PEARSON *WORDS THEIR WAY: WORD STUDY IN ACTION* INTERVENTION EFFICACY STUDY

Final Report

Submitted by:
COBBLESTONE APPLIED RESEARCH & EVALUATION, INC.

Rebecca M. Eddy, Ph.D., H. Todd Ruitman, M.A., Nancy Hankel, M.A.,
Monique H. Matelski, M.A., Mariana Schmalstig, M.A.
With special acknowledgement to Xiaoxia Newton, Ph.D.

October 18, 2011

COBBLESTONE
APPLIED RESEARCH & EVALUATION, INC.

COBBLESTONE APPLIED RESEARCH AND EVALUATION, INC.
2120 Foothill Blvd., Ste. 202 | La Verne, CA 91750 | (800) 971-3891

Acknowledgements

We would like to thank all of the participating schools sites that worked with us to study how *Words Their Way* impacted student performance, especially the intervention teachers and their students, school principals, district staff and administration. We appreciate all of the work that you do and your diligent efforts to make the study a success and your continued work to ensure a quality educational experience for children. Special thanks to all site liaisons that helped to coordinate all study activities. We would also like to thank our entire team at Cobblestone, especially Namrata Mahajan and Natasha Schuman. We also appreciate everyone from Pearson Education including Laura Egan, Deena Meyer, Judy Nelson, Linda Duncan, Madeline Coutts, Halle Smith, Anne Thomasmeyer, Marcy Baughman, and especially Ann Vilcheck. Finally, thanks to our families for their ongoing support.

With gratitude,

Rebecca M. Eddy
H. Todd Ruitman
Nancy A. Hankel
Monique H. Matelski
Mariana Schmalstig

PEARSON *WORDS THEIR WAY: WORD STUDY IN ACTION*
INTERVENTION EFFICACY STUDY
FINAL REPORT

Table of Contents

EXECUTIVE SUMMARY **6**
SECTION ONE: STUDY BACKGROUND, STUDY PURPOSE, AND PROGRAM DESCRIPTION **9**
Efficacy Study Background 9
A Response to Intervention (RTI) Context 9
Study Purpose 10
Program Description 10
Background, Study Purpose and Program Description Summary 11
SECTION TWO: DESCRIPTION OF STUDY DESIGN, SETTING, AND SAMPLE **12**
Study Design 12
Site Selection 12
Site Demographic Characteristics 12
Student Participants 12
Attrition 13
Pretest Equivalence 14
Intervention Teacher Participants 14
Design, Setting, and Sample Summary 16
SECTION THREE: DESCRIPTION OF STUDY PROCEDURES AND MEASURES **17**
Study Procedures 17
Study Incentives 17
Training Activities 17
Data Collection Measures: Curriculum Implementation 17
Weekly Logs 17
Observations 18
Teacher Focus Groups and Individual Interviews 18
Data Collection Measures: Outcomes 18
Standards-based Reading and Language Arts Assessment 19
Student Survey 19
Study Procedures and Measures Summary 20
SECTION FOUR: ASSESSMENT OF CURRICULUM IMPLEMENTION **21**
Description of Program Implementation 21
Fidelity of Implementation Results 21
Classroom Observations 22
Implementation Summary 22
SECTION FIVE: RESULTS RELATED TO STUDENTS' ATTITUDES AND ACHIEVEMENT IN LANGUAGE ARTS **23**
HLM Analyses of Outcome Measures 23
Metropolitan Achievement Tests 8 Sounds and Print 23
Metropolitan Achievement Tests 8 Spelling 25
Summary of HLM Results 26
Additional Analyses of Outcome Measures 26
Performance: Comparing Treatment and Control 26
AIMSweb R-CBM Reading Fluency Assessment Results 27
Summary of Research Findings 29
SECTION SIX: PRODUCT SATISFACTION **31**
Student Satisfaction 31
Teacher Satisfaction 32
Instructional Components 32
Response to Intervention Components 33
Product Satisfaction Summary 33
SECTION SEVEN: DISCUSSION **34**
Efficacy Study Key Findings 34
Intervention teachers and Students Liked Using the Words Their Way *Program* 34
Likeability Was Not Correlated with Achievement 34
Achievement Was Not Related to Student Attitudes about Reading 34
Study Summary and Limitations 35
REFERENCES **36**
APPENDIX A: RESPONSE-TO-INTERVENTION TIERS **37**
APPENDIX B: NCRTI RUBRIC **38**
APPENDIX C: PROGRAM QUALIFICATIONS FOR NCRTI REVIEW **39**
APPENDIX D: NCRTI QUESTIONS AND ANSWERS **40**
APPENDIX E: TEACHER TRAINING DESCRIPTION **42**
APPENDIX F: IMPLEMENTATION GUIDELINES **43**
APPENDIX G: DOSAGE OF TREATMENT IMPLEMENTATION **44**
APPENDIX H: PROGRAM COMPONENTS COMPLETED **45**
APPENDIX I: RANDOM-INTERCEPT MODELS WITH COVARIATES **46**

Tables

Table 1. School Level Demographic Characteristics for Participating Sites 13
Table 2. Sample Sizes for Participating Study Groups 14
Table 3. Sample Sizes for Students in the Program and Control Groups 14
Table 4. Pretest Academic Performance Measures 14
Table 5. Demographic Information for Participating Students 15
Table 6. Summary of Intervention Teacher Characteristics 15
Table 7. Schedule of Study Activities 17
Table 8. Distal Measures 19
Table 9. Proximal Measures 19
Table 10. Student Survey Cronbach's Alpha Reliability Measures 20
Table 11. HLM Results for MAT8 Sounds and Print Scaled Scores 24
Table 12. HLM Results for Treatment Group MAT8 Sounds and Print Scaled Scores 24
Table 13. HLM Results for the MAT8 Spelling Scaled Scores 25
Table 14. HLM Results for the MAT8 Sounds and Print Scaled Scores 25
Table 15. Illustration of the Significant Interaction Effect 26
Table 16. MAT8 Sounds and Print: Treatment and Control Scaled Scores by Grade 27
Table 17. Pretest and Posttest MAT8 Sounds and Print Results for Minority Students 27
Table 18. Pretest vs. Posttest Scores for Students Using *Words Their Way* 28
Table 19. Student Survey Pretest and Posttest Results 29
Table 20. Results for Proximal and Distal Outcome Measures 29
Table 21. Product Satisfaction Ratings of *Words Their Way* Program Components by Grade Level 32

Figures

Figure 1.Conceptual Framework for *Words Their Way: Word Study in Action* 11
Figure 2. Pretest and Posttest MAT8 Sounds and Print: Control vs Treatment 26
Figure 3. Pretest and Posttest MAT8 Sounds and Print Results for Minority Students 27
Figure 4. AIMSweb R-CBM Median Words Read Correct: Raw Score Results 28
Figure 5. AIMSweb R-CBM Median Words Read Correct: Percentile Score Results 28
Figure 6. Student Survey Pretest and Posttest Results 29
Figure 7. Students' Favorability of the *Words Their Way* Program 31
Figure 8. Teacher Feedback for *Words Their Way* Components that Enhance Student Learning 32

PEARSON *WORDS THEIR WAY: WORD STUDY IN ACTION*
INTERVENTION EFFICACY STUDY
FINAL REPORT

Executive Summary

Cobblestone Applied Research & Evaluation, Inc. was hired by Pearson Education to conduct an efficacy study of the *Words Their Way: Word Study in Action* program (developmental model) during the 2010-11 school year. The study focused on improving second and fourth grade intervention students' spelling, phonics, vocabulary, and word study skills in a small group, pull-out format. The study was designed to assess implementation of the curriculum in classrooms, answer research questions related to student reading achievement and attitudes, and to assess product satisfaction from teachers and students. This report describes all study activities and provides results related to the research questions.

