.0%	23.5%	47.1%	29.4%	17	3.06	0.75
Literature anthology	0.0%	47.1%	47.1%	5.9%	17	2.59	0.62
Leveled readers	29.4%	35.3%	29.4%	5.9%	17	2.12	0.93

Use and reaction to Reading Wonders. RW participants were asked to indicate their level of agreement with statements about aspects of RW. Participants rated their level of agreement using a five-point Likert-type scale from 1 (*strongly disagree*) to 5 (*strongly agree*). Response data can be seen in Table 5.

The majority of participants either agreed or strongly agreed with the following two statements:

- I follow the steps on how to use this program ($M = 3.94$, $SD = 0.83$)

- I feel like I am able to conduct the lessons effectively ($M = 4.12$, $SD = 0.78$)

However, 41.2% of participants reported that they disagreed with the statement reading “I have sufficient training on how to fit this program into my lessons.”

Table 5:

Rate the extent to which you agree or disagree that each statement describes your experience with Reading Wonders.

Statement	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree	N	M	SD
I follow the steps on how to use this program.	0.0%	5.9%	17.6%	52.9%	23.5%	17	3.94	0.83
I have a hard time teaching parts of this program.	11.8%	35.3%	11.8%	29.4%	11.8%	17	2.94	1.30
I feel like I am able to conduct the lessons effectively.	0.0%	5.9%	5.9%	58.8%	29.4%	17	4.12	0.78
I have sufficient training on how to fit this program into my lessons.	0.0%	41.2%	17.6%	29.4%	11.8%	17	3.12	1.11
I find the Reading Wonders professional development videos useful.	0.0%	5.9%	58.8%	23.5%	11.8%	17	3.41	0.80
I am confident that I implement this program correctly.	0.0%	0.0%	35.3%	52.9%	11.8%	17	3.76	0.66
I change parts of the program to fit them into my curriculum goals.	0.0%	11.8%	11.8%	52.9%	23.5%	17	3.88	0.93
The technological resources are easy to use.	0.0%	17.6%	11.8%	52.9%	17.6%	17	3.71	0.99

Table 6 outlines a comparison between treatment and control group responses to questions regarding fidelity and professional development.

Table 6:	Comparison Group		Treatment Group		T-C Difference	Std. Error	df	p-value	Effect Size
	Mean	Std. Dev.	Mean	Std. Dev.					
ITEM_1 The following questions ask about your experience in using your primary fourth grade English/lang...	4.11	1.05	4.24	.66	.12	.34	24	.71	.15
ITEM_2_1 I follow the steps on how to use this program.	3.67	.71	3.94	.83	.27	.33	24	.41	.35
ITEM_2_2 I have a hard time teaching parts of this program.	3.11	1.05	2.94	1.30	-.17	.50	24	.74	-.14
ITEM_2_3 I feel like I am able to conduct the lessons effectively.	4.00	.50	4.12	.78	.12	.29	24	.69	.17
ITEM_2_4 I have sufficient training on how to fit this program into my lessons.	3.11	1.27	3.12	1.11	.01	.48	24	.99	.01
ITEM_2_6 I am confident that I implement this program correctly.	3.56	.73	3.76	.66	.21	.28	24	.47	.31
ITEM_2_7 I change parts of the program to fit them into my curriculum goals.	4.44	.53	3.88	.93	-.56	.34	24	.11	-.69

Note: p-value was obtained from independent sample t-test.

Note: For ITEM1 Rating of CONT1 had been reversed in order to align to RW1 original code.

Participants were asked to indicate their level of agreement with statements regarding use of RW during the current school year. Participants rated their level of agreement using a five-

point Likert-type scale from 1 (*strongly disagree*) to 5 (*strongly agree*). Over 87% of participants either agreed or strongly agreed with the following three statements:

- I enjoy teaching this program ($M = 4.06, SD = 0.56$)
- This program enhances my knowledge of reading content ($M = 4.18, SD = 0.53$)
- I support the goals of this program (i.e., to make language arts more focused, coherent, and rigorous) ($M = 4.18, SD = 0.81$)

In addition, over 75% of participants agreed or strongly agreed that the program met the needs of most of their students ($M = 3.88, SD = 0.60$).

Table 7:

To what extent do you agree with the following statements about the program this year?

Statement	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree	N	M	SD
I enjoy teaching this program.	0.0%	0.0%	11.8%	70.6%	17.6%	17	4.06	0.56
This program enhances my knowledge of reading content.	0.0%	0.0%	5.9%	70.6%	23.5%	17	4.18	0.53
This program meets the needs of all or most of my students.	0.0%	0.0%	23.5%	64.7%	11.8%	17	3.88	0.60
I support the goals of this program (i.e., to make language arts more focused, coherent, and rigorous).	0.0%	5.9%	5.9%	52.9%	35.3%	17	4.18	0.81
My principal provides the needed support (e.g., materials, training) for the program to be used properly.	0.0%	23.5%	35.3%	29.4%	11.8%	17	3.29	0.99

Parents are actively involved in this program.	11.8%	35.3%	35.3%	17.6%	0.0%	17	2.59	0.94
Parents are supportive of this program.	0.0%	11.8%	47.1%	35.3%	5.9%	17	3.35	0.79

The next section of the survey asked participants to rate their level of agreement with statements regarding student engagement. Over 82% of respondents either agreed or strongly agreed with each of the four statements in this section, with over 94% agreeing or strongly agreeing that students use thinking/reasoning skills and strategies in the program. Response data can be seen in Table 7.

Table 7:

To what extent do you agree with the following statements?