Study Description, Design, and Measures

The study design adhered to requirements set forth by the National Center on Response to Intervention (NCRTI); a randomized controlled trial (RCT) design was used where individual students were randomly assigned to either the treatment or control condition. **Implementation measures** were collected to assess the extent to which students and teachers implemented their respective reading intervention programs in their classrooms. **Outcome measures** were administered as pretest and posttest instruments and assessed the impact on student acheivement and attitudes about academic and recreational reading.

Study Sample

Twenty-three intervention teachers across fifteen schools in six states from a combination of suburban and rural areas used the *Words Their Way* program in pull-out intervention classrooms during this efficacy study. Data were analyzed for 257 participating students with complete matched pretest-posttest scores in 54 separate groups; there was 5% attrition for participating students from pretest to posttest. The study sample was primarily Caucasian students who were not diagnosed with any learning disabilities, and most of whom had English as their first language. Intervention teachers taught English/language arts 11.8 years, on average, and most (70%) possessed a Master's level degree.

Study Results

Research Question 1:

Are intervention teachers using Words Their Way *program able to implement the program according to NCRTI guidelines (e.g., minimum of 60 minutes per week in a small-group-pull-out program?*

Answer: The level and quality of implementation varied throughout the study, but treatment teachers generally used the *Words Their Way* program in the intervention sessions with fidelity according to classroom observations, interviews and weekly logs. On average, intervention sessions were longer than expected (more than 100 minutes per week) and teachers covered about 21 sorts, on average, primarily from the *Letter Name* and *Within Word Pattern* books.

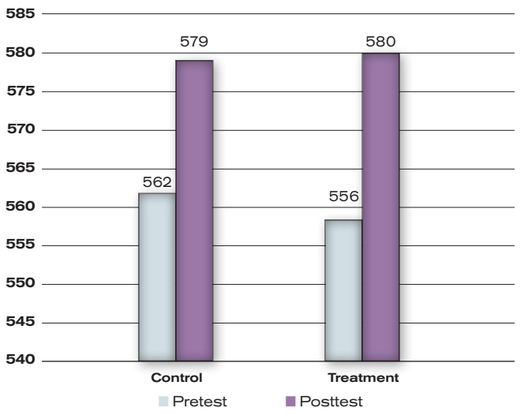
Outcome Measures	
AIMSweb R-CBM	A norm-referenced measure in which students read three passages aloud and a trained administrator recorded the number of words read correctly and the number of errors. The median score of the three passages was used as the overall score and the percentile was calculated from the overall score. This measure was administered at the beginning of the year, once in the middle of the year, and again at the end.
MAT8 – Sounds & Print	A norm-referenced measure of language arts abilities. Sounds and Print measures a student's ability to discriminate sounds and recognize letters, basic words and phoneme-grapheme relationships.
MAT8 - Spelling	A norm-referenced measure of language arts abilities. Spelling measures a student's ability to identify misspelled sight words.
Student attitude survey	Included questions regarding students' enjoyment of recreational and academic reading.
Implementation Measures	
Online logs	Completed by all participating intervention teachers weekly to report the sorts covered and specific components used in the classroom.
Classroom observations	Observed by researchers, all intervention teachers and their students participated in two observations (fall and spring).
Teacher interviews / focus groups	Completed at the end of the study, most intervention teachers participated in individual interviews or focus groups to discuss the program implementation over the duration of the year.

Research Question 2:

How do student reading performance assessments differ for those Tier II students using Words Their Way *compared with those Tier II students using another intervention program?*

Answer: The HLM analysis found that fourth grade treatment students outperformed all other study groups (i.e., fourth grade control, second grade treatment & control). When ignoring other classroom and group effects, the treatment group's increase from pretest to posttest showed that they outperformed the control by having a higher rate of increase. However, this overall effect was not seen in the more conservative HLM analysis. Ethnic minority students in the treatment group outperformed minority students in the control group when ignoring other classroom and group variables.

Pretest and Posttest MAT8 Sounds and Print: Control vs Treatment



Research Question 3:

How does student reading achievement differ from pretest to posttest for those Tier II students using Words Their Way?

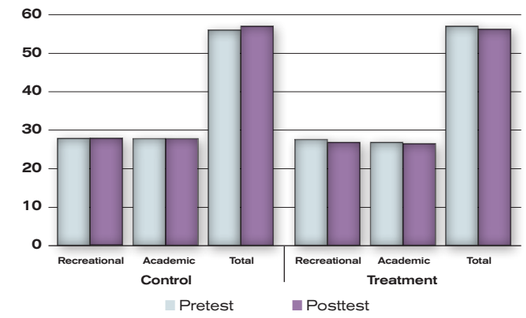
Answer: The treatment group had significant growth from pretest to posttest for the MAT8 Sounds and Print and MAT8 Spelling but not for the AIMSweb R-CBM percentile score. The following table shows scaled scores and grade equivalent scores for the MAT subtests.

Research Question 4:

How do Tier II students using Words Their Way *compare to those Tier II students using another intervention program from pretest to posttest on attitudes about reading?*

Answer: Both treatment and control groups did not show gains from pretest to posttest on the student survey when measuring attitudes about academic and recreational reading. There were no differences between treatment and control at the pretest and posttest administrations of the student survey.

Student Survey Pretest and Posttest Results



Product Satisfaction

Teachers and students reported liking the *Words Their Way* program and activities, especially for second grade students. Some reported that the repetitive nature of the program was helpful to students, whereas, others suggested that the student libraries could be improved. Most students (73%) were at least happy about these specific program components, and many were "very happy"; about 87% of all the students rated Games in the *Words Their Way* books as making them happy or very happy.