Statement	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree	N	M	SD
Students seem to enjoy the program.	5.9%	0.0%	11.8%	82.4%	0.0%	17	3.71	0.77
Students are on task for most of the time in the program.	5.9%	0.0%	11.8%	76.5%	5.9%	17	3.76	0.83
Students use thinking/reasoning skills and strategies in the program.	5.9%	0.0%	0.0%	88.2%	5.9%	17	3.88	0.78
The program is effective in engaging students in learning to read.	5.9%	0.0%	11.8%	76.5%	5.9%	17	3.76	0.83

Participants were asked to indicate the extent to which they agreed that RW is aligned with the fourth grade ELA Common Core State Standards in specific areas. All of the

respondents agreed or strongly agreed that RW was aligned with the Common Core in the following two areas:

- Reading Fluency (M = 4.24, SD = 0.44)
- Understanding key ideas and details of texts (M = 4.47, SD = 0.51)

A high level of agreement was seen each of the remaining four areas. Response data can be seen in Table 8.

Table 8:

To what extent do you agree that the reading program is aligned with the fourth grade ELA Common Core State Standards in the following areas?

Statement	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree	N	M	SD
Phonics and word recognition	0.0%	0.0%	17.6%	64.7%	17.6%	17	4.00	0.61
Reading fluency	0.0%	0.0%	0.0%	76.5%	23.5%	17	4.24	0.44
Understanding key ideas and details of texts	0.0%	0.0%	0.0%	52.9%	47.1%	17	4.47	0.51
Learning about craft and structure of texts	0.0%	0.0%	5.9%	58.8%	35.3%	17	4.29	0.59
Integration of knowledge and ideas (e.g., making text connections, comparing/contrasting themes, explaining how author uses evidence, etc.)	0.0%	0.0%	11.8%	35.3%	52.9%	17	4.41	0.71
Accessing a range of reading and levels of text complexity	0.0%	5.9%	5.9%	52.9%	35.3%	17	4.18	0.81

Participants responded to open-ended questions which asked them to indicate what parts of their language arts program had been most successful for them so far, and which had been

most challenging. RW participants provided feedback on multiple parts of the program that they had found to be most successful for them in the classroom. Successful elements of the program that were reported most frequently were:

- Technology
- Lesson Plans
- Grammar lessons
- Stories

Responses included:

I liked having the lessons already put together instead of searching for the different resources. Loved having the lessons available in electronic form making it easier to use with my Smart board. The program ensured that my scholars were receiving common core standards instruction and helped them become critical thinkers.

The lesson plans and technology are easy to use and follow. I appreciate having all materials needed for a lesson at the tips of my fingers with the "launch presentation" feature online.

Comparison participants named a broad range of program components that had proven to be successful for them in the classroom. These varied from activities for students, such as reading novels, writing journals, and doing collaborative work, to materials for teachers, such as a Teacher's manual to aid in differentiated instruction.

When it came to what parts of the program were most challenging, nearly half of the RW respondents reported 'time' as being their greatest hurdle.

Not enough time in the day to get to everything in one week. I have to use two weeks of instruction for one theme to feel like I am getting all areas covered completely.

Time! I LOVE teaching Wonders, but I feel like time is always in low supply. My kids engage in these amazing conversations that I feel like I'm constantly having to cut short to fit in what I need to do each day.

Comparison participants (N = 2, 25%) also reported that ‘time’ was a challenge to them, whether it was finding time to create a lesson plan or to try to differentiate a lesson plan.

The other primary challenge reported by RW respondents was a lack of sufficient training or professional development. Responses included:

Only received one day of professional development in October, and our Curriculum director said we would have another one later on in the year, but we didn't.

Inadequate professional development time.

I feel I need more training on the vocabulary portion and the phonics computer parts.

Comparison participants (N = 3, 37.5%) reported a lack of differentiation or a need for differentiated materials. In addition, some (N = 2, 25%) indicated that their programs were not fully aligned with the Common Core, or that they needed to adjust content to meet the Common Core State Standards. RW Participants were asked if, in their opinion, RW should be implemented next year. Of the 16 responses to this question, 14 (87.5%) were in favor of continuing use of the program. One participant commented:

I feel that the strongest reason it should be implemented each year is that kids are gaining a wealth of knowledge- not just reading, but also science, history, health, hot topics, etc. Kids who used to be so disengaged during reading are excited and constantly

looking ahead in their Wonders books. Additionally, the Common Core alignment means that our kids are beyond ready to meet those standards.

Finally, participants were asked what suggestions they have to improve RW. Suggestions varied widely, ranging from a request for fewer units to one for more instructional options for the below grade-level, struggling reader. Two participants indicated that they would like to move away from having to use as much paper with the assessment portion of the program. Responses included:

The weekly assessments are too long and too costly to copy on the copier machine.

The tests are fine, but the amount of paper they take is ridiculous! What I really want is a version of the tests we can use in the computer lab or on tablets to save on paper.

Teacher survey findings summary. Participants reported, on average, that they followed the steps on how to follow the program ($M = 3.94$, $SD = 0.83$), that they felt that they were able to conduct the lessons effectively ($M = 4.12$, $SD = 0.78$) and that the program enhanced their knowledge of reading content ($M = 4.18$, $SD = 0.53$). In addition, teachers reported, on average, that their students used thinking/reasoning skills and strategies in the program ($M = 3.88$, $SD = 0.78$) and that they believed that the program was aligned with CCSS in understanding of key ideas and details of texts ($M = 4.47$, $SD = 0.51$). Teachers identified the technology and lesson plan components of the program as being particularly successful and indicated that a lack of time to make full use of the program, as well as a lack of training were the program's greatest challenges. The majority of respondents would choose to implement RW again.

Classroom Observations

A comparison of means for the treatment and control groups was conducted by item across all observations using a two-tailed t-test in STATA. The results of this analysis are displayed in Table 9. Statistically significant differences were observed between the treatment and control group means within each of the constructs measured by the COR; instructional orientation, instructional strategies, behavior management, and student engagement.