Pretest vs. Posttest Scores for Students Using *Words Their Way*

Test	Pretest Mean	Posttest Mean	*n*	*t*	*p-value*
MAT8 Sounds and Print Scaled Score (Grade Equivalent)	556 (1.8)	580 (2.3)	135	11.34	< .001
MAT8Spelling Scaled Score (Grade Equivalent)	561 (2.5)	592 (3.0)	138	10.58	<.001
AIMSweb R-CBM Percentile	25	27	138	1.72	.087

PEARSON *WORDS THEIR WAY: WORD STUDY IN ACTION*
INTERVENTION EFFICACY STUDY
FINAL REPORT

> *"I think I have become a better teacher with word study…"*
> *"They [students] love it. They don't want it to end."*
> *"They look forward to coming. They really like the activities."*
>
> **–Teacher comments feedback on using *Words Their Way***

Conclusions

Overall, the students using *Words their Way* (treatment group) outperformed the control group from pretest to posttest on the MAT8 Sounds and Print subtest when not taking group effects into account. This finding was mainly due to the performance of the fourth grade students and the performance of ethnic minority students in each grade level. Students and teachers liked the various components of *Words their Way* and enjoyed using it as an intervention program although likability did not necessarily translate into higher student scores. The instances when teachers reported not liking the program were more related to coordination of ancillary materials rather than the core program components. Despite the fact that there were gains from pretest to posttest for students using *Words their Way*, they did not show more positive attitudes about either academic or recreational reading during the study according to the student survey. Future studies should examine how *Words their Way* functions as both a whole-group intervention as well as a pull-out intervention to see whether the group structure accounts for performance.

Section One: Study Background, Study Purpose, and Program Description

Cobblestone Applied Research & Evaluation, Inc. was hired by Pearson Education to conduct an efficacy study of the *Words Their Way: Word Study in Action* (herein referred to as *Words Their Way*) program (developmental model) during the 2010-11 school year. The study focused on improving second and fourth grade intervention students' spelling, phonics, vocabulary, and word study skills in a small group, pull-out format. The first section of this report includes a description of the study background, purpose and a review of the main program components for *Words Their Way.*

Efficacy Study Background

In schools throughout the United States, many children struggle to read while in their elementary education. In 2009, a national representative sample of approximately 178,000 fourth-graders took the National Assessment of Educational Progress (NAEP) in reading and, of those fourth-graders, one out of three performed below the "Basic level." Basic level is defined as "partial mastery of prerequisite knowledge and skills that are fundamental for proficient work" (NCES, 2009, p. 5). This statistic is alarming and indicates an immediate need to provide children with educational programs that will improve their reading achievement.

Competent reading and writing skills are essential for successful matriculation through school and entry into the workforce (Apel, 2009). Illiteracy can have long-term repercussions for children's developing self-esteem and motivation to learn, as well as their later academic performance (Savolainen, Ahonen, & Aro, 2008). As such, state and federal governments fund initiatives or provide mandates to increase student success in reading (e.g., No Child Left Behind). Although there are no quick solutions to ameliorate this pervasive problem, extensive research suggests that development of a number of essential skills can aid student literacy. Linguistic factors that have been found to predict children's successful early reading and spelling development include phonemic awareness (Ehri, Nunes, Willows, Schuster, Yaghoub-Zadeh, & Shananhan, 2001), phonological awareness (Bowey, 2005), spelling ability (Ehri & Rosenthal, 2007), and breadth of vocabulary (NRP; National Reading Panel, 2001).

Literacy acquisition centers on learning the relationship between the letters of written language and the sounds of spoken language (Nagy & Scott, 2000). However, children must also learn individual words and their meanings, as well as the spelling of these words (Ehri & Rosenthal, 2007). Only after they attain these basic reading skills can children begin to develop reading fluency, as well as reading comprehension (Perfetti, 1985). Pearson's *Words Their Way: Word Study for Phonics, Vocabulary, and Spelling Instruction, 4th edition* (Bear, Invernizzi, Templeton, & Johnston, 2008) program is a curriculum used to increase student reading fluency and reading comprehension by teaching students fundamental reading skills. *Words Their Way* provides language activities related to phonics, spelling, and vocabulary as well as individualized tutoring in a small group format.

A Response to Intervention (RTI) Context

Within a Response to Intervention (RTI) context, teachers are responsible for identifying and monitoring students' abilities and providing those students with targeted evidence-based interventions depending on students' responsiveness. Burns (2008) writes that "research has consistently found that RTI initiatives lead to gains in student achievement and school wide improvements, such as reduced referrals to and placements in special education and a higher rate of students scoring proficiently on state tests (Burns, Appleton, & Stehouwer, 2005)" (p. 14). Typically, RTI is administered as a three-tier model (see Appendix A).

The National Center on Response to Intervention (NCRTI) evaluates research on RTI interventions through a technical review committee (TRC). The TRC looks specifically at the study participants, study design, fidelity of treatment implementation, and study measures in determining the quality of the research study. For a more in-depth look at the rubric used to evaluate research see Appendix B. The U.S. Department of Education (2004) has emphasized that RTI be supported by researched interventions. Pearson will submit the results of the proposed study of *Words Their Way* to the NCRI TRC. See Appendix C for more information about the qualifications for an intervention to be reviewed by NCRTI. Finally, we have summarized all NCRTI questions, responses to questions, and accompanying page numbers in this report in Appendix D.

This report describes the evaluation of the *Words Their Way* program that was conducted during the 2010-2011 school year. NCRTI questions about the study participants, study design, fidelity of treatment implementation, and

NCRTI Q1 example
How was the program delivered?

other study measures are answered throughout the report. Specifically, NCRTI protocol questions are highlighted in purple text boxes throughout the document. The corresponding answers to these questions are reflected within the text.

Study Purpose

An efficacy study of *Words Their Way* within a Response-to-Intervention (RTI) framework was conducted in six states (California, Indiana, Kentucky, Massachusetts, Michigan, and New Jersey) during the 2010-2011 academic school year. Second and fourth grade Tier II students who required a reading intervention program were recruited to participate in the study in a small-group, pull out format with an intervention teacher for 20 minutes per day, five days per week. For the purpose of this study, a number of student outcomes were assessed. In addition, we systematically tracked curriculum implementation to determine if program implementation also impacted student outcomes. The primary questions motivating the research for the study include the following:

Research Question 1:
Are intervention teachers using Words Their Way *program able to implement the program according to NCRTI guidelines (e.g., minimum of 60 minutes per week in a small-group-pull-out program?*

Research Question 2:
How do student reading performance assessments differ for those Tier II students using Words Their Way *compared with those Tier II students using another intervention program?*

Research Question 3:
How does student reading achievement differ from pretest to posttest for those Tier II students using Words Their Way?