	Kappa	% Agree	P	Kappa	% Agree	P	Kappa	% Agree	P
	Observation 1			Observation 2			Observation 3		
Instructional Orientation									
Lecture	0.92	97.6%	.00**	1.00	100%	.00**	1.00	100%	.00**
Small Group	0.91	97.6%	.00**	0.92	97.6%	.00**	0.8894	94.7%	.00**
Individual Tutoring	1.00	100%	.00**	0.86	97.6%	.00**	0.65	94.9%	.00**
Instructional Strategies									
Purpose of Lesson Clearly Communicated	0.73	92.9%	.00**	0.59	90.2%	.00**	0.81	94.9%	.00**
Use of high-level questioning strategies	0.74	92.9%	.00**	0.84	95.1%	.00**	0.70	89.7%	.00**
Students receive regular feedback on learning	0.84	97.6%	.00**	0.60	92.7%	.00**	1.00	100%	.00**
Behavior Management									
Rules/expectations appear to be understood	0.77	97.6%	.01*	0.48	92.7%	.03*	1.00	100%	.00**
Behavior management is preventative	0.77	95.2%	.00**	0.86	97.6%	.00**	1.00	100%	.00**
Students receive positive feedback	0.79	92.9%	.00**	0.64	87.8%	.00**	0.68	89.7%	.00**
Student Engagement									
At least 80% consistently on task/engaged	0.73	97.6%	.04*	1.00	100%	.00**	1.00	100%	.00**
Students enjoy the work in class	0.86	97.6%	.00**	0.85	97.6%	.00**	1.00	100%	.00**
Students think deeply about material	1.00	100%	.00**	1.00	100%	.00**	1.00	100%	.00**

*Indicates Cohen's kappa coefficient is significant at .05 level

**Indicates Cohen's kappa coefficient is significant at .01 level

Instructional orientation. Within the construct of instructional orientation, significant differences were observed with regard to the use of lecture and individual

tutoring. In this sample, lecture was used significantly more in the treatment classrooms ($M = 1.86$, $SD = 1.36$), $t(242) = 2.97$, $p = .00$. By contrast, individual tutoring was used significantly more in control classrooms ($M = 0.53$, $SD = 1.02$) than treatment classrooms ($M = 0.23$, $SD = 0.55$), $t(242) = 2.97$, $p = .00$.

Instructional strategies. Within the construct of instructional strategies, significant differences were observed with regard to the communication of the purpose of the lessons and the use of high-level questioning. In this sample, clear communication of the purpose of the lesson was observed significantly more often in the treatment classrooms ($M = 2.32$, $SD = 0.86$) than in the control classrooms ($M = 1.92$, $SD = 1.23$), $t(242) = -2.80$, $p = .01$. Similarly, the use of high-level question techniques was observed significantly more often in treatment classrooms ($M = 2.21$, $SD = 0.98$) than in control classrooms ($M = 1.72$, $SD = 1.19$), $t(242) = -3.23$, $p = .00$.

Behavior management. Within the construct of behavior management, significant differences were observed with regard to the use preventative techniques. In this sample, preventative behavior management was observed significantly more often in the control classrooms ($M = 2.42$, $SD = 0.94$) than in the treatment classrooms ($M = 2.16$, $SD = 0.81$), $t(242) = 2.11$, $p = 0.04$.

Student engagement. Within the construct of student engagement, significant differences were observed with regard to the level of student enjoyment of class work. In this sample, student enjoyment was observed significantly more often in the control classrooms ($M = 2.50$, $SD = 0.80$) than in the treatment classrooms ($M = 2.24$, $SD = 0.77$), $t(240) = 2.32$, $p = .02$.

The remaining items on the COR did not demonstrate significant differences

between the treatment and control groups.

Outcome		Treatment N=186		Control N=66		Estimated Impact				
		Posttest Mean	Std. Dev.	Posttest Mean	Std. Dev.	T-C Difference	Std. Error	df	P-Value	Effect Size
Instructional Orientation	Lecture	3.06	1.13	1.86	1.36	1.20	0.08	242	.00**	0.92
	Small Groups	0.90	1.14	1.20	1.31	-	0.08	242	.08	0.25
	Individual Tutoring	0.23	0.55	0.53	1.02	-	0.05	242	.00**	0.42
Instructional Strategies	Purpose of lesson clearly communicated	2.32	0.86	1.92	1.23	0.40	0.06	242	.01*	0.41
	Use of high-level questioning strategies	2.21	0.98	1.72	1.19	0.49	0.07	242	.00**	0.46
	Students receive regular feedback on learning	2.29	0.80	2.22	1.11	0.07	0.06	242	.06	0.08
Behavior Management	Rules/expectations appear to be understood	3.02	0.74	3.19	0.61	-	0.05	242	.10	0.24
	Behavior management is preventative	2.16	0.81	2.42	0.94	-	0.05	242	.04*	0.30
	Students receive positive feedback	1.14	1.01	1.27	1.24	-	0.07	242	.42	0.12
Student Engagement	At least 80% consistently on task/engaged	2.88	0.73	2.78	0.77	0.10	0.05	240	.38	0.14
	Students enjoy the work in class	2.24	0.77	2.50	0.80	-	0.05	240	.02*	0.33
	Students think deeply about material	2.08	0.73	2.09	0.77	-	0.05	240	.93	0.01

** Indicates significant difference at the 99% confidence level

* Indicates significant difference at the 95% confidence level

The above findings suggest that RW teachers were more likely to use lecture as an instructional orientation, clearly state the purpose of the lesson, and use high-level questioning techniques. By contrast, control teachers were more likely to use individual tutoring as an instructional orientation and preventative behavior management techniques. Student enjoyment of class work also appeared to be higher in control classrooms.

The results of the data analysis presented in this report should be interpreted with caution due to multiple limitations of the data in this sample. Several factors must be taken into

consideration before drawing conclusions based on these results. First, the goals of the RW curriculum should be considered in addition to the degree of alignment between the constructs measured by the items on the COR with that curriculum. Second, the validity of the COR, as a measure of the desired constructs, should be explored. Third, the small number of classrooms in the control group may have produced skewed results; these phenomena might not persist given a larger control group sample. Furthermore, the analysis of the data might be conducted differently depending on the research questions to be addressed by the evaluation study. For example, rather than pooling the data from equivalent items across the three observation intervals for each classroom, data from each 15-minute interval could be analyzed separately to produce finer-grained results within each construct.