Research Question 4:
How do Tier II students using Words Their Way *compare to those Tier II students using another intervention program from pretest to posttest on attitudes about reading?*

Program Description

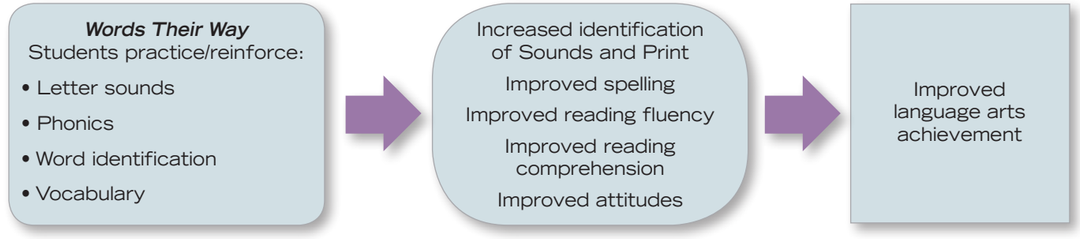
Words Their Way is a powerful approach used to teach students phonics, spelling and vocabulary. The study used the print developmental version (4th edition) of the *Words Their Way* program as the primary source of instruction for intervention teachers and their students. The program focuses on word study using alphabetic (phonics), patterns (spelling), and meaning (vocabulary) to instruct students. There are five student books available which relate to various developmental stages of the program. Each book contains a number of "sorts" used as part of instruction; instructions for implementing lessons are found in the *Developmental Model Teacher Resource Guide.*

- Early Emergent Letter Name: Sort 1 (Concept Sort *Fruit/Not a Fruit*) through Sort 51 (Word families *–an, -un, -in*)
- Letter Name: Sort 1 (Beginning consonants *b, m, r, s*) through Sort 49 (Preconsonantal Nasals *-nt, -nd, -nk*)
- Within Word Pattern: Sort 1 (Short and Long *a* Pictures) through Sort 51 (Short and Long *i Homophones* #2)
- Syllables and Affixes: Sort 1 (*Compound Words*) through Sort 53 (*Homophones*)
- Derivational Relations: Sort 1 (Prefixes: *pre-, fore-, post-, after-*) through Sort 38 (Prefix Assimilation: *Prefixes in-, com-, ad-, sub-*)

Picture and word cards, sorting grids, and game boards direct students to engage closely with the words in order to identify conventions of the English language that are necessary to understand to read and spell. Students receive intervention based on their spelling aptitude, not their grade level. In order to assess the students' developmental spelling stage, students are given a spelling inventory to identify the student book and sort where instruction should be delivered at the appropriate developmental stage.

The *Words Their Way: Word Study in Action* program drives the conceptual model shown below in Figure 1. We hypothesized that students would advance in their reading skills when using *Words Their Way: Word Study in Action*. We expected students to experience these gains because of the combined focus on phonics, spelling, and vocabulary. The success of the materials also depended on the extent to which the curriculum was implemented as intended, and therefore close tracking of curriculum implementation was included as part of the study.

Figure 1. Conceptual Framework for *Words Their Way: Word Study in Action*



Background, Study Purpose and Program Description Summary

An efficacy study of the *Words Their Way* program was conducted during the 2010-11 school year. This program includes numerous components designed to increase student reading fluency and reading comprehension by engaging students and teaching them fundamental reading skills. *Words Their Way* provides language activities related to phonics, spelling, and vocabulary as well as individualized tutoring in a small group format. The study was designed to assess implementation of the curriculum in classrooms, answer research questions related to student reading achievement and attitudes, and to assess product satisfaction from teachers and students.

PEARSON *WORDS THEIR WAY: WORD STUDY IN ACTION*
INTERVENTION EFFICACY STUDY
FINAL REPORT

Section Two: Description of Study Design, Setting, and Sample

The following section provides information on the study design and sample. When necessary, we also provide questions and corresponding answers to questions required for NCTRI review.

Study Design

The *Words Their Way* study was conducted during the 2010-2011 school year.

The efficacy study was designed as a Randomized Controlled Trial (RCT) in which qualified students were randomly assigned to either the treatment group, using the *Words Their Way* program or a control group (using the existing reading intervention program, if any, at their schools or continuing with business as usual). Teachers and their students used their respective language arts programs in their classes for the duration of the 2010-11 school year.

Design Q1
Was random assignment used?

An experimental design was selected because it is well-regarded as the strongest in terms of internal validity (appropriately assigning cause to a particular treatment) while having the highest probability for ruling out alternative explanations of cause (Shadish, Cook, & Campbell, 2002). In addition to collecting information related to program outcomes (e.g., student achievement data), we also collected information related to program implementation, given that varying levels of implementation can have differential impacts on related outcomes (Sechrest, et al., 1979). The study design is also considered a cluster-type design in which a cluster (small group of intervention students) is nested within one intervention teacher, hence allowing analyses to be conducted on multiple levels to more specifically identify potential treatment effects.

Site Selection

Site selection began in June 2010 and continued through the fall of 2010. Cobblestone researchers identified potential sites throughout the United States by selecting specific criteria from districts listed in the National Center for Education Statistics (http://nces.ed.gov/ccd/schoolsearch/ and http://nces.ed.gov/ccd/districtsearch/). Several hundred school districts were contacted through phone and email. It is important to note that schools with diverse student ethnicity and lower socioeconomic status individuals were targeted specifically for inclusion in the study to determine the impact of the program in a variety of settings. Ultimately, a majority of districts that had the most diverse group of students declined to participate in the study. This was not unexpected, as the most diverse districts tend to be concentrated in urban areas where students typically have high mobility, district research protocols are particularly stringent, and numerous competing district initiatives does not allow participation in a research study to be a priority.

Of the schools that met the inclusion criteria, securing their participation occurred through initial contact with teachers or district supervisors. In total, fifteen sites across six states were confirmed for participation in the study. The participating sites were identified and recruited by Cobblestone researchers. All participating teachers, site liaisons, district personnel, and Cobblestone researchers signed a Memorandum of Understanding (MOU) document to formally secure each school's participation.

Site Demographic Characteristics

As indicated earlier, a total of fifteen schools across six states participated in the study. The six states consisted of California, Indiana, Kentucky, Massachusetts, Michigan, and New Jersey. Table 1 provides full detailed information about each site, including school-based and community indicators. The average school had 400 students and a majority of students were identified as Caucasian. An average of 33% of students was eligible for Free or Reduced Lunch. On average, 40% of adults 25 and over living in the community in which the school was located had a college degree and the average median household income was $60,277.

Student Participants

This section describes how participants were selected, sample sizes, and demographic characteristics of student participants. Students were screened and then selected for participation in the study if they met inclusion criteria. First, schools recommended students for screening based on their performance on reading or language arts state standardized test. Schools were asked to specifically not include Tier III students (those requiring individual, intensive interventions) in the screening.

Participants Q1
How were students selected to participate in the study?

Participants Q5
What was randomly assigned?
What unit was used for data analysis?

Table 1. School Level Demographic Characteristics for Participating Sites

State	Site	School Size*	Ethnicity*				Economic Measure*	Community Measure**	
			% Caucasian	% Hispanic/ Latino	% African American	% Other Ethnicity	% Free & Reduced Lunch	% Age 25+ With College Degree	Median House-hold Income
CA	Site 1	369	5%	86%	<1%	8%	57%	21.9%	$49,256
	Site 2	390	18%	64%	2%	16%	55%	83.1%	$47,467
	Site 3	446	45%	24%	3%	28%	15%	83.1%	$47,467
IN	Site 4	453	97%	2%	<1%	<1%	26%	48.9%	$85,829
	Site 5	521	98%	<1%	<1%	<1%	29%	48.9%	$85,829
	Site 6	380	66%	11%	22%	2%	65%	14.9%	$37,234
	Site 7	508	92%	2%	5%	<1%	20%	14.9%	$37,234
	Site 8	514	72%	4%	21%	5%	50%	14.9%	$37,234
KY	Site 9	366	82%	2%	12%	4%	25%	7.7%	$27,374
	Site 10	304	100%	-	-	-	69%	26.9%	$29,047
MA	Site 11	304	91%	4%	2%	2%	10%	42.9%	$80,944
	Site 12	353	84%	5%	5%	6%	10%	42.9%	$80,944
	Site 13	224	78%	9%	7%	6%	21%	42.9%	$80,944
MI	Site 14	409	59%	4%	7%	30%	13%	50.0%	$77,538
NJ	Site 15	425	32%	49%	14%	6%	33%	46.5%	$99,817

**Information obtained from each state's department of education or district websites; ** US Census 2005-2009 Community Survey*

Recommended students then completed the AIMSweb Reading Curriculum- Based Measurement (R-CBM) diagnostic assessment. (A full description of the AIMSweb R-CBM can be found in Section Three of this report.) Once students' AIMSweb R-CBM scores were calculated, they were then selected for participation if they had a "low" pretest score. Scores were considered "low" if they were below the 30th national percentile.

Of those students that qualified for participation, half were randomly assigned to the "treatment" condition (using *Words Their Way*) and half were randomly assigned to the "control" condition (using the existing intervention program, if applicable, or continuing with business as usual). Consequently, random assignment was done at the student level; we also used the intervention teacher cluster as the unit of analysis given that students were nested within intervention teacher at each school.

Students were identified as at risk for academic failure based on their scores on state standardized tests and teacher recommendations.

The program used for the treatment condition was *Words Their Way*. Control students used a variety of other phonics, spelling, and vocabulary programs, depending on the school site. Some control students were considered part of a "true control" group in which they stayed in their regular language arts classroom and did not receive any intervention program that was comparable to *Words Their Way*. However, other students used programs such as *Read Naturally, Passports, PLAID Phonics, Voyager, Literacy Place, Leveled Literacy, and Phonics for Reading*.

Participants Q2
How were students identified as at-risk for academic failure?

Participants Q2a
Clarify which treatment is the submitted program.

Participants Q2b
Clarify which condition is the control condition

The following Table 2 summarizes the sample sizes for each participant type: schools, intervention teachers, classrooms (defined as pull-out intervention groups for the treatment group and corresponding control groups for the control group), and students.

Participants Q3
Please provide the sample sizes for your study, for all types of participants (schools, instructors, classrooms, and students) and relevant conditions (intervention and control).

Attrition

Sample attrition is defined as those students who completed pretests on any of the primary outcome

Table 2. Sample Sizes for Participating Study Groups

Participant Type	Sample Size – Program	Sample Size – Control
Schools	15	
Intervention teachers	23	18
Classrooms	29	25
Students	138	119

Participants Q4
How many program students were pretested? Posttested?
How many control students were pretested? Posttested?

measures (i.e., MAT8 Spelling and MAT8 Sounds and Print), yet did not complete a posttest on any of these measures. The total amount of attrition was only five percent of the original sample. There were 257 participating students that completed at least one matched pretest and posttest. The original pretest sample (i.e., students that completed a pretest measure) included 271 students (135 second grade; 136 fourth grade). The difference between the original sample of students that took any or all of the pretests and the final number of students is 14 (8 control; 6 treatment). Of the 14 students, we were informed by intervention teachers that 11 had moved out of school boundaries and one student chose to drop from the study. The remaining two students were either absent during posttesting or had moved during the school year. Because of the small number of students that dropped from the study, we can be confident that attrition did not affect the results found in this study. Therefore, the students that are considered "participating", as displayed in the Table 3.

Pretest Equivalence

Despite the fact that this was an RCT study where students were randomly assigned to each group, it was important to verify that the groups were equivalent at pretest to not provide an advantage of one group over another. Therefore, pretest

Design Q2
Was the program group comparable to the control group on pretest performance measures?
Was the program group comparable to the control group on demographic variables?

academic performance was compared across program and control groups (see Section Three for a full description of outcome measures). Additionally, demographic variables were compared across groups. Tables 4 and 5 demonstrate comparability of the program and control groups on pretest performance and demographics. Given that each comparison of the treatment and control groups yielded non-significant *p*-values for all academic performance measures and demographic characteristics, comparability between program and control groups should be assumed.

Intervention Teacher Participants

Fidelity of Implementation Q3
What were the background, experience, training, and ongoing support of the instructors?

There were a total of 23 intervention treatment teachers who participated in the study, four of whom had intervention groups at both the second and fourth grade levels. Intervention teachers' number of years teaching reading ranged from 0 to 36 with an average of 11.8 years. The average number of years teachers taught at the K-12 level was 15.3, which ranged from 1

Table 3. Sample Sizes for Students in the Program and Control Groups

Participant Type	Pretest	Posttest
Program students	144	138
Control students	127	119

Table 4. Pretest Academic Performance Measures

Measures (Name)	Program (*n*=144)		Control (*n*=127)		*p*
	Mean	Standard Deviation	Mean	Standard Deviation	
AIMSweb R-CBM	25.0	12.4	24.7	12.6	.876
MAT8 Sounds & Print	555.8	35.8	560.7	44.9	.310
MAT8 Spelling	561.8	51.7	555.5	53.5	.321

Table 5. Demographic Information for Participating Students

	Program		Control		*p* of chi square
	Number	Percentage	Number	Percentage	
Grade level					
Grade 2	70	51	58	49	.751
Grade 4	68	49	61	51	
Race-ethnicity					
African-American	13	9	14	12	.597
American Indian	1	1	1	1	
Asian/Pacific Islander	2	1	4	3	
Hispanic	39	28	23	19	
White	81	59	75	63	
Other	2	1	2	2	
Socioeconomic status					
Subsidized lunch	63	46	59	50	.530
No subsidized lunch	75	54	60	50	
Disability status					
Speech-language impairments	6	4	2	2	.086
Learning disabilities	1	1	4	3	
Behavior disorders	0	0	0	0	
Mental retardation	0	0	0	0	
Other	2	1	0	0	
Not identified with a disability	126	91	113	95	
Multiple	3	2	0	0	
ELL status					
English language learner	35	25	40	35	.112
Not English language learner	103	75	76	66	
Gender					
Female	64	46	53	45	.768
Male	74	54	66	55	

Table 6. Summary of Intervention Teacher Characteristics

Highest Degree Obtained				Teaching Experience	
Associates	Bachelor of Arts/ Science	Teaching Credential/ Certificate	Master of Arts/ Science	Number of years teaching (average)	Number of years teaching English (average)
2 (9%)	4 (17%)	1 (4%)	16 (70%)	15.3	11.8

to 36 years. Over two-thirds of the intervention teachers held master's degrees. Table 6 summarizes teacher characteristics of teaching experience and education level. Additional information about training and support can be found in Section Three.