Principal Interviews

Principals (10 RW, 3 comparison principals) from the participating schools shared their perceptions of the RW program in response to questions about the following broad topics: (1) principal attitudes towards the RW program, (2) student response to the RW Program (3) decision to implement the program, and (4) suggestions for improvement. Data analysis included systematic coding of interviews using a qualitative data analysis package. The interviews were segmented into “quotations,” or sections of data that were relevant to codes. Codes were thematic groupings that arose in response to the four categories of questions. Not all sections of transcripts were assigned codes and some quotations received multiple or overlapping codes. Codes were created iteratively, with codes created, combined, and deleted based on how well the patterns reflected the data. Subsequently, the frequency with which each code was assigned, or its “groundedness,” was calculated to ascertain which patterns of responses were salient in the

conversations. The first section summarizes RW principal responses to the four broad topics and the following section summarizes the same topics for the comparison principals.

The first portion of the interview probed treatment principals about their perceptions of the RW Program.

Is Reading Wonders beneficial? A question embedded in this first section asked whether principals felt the RW program was beneficial. In response to this question, all 10 (100%) RW principals agreed that the program was beneficial. The opinion that RW helped prepare for the Common Core State Standards was overwhelmingly salient in the interviews, with 9 RW principals (90%) mentioning alignment with the Common Core a reason for RW being a benefit. One principal offered, “but I think Wonder[s] does a good job of honoring the intent of Common Core instead of just slapping a Common Core stamp on something they’ve already done.” (P05)

RW principals found the program to be beneficial because teachers appreciate the support that the program provides with the Common Core. Eight stated that teachers felt more at ease knowing that their curriculum was aligned with the Common Core. In addition, teachers were noted as enjoying teaching the program, finding the program easy to use, and using the program regularly. Principals deferred heavily to how teachers felt about the program to inform their own opinions. One of the principals commented, “I feel it is beneficial, because the teachers feel that way and they’re the front line.”(P09)

A related sentiment was that RW was beneficial, depending on the teacher’s delivery. Comments such as “depends on the teacher” (P08) and “it’s up to how the teacher presents it” (P07) were recurring remarks.

To what degree and how are you and other school administrators involved? Eight out of 10 RW principals indicated that they were not very involved personally with the implementation of the program. Instead, a common trend was to defer to the instructional coach or district leader for more information, from clarification of curricular components to assistance with professional development. One principal noted that “we have a coordinator for the district and she presents the curriculum to the teachers who are piloting. So she’s been kind of the one who has been kind of meeting the teachers and teaching them and PD with the staff piloting” (P03) and another principal similarly stated, “We have an elementary curriculum person at the district office level and she would be the one to talk to about that.” (P09)

One principal expressed that she didn’t “even think that she was invited to them [professional development sessions]” and another stated that while she knew there was professional development for teachers, she herself was not involved:

I did not [attend] as an administrator, but that could be the way my district did it or maybe I wasn’t there that day because I didn’t receive any training in it so I’ve only seen it through the implementations that teachers embed). (P04)

Others who claimed that they were involved admitted that their involvement was minimal and embedded within their daily interaction with teachers. For instance, one principal commented, “I’m not very involved other than my informal walkthroughs and any feedback that they might give me in an informal setting.” (P08)

Only two of the 10 RW principals expressed that they felt involved in the process:

Probably attended more professional development than other administrators in other buildings. I attend most of the PD that my teachers also attend. I probably don’t have as

much knowledge base as my instructional coach about the pieces of Wonder, but I definitely would consider myself pretty knowledgeable and I interact pretty frequently. (P02)

To what degree and how are parents involved? Nine out of 10 RW principals answered that parents were either not involved or were minimally and indirectly involved with the RW program, such as helping with homework or reading to children at night. Principals were not very aware of what had been shared with parents regarding the program, or what parents were doing specifically in regards to the RW program.

Well to be honest with you, I haven't done any assessment or feedback with the parents. I know things are shared with the parents, I know...they've been given information. I don't know to what extent they're actually using the material at home. I know we send books home to read, you know read to the students, but I'm not sure). (P02)

A great deal of uncertainty about what was being done at home pervaded the conversation.

How does the McGraw-Hill program differ from other reading programs? When comparing the RW program with other reading programs, the most common response was its alignment with the Common Core. Five out of 10 RW principals mentioned that the Common Core alignment differentiated RW from others. One principal commented on how prior to RW, much of the onus to raise the rigor of instruction to what Common Core required was on the teacher: "Oh I like it much better. Because like I said those [previous curriculum] weren't at all tied to the Common Core and the levels of questioning were low. So the teachers had to...tried to bump up the levels of complexity themselves." (P05)

The next section of the interview focused on how principals perceived the program impacted students.

To what degree does the program meet the needs of most of your students?

Six out of 10 RW principals agreed that the RW Program met the needs of the students.

But they're not the low, low kids, they're below average up to gifted. And their needs seem to be met. And the teachers haven't said anything lacking and I haven't seen anything...(P07)

I think that it teaches the standards and for where we're at right now its meeting their needs. (P04)

Yet, not all principals were satisfied with the program's ability to meet their students' needs. One principal commented that the curriculum was not rigorous enough to challenge their high achieving students. The principal mentioned that she wanted to see more novels being taught as part of the curriculum noting that:

(T)here's these leveled readers that the teachers have been using for kids that are struggling readers but then I feel like we have a lot of kids that are on the other end, like GT students and things like that...and I haven't seen my teachers using anything to really challenge students on the other end. (P04)

To what degree are students engaged in this program? Nine out of 10 RW principals indicated that students enjoy participating in the program. In fact, the observation that students enjoyed the program was coded 17 times throughout the interviews. In discussing this topic, 3 principals mentioned the high interest stories that were also culturally representative: "I think the stories are high interest and even those that are low-leveled readers have that same high interest."(P02)