Design, Setting, and Sample Summary

Twenty-three intervention teachers across fifteen schools in six states from a combination of suburban and rural areas used the *Words Their Way* program in pull-out intervention classrooms during this efficacy study. Data were analyzed for 257 participating students with complete matched pretest-posttest scores in 54 separate groups; there was 5% attrition for participating students from pretest to posttest. The study sample was primarily Caucasian students who were not diagnosed with any learning disabilities, and most of whom had English as their first language. Teachers taught English/language arts 11.8 years, on average, and most (70%) possessed a Master's level degree.

Section Three: Description of Study Procedures and Measures

This section includes a description of study procedures and implementation and outcome measures used in the study. We also describe other aspects of the study including compensation and training.

Study Procedures

Once sites were recruited to participate in the study, school administrators identified students that were in need of reading intervention. One or more individuals at each school site were appointed to administer the AIMSweb R-CBM assessment (described in the next section) to determine which students would be eligible to participate. After eligible students were identified based on their AIMSweb R-CBM score, students were randomly assigned by researchers to either the program/treatment or control group. All treatment and control group students then completed pretesting for all outcome measures (and again at the end of the year for posttesting). Intervention teachers received training on the *Words Their Way* program and began using the program with treatment students after all pretesting was completed. The groups used their respective programs for most of the 2010-11 school year until just prior to posttesting.

Study Incentives

Compensation for participation in the study was a $300 cash stipend for intervention teachers. In addition, all treatment intervention teachers received free *Words Their Way* materials as well as free training and product orientation for all intervention teachers. Each participating school received enough materials for the participating students in the treatment group, and received the balance of materials (equivalent to the number of control students) at the end of the study.

Training Activities

A summary of study activities and corresponding dates can be found in Table 7. Treatment intervention teachers used the *Words Their Way* program during the course of the 2010-11 school year. Noting that not all schools began and completed the school year at the same time, the sequence of study activities was similar across all sites. A detailed description of teacher training is documented in more detail in Appendix E.

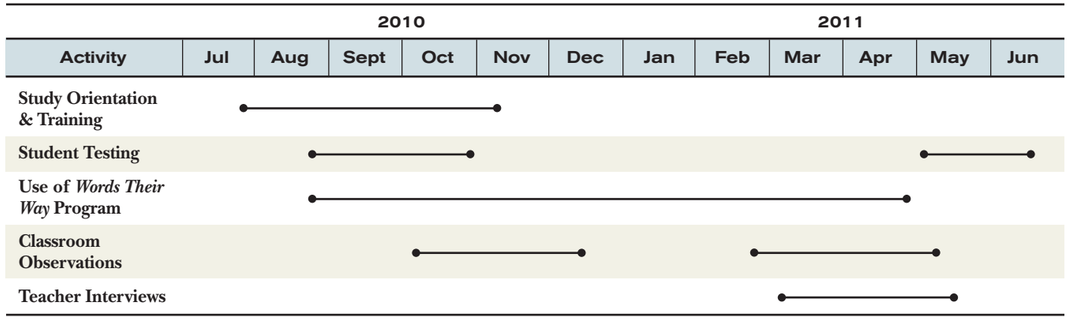
Data Collection Measures: Curriculum Implementation

Implementation measures were developed to monitor and assess the activities in participating classrooms throughout the year. Implementation measures included weekly teacher logs, classroom observations, and intervention teacher interviews / focus groups. Teachers also communicated informally with the researchers via email, phone, open-ended sections of the teacher logs, and during informal interviews conducted during the fall and spring classroom observations. Intervention teachers were expected to use the program 20 minutes per day for five days per week. Treatment intervention teachers were continuously encouraged to provide feedback about the *Words Their Way* program throughout the study.

Weekly Logs

Each week treatment intervention teachers were required to complete online logs that addressed which book was used, the sort the group completed, the daily components of the program covered, materials used, and homework assigned. In addition, teacher logs were useful as a source of teacher reflection on their own practice

Table 7. Schedule of Study Activities



or for providing informal feedback regarding use of the products or other issues with program implementation. The primary goals in utilizing an online teacher log system were:

- Capture as accurately as possible both the *Words Their Way* content covered in intervention periods.
- Allow intervention teachers to report any activities or events whether at school or within the classroom that might have impacted their teaching or student learning.
- Reduce the strain on intervention teachers by making the process user-friendly and efficient.
- Collect data in a way that was meaningful to researchers and could be reported back easily.

Observations

All intervention groups were observed by at least one member of the research team on two occasions during the 2010-2011 school year. One member of the research team served as the primary observer, while other members of the research team conducting observations were required to calibrate scores based on live observation sessions with the primary observer. To calibrate multiple observer ratings, after each classroom observation, the raters discussed scores provided for the protocol. After carefully reviewing the scoring rubric, most teacher rating scores remained within a 1-point difference, and we ultimately established inter-rater agreement above 95%. Once sufficient inter-rater agreement was established, additional observations proceeded at all study sites. The observation protocol used included descriptive information such as which book and sort comprised the lesson, types of sorts used, student engagement, classroom management, and student and teacher rapport. This protocol, along with the weekly logs and focus groups / interviews allowed researchers to understand the activities that occurred in participating classrooms throughout the efficacy study. Scores were compiled per teacher and the average score became the associated variable for that teacher in subsequent analysis.

Teacher Focus Groups and Individual Interviews

Lastly, brief interviews (or focus groups with multiple intervention teachers, where possible) were conducted toward the end of the study with each treatment intervention teacher. Questions addressed both intervention teacher and student satisfaction with the *Words Their Way* program, average weekly implementation practices, differences in student responses to the program based on gender, ethnicity, ELL status, etc., and differences between *Words Their Way* and other similar intervention programs. Twenty-two of the twenty-three participating intervention teachers participated in a focus group or an individual interview.

Data Collection Measures: Outcomes

Participation in the study required students to complete two measures at pretest and posttest as well as a progress monitoring assessment three times during the course of the study. The progress monitoring assessment used was the AIMSweb R-CBM and the student outcome measures in this study were the Metropolitan Achievement Test, 8th Edition (MAT8)— which consisted of two subtests, Sounds and Print and Spelling; and a student attitude survey. The goal of the MAT8 assessment was to obtain an objective measure of student achievement in language arts skills to compare across schools in multiple states. The following includes a description of AIMSweb R-CBM and the other outcome measures used in the study.

Progress Monitoring Assessment / Distal Outcome Measure: AIMSweb R-CBM was used as the diagnostic and progress-monitoring tool, also known as a distal measure. The AIMSweb R-CBM is considered a distal measure because it assesses areas of competence related to skills targeted by *Words Their Way* but not taught directly in the program. The AIMSweb R-CBM assessment requires students to read passages aloud while a trained administrator records the number of words read correctly and the number of errors made during the reading, giving the students a score. The student reads three passages and the median of the three scores (words read correct) is used as the overall score. The student's percentile is then calculated using the median score. AIMSweb R-CBM was administered as soon as a school was to participate as a diagnostic tool to determine which students should receive the intervention. This assessment was administered twice more as a performance assessment, once mid-school year and again toward the end of the school year. AIMSweb R-CBM is a preferred diagnostic and progress monitoring tool because the NCRTI TRC reviewed its use as a progress-monitoring tool and rated the tool high on reliability, validity, disaggregated reliability and validity, alternate forms, end-of-year benchmarking, and rates of improvement; although, sensitivity to student improvement was rated lower than the other criteria. Table 8 describes the AIMSweb R-CBM as a distal outcome measure.