Moreover, 3 principals cited that the online components of the program were particularly engaging to students. "When I was watching them, some of the kids were online doing the vocabulary – they love the games. Um they were very into that." (P05)

Not only did principals think students were enjoying the program, but they also perceived students as being more cognitively engaged during instruction. Three principals described the degree of deep thinking and focus that they observed in RW classrooms. One principal described:

We're seeing children on task, we're seeing the teachers are you know doing a lot more of the peer pair-share and group activities with that with Wonders. And I think that that has helped also, because we're trying to get away from the lecture type of school and getting more into the teacher just being the lead teacher and facilitator of learning and the children doing most of the talking within the classroom. (P07)

Another principal commented, "I was watching the discussion, I was so impressed when they were doing the child labor article, or the slides and then going through the reading and then having those discussions because what I see is that it is really going deeper." (P05)

To what degree has the program had a positive impact on student achievement (e.g., grades, test scores)? None of the respondents could answer definitively that RW had had a positive impact on achievement, since they had not received up-to-date data on test scores at the time of interviews. Although no formal assessment data was available, principals could obtain a sense of achievement through informal conversations: "(T)he initial assessments we did for the standards were pretty poor but we have made some growth, like talking with the teachers they [have] definitely seen growth in the kids meeting the standards. Definitely room for improvement though." (P01)

Many schools were also implementing a variety of different strategies and new programs that could be confounded with the RW program, making it difficult to isolate the effect of the program.

I think I would like to say that Wonders has had a huge impact, but we've done a lot of professional development around just better reading strategies that I'd say along with the Wonders curriculum and the professional development around teaching have both made a significant impact for us. (P02)

Finally, despite the lack of information and clear effects, 3 treatment principals were hopeful that RW would help their students perform better on achievement tests.

I think the kids are doing well with it. I think the informational texts, that kind of thing, they're going to need to be prepared for their state tests, so you know the areas that aren't in the California standards that are now in the Common Core, I think the kids in the Wonders program are getting more prepared because it's right there for the teacher).
(P04)

Why did you decide to implement Reading Wonders in your school? For 4 of the RW principals, the decision to implement RW was directed by the district. For 6 principals however, the schools were approached by the district and principals agreed to participate predominantly because they were offered Common Core aligned material.

We were basically offered the materials for free, a teacher training if we participated in the research study, so me as an administrator to take something that is at least aligned with our current standards works better than the old. For me, it was a decision of is this better than what we currently have, which the answer is yes, and so for me that's why.
(P04)

Would you recommend this program to other educators? Six out of the 10 RW principals replied that they would absolutely recommend the program to others. In fact, one of the principals interviewed mentioned that she had already done so:

I would definitely recommend it, and I have. One of my principal associations has a blog and they ask what programs, you know, whether it's math or reading or science; I have expressed that the program that we have adopted was Wonders and that I enjoy it, everything was there and there's nothing I'm not happy with. (P02)

Two principals were more cautious about their recommendations and replied yes, under certain conditions. The principals mentioned:

Yes, I think I would recommend the program you know with a few modifications we've discussed. (P09)

I would recommend it to teachers who are really willing to put their all into it, yes. (P07)

Finally, 2 principals voiced that they could not give a recommendation of the program because it was too early to tell. One commented, "I at this point could not, I would not go and recommend it, but I wouldn't speak to it at all. I don't have enough data or information." (P04)

There were varied suggestions for improving the program. Three RW principals voiced that the writing component was not as developed or integrated into the program as needed, especially in light of new tests like the SBAC and PARCC assessments that require extensive writing:

The writing components, I would really like to take a look at those, because we have the SBAC, we've been looking into the practice assessments for the SBAC. The rigor of the writing and stepping out to the district or the school so we would like to see those definitely tied in. (P03)

As aforementioned, many cited the differentiated and comprehensive components of the RW program as a point of praise. Yet others found that the amount of material and activities

included could be overwhelming. Three principals commented that the pacing and timing of the program needed more clarity. One principal stated: “They [teachers] do feel there’s too much to do in one week, so they’re spending two weeks on a story, so that the pacing isn’t...they didn’t feel comfortable with the pacing in the sense of wanting to go deeper with Common Core.”

(P08)

Others pointed to the need for more differentiation of the materials to incorporate high-achieving and severely struggling students. The comments of a principal of a high achieving school illuminates the needs of such a student population: “So it’s kind of good for the low-achieving, struggling students, but not so many materials for the high-achieving students.” (P04) In contrast, the comments of a principal from a school of more diverse abilities highlights the need for stronger support: “What we would like to see is a little more differentiation I guess for RTI, so response to intervention.” (P03)

Finally, 4 intervention principals mentioned the need for more professional development. Principals perceived that there were numerous components to the RW program and that the brief professional development sessions did not suffice. A principal expressed her frustration:

That was part of the problem too was it was they didn’t feel...it was like a real basic one, so that’s why it took them so long...they felt like they had to go back and totally spend their whole vacation and weekends trying to find pieces of stuff. (P05).

Comparison principals. Similar questions were asked of 3 comparison group principals. Two of the comparison schools used a reading curriculum that had been mandated by the district (P12, P13), while one of the comparison schools used a combination of chapter books and supplementary reading curricula (P11). The following section mirrors the analysis of the RW principals’ comments. Generally, the 3 principals showed a consistent trend: 1 principal (P11)

was optimistic and satisfied with his current reading program, another (P12) was ambivalent, and the last principal (P13) expressed negative perspectives towards the current program at his school.

Is the current reading program beneficial? Two comparison principals used Houghton Mifflin. One said it was “outdated” (P13) and the other said they needed a lot of supplementing for the Common Core, commenting:

The program we’ve had probably for 14 to 15 years. The assessment piece never good you know from the....test to what we had adopted, which was more standards-based aligned assessment. It’s always been a struggle. Never really gave us a true continuation of how students were performing on the ESP based on the embedded assessment. (P13)

Another comparison principal used chapter books and Macmillan/McGraw-Hill’s Treasures for the second year. He found that this program was “going pretty well” because teachers have more flexibility in terms of what they want to do.