Measures Q2
What is the distal outcome measure?

Proximal Outcome Measures: The study also tracked student proximal outcome measures; specifically, two reading assessments were administered as pretest and posttest measures. Table 9 describes the proximal outcome measures followed by a detailed description of each.

Measures Q1
What is the proximal outcome measure?

Table 8. Distal Measures

Distal Measure (name)	Score type and range of measure	Reliability statistics (specify type of reliability)	Relevance to program instructional content
AIMSweb Reading Curriculum-Based Measurement	Raw Score: 0 – 199 Percentile Rank: 1 – 99	Test-retest: Grade 2 Fall-Winter (.93); Grade 2 Winter-Spring (.94) Grade 4 Fall-Winter (.95), Grade 4 Winter-Spring (.95)	Measure establishes reading fluency (speed and accuracy).

Standards-based Reading and Language Arts Assessment

A standards-based, nationally recognized reading and language arts assessment was identified to measure student learning in language arts class. We assessed student reading achievement using two subtests of the the Metropolitan Achievement Test, Eighth Edition (MAT8): Sounds and Print and Spelling. Students completed the test appropriate for their grade level. We converted raw scores obtained from testing into scaled scores reported in Section Five; we also converted these scores in national percentile ranks and grade-equivalence scores for comparison purposes. The Sounds and Print subtest included 40 multiple-choice questions for second graders and 30 multiple choice questions for fourth graders. The Spelling section contained 30 multiple-choice questions for both grades.

Student Survey

Student surveys were administered as both a pretest and a posttest to assess attitude change over the duration of the study. All students participating in the study were required to complete a self-report survey that addressed attitudes towards both recreational and academic reading. The survey utilized was the *Elementary Reading Attitude Survey*, developed by Michael McKenna and Dennis Kear (McKenna & Kear, 1990). The survey consists of twenty questions and was developed specifically for elementary grade students. The response scale, consequently, consists of four pictures of the cartoon cat Garfield, ranging from looking very happy to very upset. Students were instructed to circle the picture that most closely represented their feelings in relation to each question.

The first ten questions relate to recreational reading (e.g. "How do you feel about starting a new book?") and the second ten questions relate to academic reading (e.g. "How do you feel when you read aloud in class?"). A detailed description of the development of the measure as well as its subscales can be found in McKenna & Kear, 1990.

Reliability analyses (i.e. Cronbach's alpha) were conducted to measure the internal consistency of attitude scales. These analyses were conducted for both subscales, recreational and academic reading, and for the overall survey. Table 10 shows the results of the reliability analyses,

Table 9. Proximal Measures

Proximal Measure (name)	Score type and range of measure	Reliability statistics (specify type of reliability)	Relevance to program instructional content
Metropolitan Achievement Tests Eighth Edition (Form V): Sounds and Print (Primary 1)	Raw Score: 1 – 40 Scaled Score: 352 – 679 Percentile Rank: 1 – 99 Grade Equivalent: K.0 – 12.9	Internal consistency (Kuder-Richardson) $r = .91$ Test-retest $r = .91$	Measures ability to discriminate sounds and recognize letters, basic words and phonemegrapheme relationships.
Metropolitan Achievement Tests Eighth Edition (Form V): Spelling (Primary 1)	Raw Score: 1 – 30 Scaled Score: 348 – 634 Percentile Rank: 1 – 99 Grade Equivalent: K.0 – 12.9	Internal consistency (Kuder-Richardson) $r = .88$ Test-retest $r = .87$	Spelling is assessed in context. Students identify a misspelled word in a sentence. The spelling of sight words is measured.
Metropolitan Achievement Tests Eighth Edition (Form V): Sounds and Print (Elementary 1)	Raw Score: 1 – 30 Scaled Score: 447 – 731 Percentile Rank: 1 – 99 Grade Equivalent: K.0 – 12.9	Internal consistency (Kuder-Richardson) $r = .83$ Test-retest $r = .85$	Measures ability to discriminate sounds and recognize letters, basic words and phonemegrapheme relationships.
Metropolitan Achievement Tests Eighth Edition (Form V): Spelling (Elementary 1)	Raw Score: 1 – 30 Scaled Score: 454 – 732 Percentile Rank: 1 – 99 Grade Equivalent: K.0 – 12.9	Internal consistency (Kuder-Richardson) $r = .82$ Test-retest $r = .79$	Spelling is assessed in context. Students identify a misspelled word in a sentence. The spelling of sight words is measured.

PEARSON *WORDS THEIR WAY: WORD STUDY IN ACTION*
INTERVENTION EFFICACY STUDY
FINAL REPORT

Table 10. Student Survey Cronbach's Alpha Reliability Measures

	2nd Grade	4th Grade	Overall
Academic	.84	.83	.84
Recreational	.80	.85	.83
All questions	.90	.91	.90

which correspond to the results demonstrated in the original publication (McKenna & Kear, 1990).

Study Procedures and Measures Summary

The efficacy study was designed to assess implementation of the curriculum in classrooms, answer research questions related to student achievement and attitudes, and to assess product satisfaction from teachers and students. **Implementation measures** were collected to assess the extent to which students and teachers implemented their respective language arts programs in their classrooms. **Outcome measures** were administered as pretest and posttest instruments and assessed the impact on student attitides and acheivement. The following table summarizes the implementation and outcome measures used in the *Words Their Way* study.

Outcome Measures	
AIMSweb R-CBM	A norm-referenced measure in which students read three passages aloud and a trained administrator recorded the number of words read correctly and the number of errors. The median score of the three passages was used as the overall score and the percentile was calculated from the overall score. This measure was administered at the beginning of the year, once in the middle of the year, and again at the end.
MAT8 – Sounds & Print	A norm-referenced measure of language arts abilities. Sounds and Print measures a student's ability to discriminate sounds and recognize letters, basic words and phoneme-grapheme relationships.
MAT8 - Spelling	A norm-referenced measure of language arts abilities. Spelling measures a student's ability to identify misspelled sight words.
Student attitude survey	Included questions regarding students' enjoyment of recreational and academic reading.
Implementation Measures	
Online logs	Completed by all participating intervention teachers weekly to report the sorts covered and specific components used in the classroom.
Classroom observations	Observed by researchers, all intervention teachers and their students participated in two observations (fall and spring).
Teacher interviews / focus groups	Completed at the end of the study, most intervention teachers participated in individual interviews or focus groups to discuss the program implementation over the duration of the year.