To what degree and how are you and other school administrators involved?

None of the comparison principals expressed extensive involvement with the program. Two of the principals mentioned their involvement with professional development (P11, P12) of new adoptions of curricula or Common Core preparation. One principal took a more hands-off approach, viewing the reading program as just a tool and that he does not “do feedback about the program itself when I’m doing my observations.” (P13)

To what degree and how are parents involved? Two out of the 3 comparison principals stated that parents were involved minimally or not at all. One principal commented that he did not think that parents “have any interaction with it [reading program],” (P13) while 1

principal mentioned that there were opportunities for involvement but that this was variable based on the parent:

I think they're involved in their child's education at home, so they're seeing aspects of the program in work that their children are doing at home. That's something that is talked about during parent conference time. There are some electronic components that kids can access from home, so I think depending on the level of the individual parent and their involvement, that sort of dictates how much, you know, involvement they have with the program itself. (P12)

How does the current reading program differ from other reading programs?

All 3 comparison principals had used the curriculum from McGraw-Hill's Open Court program before and mentioned the Open Court program as a point of comparison to their current programs. However, they expressed differing opinions about the program. One principal (P11) disliked the prescriptive nature of Open Court and indicated teacher discontent towards the program.

Another principal found that the previous Open Court material provided a better foundation in phonics and initial reading instruction than what their current program offered. However, the principal expressed more ambivalent feelings toward both the previous and current programs, attributing more of the work to the teachers:

I'm almost at 30 years in being in public education and you know there are pluses and minuses to any program out there and you know none of them again are the curriculum. They are resources to help teachers teach the curriculum, and so you know they're all going to have some pluses and they're all going to have some minuses. (P12)

Finally, the third principal found that the Open Court curriculum his school adopted before the current one was very structured and worksheet based. In regards to comparing the current program with Open Court, he stated, "I'm not impressed with either program." (P13)

To what degree does the program meet the needs of most of your students? The trend of satisfaction, ambivalence, and dissatisfaction with the programs were clearest in responses to this question. Namely, 1 principal found that generally the program was meeting their needs, although not completely: "I think so far meets it pretty well. They're having a little bit of difficulty aligning...so they use Treasures, the grammar and language development, and so then that isn't always aligned with what's in the chapter books." (P11)

In contrast, another principal found that there are stronger areas of the program and weaker areas because it was an outdated curriculum. He expressed:

I think because our adoption is rather old now, I don't think it was as strong in that area as the newer programs are, and I think there's a lot more intervention in those types of programs to the newer program than what used to be there. So, you know, it meets our needs to some degree, but I think it could be stronger in terms of providing intervention programs. (P12)

The third principal flatly replied, "*it doesn't*" (P13) in response to the question, citing the mismatch with Common Core standards, the lack of high interest and nonfiction texts, weak grammar and mechanics pieces that were inadequately connected with the reading as reasons for why the current program was not meeting their needs. The principal seemed keenly aware of the standards and skills necessitated by the Common Core and felt that these were not adequately addressed with their current program.

To what degree are students engaged in this program? One principal felt that students were very much engaged with their theme-based, literature based program (P11). In contrast, P12 related student engagement to the teacher not the program. That is, “I don’t think that’s related to the program as much as it is you know what teachers are doing.” (P12) The remaining principal commented that students were engaged at a “*very low level.*” (P13)

To what degree has the program had a positive impact on student achievement (e.g., grades, test scores)? One comparison principal was very optimistic about the results of the program: “I think it’s had a positive effect on national testing. We use NWEA and have positive results on that, we use the CST, and expect positive effects on the Smarter Balance. Students will be taking that this spring.” (P11)

Another comparison principal expressed that they do not have district wide assessments and that the lack of consistent tests makes it difficult to ascertain program effects. “We’re monitoring student progress throughout the school year but in terms of can I say kids have made more progress this year than they did in the past, it’s very hard to do that because we don’t have like data to compare it to.” (P12) He also added that teachers were implementing a number of different programs and strategies in addition to the reading program, making it difficult to isolate the effect of the reading program.

The last principal answered “no” (P13) to this question. Like the previous principal, his reading teachers picked and chose from a variety of programs such as *I Excel* and *Accelerated Reader*. The teachers seemed to avoid using the reading program, instead choosing to take parts from other curricula.

Why did you decide to implement the reading program in your school? For 2 of the comparison principals (P12 & P13), the reading program was a result of a district-wide decision

that was made many years ago. For the principal who was using a combination of chapter books and supplemental materials (P11), the decision was made with teachers who were the main drivers of curricular decisions together.

Would you recommend this program to other educators? One of the 3 comparison principals said he would recommend the school's current program, which consisted of chapter books and the Treasures series. He cited the fact that students could enjoy literature and experience it on many levels as a reason for his recommendation. Additionally, he commented on the flexibility of the program: "We've had positive results so I would recommend it to any other teacher or school that has their group of very engaged teachers who really would develop a lesson and curriculum related to chapter books, not a prepackaged kind of program." (P11)

In contrast, the other two comparison principals did not recommend their program, the Houghton Mifflin curriculum. One principal mentioned that while he would not recommend the current program he was using, he would recommend the program's publisher in general because of the support they received in training and adopting the program. The third principal expressed a less optimistic opinion: "It's just old. You know, the next standards and Common Core I don't think it's meeting that and where we need to be with the Common Core."(P13)

Suggestions for improvement. All comparison principals mentioned the difficulty of finding alignment with the new standards. One comparison principal commented,

(T)he selections are short, not a lot of nonfiction text, which is you know the direction that the Common Core [is going] ...a lot of it was designed to be worksheet based sort of workbook based, busy work. It wasn't really personal reading we needed for student performance. (P13)

The principals who used the district mandated curriculum expressed frustration in that they felt the curriculum was outdated: “You know it’s [the district mandated curriculum] so old...but obviously what would need to happen, or my recommendations for this particular series, is that it needs to be more Common Core aligned.” (P12)

Even for principals that had more positive views towards their current reading program, integrating different pieces was in particular challenging.

I think what I said, having materials that will allow teachers to differentiate for their advanced learners, their struggling learners, their English language learners, and have those things easily embedded in the program so that they don’t have to have like 20 different pieces they’re trying to pull together and where to use them and where not to use them, that kind of thing. (P12)

Another principal wanted to give teachers more time to develop ancillary type activities that would accompany chapter books related to language and grammar. Comprehensive reading instruction requires a number of components including rich literature, grammar, vocabulary, and writing, and the comparison group principals agreed that their current programs needed improvement in integrating these pieces cohesively and ensuring that they were Common Core appropriate.

Conclusions

The present quasi-experimental study, involving 17 elementary schools (13 RW and 4 comparison) in three states, was conducted to evaluate the Reading Wonders Program with regard to the following research questions:

1. What is the impact of the RW program on fourth grade students’ reading achievement after one year of exposure compared to students in non-RW business-as-usual schools?

2. Are there measurable differences in instructional practices, professional development, and beliefs between RW teachers and comparison teachers?
3. How do principals and teachers perceive the effectiveness of RW and the professional development?

As a capsule of evidence obtained from multiple measures (student reading achievement, teacher survey, principal interview, classroom observations), our findings were inconclusive regarding program effects on student reading performance but generally quite positive for a first-year initiative regarding program implementation quality and particularly, the reactions and experiences of teachers and principals in participating schools.

Focusing first on student achievement, quantitative data analyses revealed that there were comparable performances and no statistically significant differences in 2015 GRADE outcomes between the RW program students and non-RW program students after controlling for 2014 GRADE and free lunch status. These results, while failing to support achievement benefits of RW, need to be interpreted relative to contextual conditions of the rollout and associated limitations of the study. Specifically, this was only the first year of the RW implementation for the participating teachers and schools. In many cases, full participation and usage of the materials and curriculum did not commence until mid-Fall semester, therefore restricting the amount of time for teachers to become familiar with the instructional strategies and employ them prior to posttest administration the following spring. If the schools were followed over a second year, a more conclusive test of RW's impacts in raising student achievement could be conducted.

Teaching Practices, Beliefs, and Attitudes

Observation and participant perception findings address Evaluation Questions 2 and 3. During the present start-up year, program implementation seemed fairly strong. Over 80% of

the teachers reported that they were able to follow the steps of the program and teach the lessons effectively. Principals also voiced strong support of the program in general and the progress made in the first year. While observations were limited in number and overall classroom time, those results further supported implementation fidelity by showing conformity of selected teaching activities to expected RW practices. Specifically, relative to comparison classes, teachers more frequently presented lesson-aligned content, clearly stated the purpose of the lesson, and used high-level questioning techniques.

However, perhaps due to the late and hurried start at many schools, less than 50% of the teachers felt that they were adequately trained to integrate the program into lessons or that the professional development videos were useful. Further, relative to the comparison group, the RW teachers were less likely to use the program “every day—for more than 50% of the instructional time.” These findings and the educator perception outcomes reviewed below collectively depict a start-up year in which teachers were receptive to and positive about RW, but tended to use it as less intensively or largely as a supplementary rather than main program than were their counterparts employing existing reading programs. Logically, reduced program “dosage” would operate to limit program potential for impacting reading achievement.

Given these considerations, it nonetheless is auspicious that teachers and principals viewed RW so positively as a means of improving students’ reading achievement. Highly telling is that over 80% of the teachers either agreed or strongly agreed that:

- they enjoyed teaching the program.
- their knowledge of reading content was enhanced.
- they supported the goals of the program (i.e., to make language arts more focused, coherent, and rigorous).

- students enjoyed the program.
- students were engaged in the processes of learning to read.

In addition, over 75% of participants agreed or strongly agreed that the program met the needs of most of their students.

Again, the picture that emerges is one of high enthusiasm for RW's potential to raise achievement, support for its structured lesson design, but frustration about conditions (hurried or delayed start-up) impeding its optimum use this first year. The following quotes, repeated from the Results section, convey these somewhat discordant feelings:

I liked having the lessons already put together instead of searching for the different resources. Loved having the lessons available in electronic form making it easier to use with my Smart board. The program ensured that my scholars were receiving common core standards instruction and helped them become critical thinkers.

Time! I LOVE teaching Wonders, but I feel like time is always in low supply. My kids engage in these amazing conversations that I feel like I'm constantly having to cut short to fit in what I need to do each day.

Only received one day of professional development in October, and our Curriculum director said we would have another one later on in the year, but we didn't.

The principals predictably were less aware than teachers of the day-to-day challenges of finding time to implement RW as a primary and full curriculum. They were also more reserved, knowing that academic programs require time to take hold, in evaluating RW as effective for raising student achievement. All or nearly all were clearly positive about adopting RW and its future potential to make a positive impact on student learning. Notably, all 10 principals interviewed perceived the RW program as beneficial, with salient appreciation that it is so

strongly aligned with the Common Core State Standards. The majority of principals already felt inclined to recommend it to colleagues and one had already done so.

The teacher and principal respondents also offered suggestions to improve the program for future use in their schools. All or most of the concerns seemed minor and directly addressable, including, for example, the needs for expanded professional development, greater differentiation of lessons for high and low achievers, and increased emphasis on writing.

In conclusion, and as noted throughout this discussion, the evaluation results present an overall supportive picture of the first-year usage of RW at participating schools. Given the challenges of introducing a new program, implementation efforts appeared fairly strong, and the majority of teachers and principals were positive about the program design and its potential to improve reading achievement in future years. The first-year achievement results were inconclusive, but the reduced implementation period, teachers' and students' lack of familiarity with the program, and its partial (supplemental or reduced hours) rather than full use by some teachers conceivably would have operated to attenuate treatment effects. In this sense, a multi-year study is clearly recommended as a stronger, more valid test of the student achievement research question.

Appendix

Multilevel Modeling

In light of the clustered data structure with the student data nested within schools, multilevel modeling was conducted with the SPSS linear mixed models routine (Heck et al., 2014; Norusis, 2012) to address the issue of inaccurate standard error estimates due to the violation of independence assumption in ordinary least-squares regression. In the level-1 (i.e., student level) model, one type of GRADE 2015 reading achievement scores (i.e., Vocabulary, Comprehension total, & GRADE total, one at a time) served as the dependent variable. On the other hand, the same type of GRADE 2014 reading achievement scores and free lunch status were the level-1 covariate. As to the level-2 (school-level) model, Reading Wonders (RW) program status was the focal independent variable for the random level-1 intercept to assess if there was a difference between the RW students and the non-RW students in terms of 2015 reading achievement measured by a particular type of GRADE scores after controlling for 2014 reading achievement and free lunch status. The actual models were specified as the follows:

Level-1 model: $Y_{ij} (\text{Reading 2015}) = \beta_{0j} + \beta_{1j} \text{Reading2014}_{ij} + \beta_{2j} \text{Free_Lunch}_{ij} + \varepsilon_{ij}$

Level-2 model: $\beta_{0j} = \gamma_{00} + \gamma_{01} R_W_Program_j + u_{0j}$

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

$$\beta_{2j} = \gamma_{20}$$

The Restricted Maximum Likelihood (RML) estimation method was adopted to avoid the possible underestimated variance components and then overly liberal hypothesis tests

(Raudenbush & Bryk, 2002). The t test of the fixed-effect coefficient of RW program status examined the possible difference between the RW students and non-RW students in terms of 2015 reading achievement after controlling for 2014 reading achievement and free lunch status. The alpha level was set at .05. The proportion of the reduction in between-group variance after the addition of the level-2 independent variable (Raudenbush & Bryk), RW program status, served as the effect size index for the last entered level-2 independent variable. The above effect size index is also called the pseudo- R^2 and can be negative (Singer & Willet, 2003)

Descriptive statistics of other items (not in T-C comparison)

Statistics

	N		Mean	Std. Deviation
	Valid	Missing		
RW_3_1 RW_4-I enjoy teaching this program.	26	8	4.08	.560
RW_3_2 RW_4-This program enhances my knowledge of reading content.	26	8	3.96	.599
RW_3_3 RW_4-This program meets the needs of all or most of my students.	26	8	3.85	.543
RW_3_4 RW_4-I support the goals of this program (i.e., to make language arts more focused, coherent, and rigorous).	26	8	4.15	.675
RW_3_5 RW_4-My principal provides the needed support (e.g., materials, training) for the program to be used properly.	26	8	3.19	1.132
RW_3_6 RW_4-Parents are actively involved in this program.	26	8	2.42	.902
RW_3_7 RW_4-Parents are supportive of this program.	26	8	3.27	.667
MTH_6_SC_1 RW_5-Students seem to enjoy the program.	26	8	3.88	.711
MTH_6_SC_2 RW_5-Students are on task for most of the time in the program.	26	8	3.88	.711
MTH_6_SC_3 RW_5-Students use thinking/reasoning skills and strategies in the program.	26	8	3.92	.688
MTH_6_SC_5 RW_5-The program is effective in engaging students in learning to read.	26	8	3.81	.749
MTH_7_SC_6 RW_6-Phonics and word recognition	26	8	3.69	.736
MTH_7_SC_7 RW_6-Reading fluency	26	8	4.04	.662
MTH_7_SC_8 RW_6-Understanding key ideas and details of texts	26	8	4.35	.562
MTH_7_SC_9 RW_6-Learning about craft and structure of texts	26	8	3.96	.871
MTH_7_SC_10 RW_6-Integration of knowledge and ideas (e.g., making text connections, comparing/contrasting themes, explaining how author uses evidence, etc.)	26	8	4.12	.952
MTH_7_SC_11 RW_6-Accessing a range of reading and levels of text complexity	26	8	3.77	.992
RW_6_1 RW_2-Reading/Writing workshop	17	17	3.06	.748
RW_6_2 RW_2-Literature anthology	17	17	2.59	.618
RW_6_3 RW_2-Leveled readers	17	17	2.12	.928

References

- Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational psychology*, 88(1), 3.
- Heck, R. H., Thomas, S. L., & Tabata, L. N. (2014). *Multilevel and longitudinal modeling with IBM SPSS* (2nd ed.). New York, NY: Routledge.
- Neuman, S. B., & Dickinson, D. K. (Eds.). (2003). *Handbook of early literacy research* (Vol. 1). Guilford Press.
- Norusis, M. J. (2012). *IBM SPSS Statistics 19 advanced statistical procedures companion*. Upper Saddle River, NJ: Prentice Hall.
- Organisation for Economic Cooperation and Development (2012). PISA 2009 Technical Report. Paris, France: OECD.
- Pianta, R. C., Karen, M., Paro, L., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS) Manual: K-3*. Baltimore, MD: Paul H. Brookes Publishing Company.
- Publishers Weekly* (2014). The world's 56 largest book publishers, 2014. Retrieved from <http://www.publishersweekly.com/pw/by-topic/industry-news/financial-reporting/article/63004-the-world-s-56-largest-book-publishers-2014.html>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Rentner, D.S. & Kober, N. (2014). Common Core State Standards in 2014: District implementation of consortia-developed assessments. Center on Education Policy.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.
- Slavin, R. E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research*.
- Williams, K.T. (2001). Group Reading Assessment and Diagnostic Evaluation (GRADE). Retrieved from <https://www.sedl.org/cgi-bin/mysql/rad.cgi?searchid=217>