Section Four: Assessment of Curriculum Implemention

Reviewing implementation is a key factor in a curriculum study since it is possible for implementation of a particular program to vary across sites and teachers. This study tracked program implementation from the initial training through the final assessment. Through the classroom observations and online teacher logs, we were able to examine the breadth of the content covered as well as the quality of implementation. The following section provides an analysis of the implementation of the treatment curriculum, specifically focusing on the extent of coverage and fidelity to implementation guidelines. We also address the efficacy study's first research question related to implementing the *Words Their Way* program.

Research Question 1:

Are intervention teachers using Words Their Way *program able to implement the program according to NCRTI guidelines (e.g., minimum of 60 minutes per week in a small-group-pull-out program?*

Description of Program Implementation

Intervention teachers were required to adhere to specific implementation guidelines requiring the integration of specific components of the *Words Their Way* program into their classes. Guidelines for using the *Words Their Way* curriculum were reviewed during the study orientation sessions. Appendix F includes the implementation guidelines for all participating groups. The purpose of the implementation guidelines was to ensure that treatment teachers would fully implement the *Words Their Way* curriculum as intended by the developers. These guidelines were developed with the cooperation of the research team and Pearson's editorial/product management team.

Fidelity of Implementation Q1
How was the program delivered?

Intervention teachers used the *Words Their Way* program with small groups of students in a pull-out format, where students would leave their normal classrooms to use the *Words Their Way* program in another room, away from the other students in the class. This was often done in a resource room, library, or other empty classroom. Teachers often used a word wall to display sorts and had other supplies available such as scissors, glue, and pencils, to allow students to use the sorts in a variety of ways. The *Words Their Way* program was implemented with small groups of students. The average group size was 5 students and ranged in size from 2 to 8.

Fidelity of Implementation Q2
What was the duration of the intervention?

The average number of implementation weeks was 18.2, with a typical week including a 3.8 mean number of sessions that were recommended to last 20 minutes each. The total number of weeks implemented ranged from 16 to 28.

Fidelity of Implementation Q4
Describe when and how fidelity of implementation information was obtained.

Participating teachers were required to complete weekly online logs that detailed classroom activities and book usage. In addition, we conducted classroom observations two times during the year (often fall and spring). A full summary of how teachers implemented *Words Their Way* during intervention sessions can be found in Appendices H and I. We also asked teachers to self-rate their level of adherence to the program on a scale of 1 to 9 each week (1 = not at all, 5 = somewhat, 9 = fully). Finally, we conducted informal interviews during the first observation sessions to inform the progress of using the program with students; formal interviews were conducted during the second observation site visits. Results of the teacher interviews can be found in Section Six.

Fidelity of Implementation Q5
Provide documentation (i.e., in terms of numbers) of fidelity of treatment implementation.

Fidelity of Implementation Results

As Appendix G shows, the average number of minutes teachers were able to implement every week ranged from about 74 to 138 minutes. The percentage of average minutes per week demonstrates the proportion of time that the program was implemented each week when the total amount of recommended minutes (100) is taken into account. Therefore, teachers were able to implement, on average, 104% of the recommended time each week, demonstrating that most teachers implemented more than was expected. The table also reports a percentage of

the number of minutes each intervention teacher could implement depending on the total amount of possible minutes, given that snow days, assemblies, and holidays would inevitably prevent implementation each regular school day. Generally, intervention teachers implemented the *Words Their Way* program 100% of the time, an average of the total percentages of implementation. In addition, Appendix H includes the total number and type or "sorts" they implemented during the year, separated by each of the five books (i.e., *Emergent Early Letter Name; Within Word Pattern*). On average, teachers covered 20.7 sorts, and most of these were from the *Letter Name* and *Within Word Pattern* books. The average self-reported degree of implementation was 6.7 on a scale from 1 to 9. This indicates that intervention teachers believed that, to some extent, they implemented the program with fidelity.

Classroom Observations

Researchers from Cobblestone and representatives from Pearson conducted two observations of each intervention teacher and their students during the year. The first set of observations was scheduled for one to two months after implementation began, and the second set of observations was scheduled during the final weeks of the school year. During the observations, researchers documented classroom activities carefully and completed an observation protocol form. Observation protocol forms prompted the research team to gather information about the students in the classroom, scheduled and actual start and end times for the sessions, coverage of instructional materials (types of sorts and types of activities), and classroom variables such as student engagement and student and teacher rapport.

Overall, treatment teachers and their students were engaged in a variety of activities during observations. The most observed types of sorts were Pattern sorts, Speed sorts, and Sound sorts. Most teachers implemented the program that resembled training sessions and implementation guides, with some exceptions. Some teachers had students engage in other types of games and activities that were not part of recommended guidelines; however, they used the words and pictures from the program faithfully. After carefully calculating length of each session, most sessions exceeded 25 minutes and several sessions that were more than 35 minutes, which provided additional support for teachers' estimates of session length reported in the logs. In observing the interactions with teachers and students, most had high rapport, good classroom management, and students were usually highly engaged given the small size of the intervention groups.

Implementation Summary

To establish implementation fidelity, we assessed teachers in a variety of ways including self-reported online teacher, interviews, and classroom observations. The level and quality of implementation varied throughout the study, but treatment teachers generally used the *Words Their Way* program in the intervention sessions with fidelity, which were also reported in weekly logs. On average, intervention sessions were longer than expected (more than 100 minutes per week), however, the number of sorts covered was at a slower pace than recommended given the number of sorts that were covered by each intervention teacher, where the average number of sorts teachers covered during the year was about 21.

Section Five: Results Related to Students' Attitudes and Achievement in Language Arts

In this section, we answer the major research questions involving student outcomes in achievement and attitudes (i.e., Research Questions 2 – 4). Each research question addressed in this section is listed, followed by a detailed explanation of the results obtained from the outcome measures (i.e., MAT8 Sounds and Print, MAT8 Spelling, AIMSweb R-CBM, and student survey).

Research Question 2:

How do student reading performance assessments differ for those Tier II students using Words Their Way *compared with those Tier II students using another intervention program?*

HLM Analyses of Outcome Measures

Given that we randomly assigned students to the treatment and control conditions, and students were nested within different groups (i.e., random assignment of students into different classrooms), we used hierarchical linear modeling (HLM) to examine differences in achievement between the treatment and control groups, taking into account various key student and teacher characteristics. HLM models were particularly appropriate for analyzing data of this kind (i.e., students within different groups) because they simultaneously examined the effect of student background variables (e.g., ethnicity) and teacher/instructional characteristics (e.g., rapport with students) on students' reading achievement. In other words, HLM analysis is used to account for the differences between the teachers across all schools in order to better detect the actual differences between students in the treatment and control groups. For a complete discussion of the rationale and theory underlying HLM models, please see Raudenbush and Bryk (2002).

Appendix I describes the HLM statistical model (i.e., random intercept model in STATA) and includes a list of variables and their operational definitions associated with student background characteristics and teacher/classroom/school characteristics that were used in the HLM models. These variables fell into the following four categories: (1) key student demographic background characteristics (e.g., gender, ethnicity); (2) baseline measure on the key outcomes (i.e., pretest scores); (3) teachers' years of teaching experiences; and (4) the key variables of interest, namely, the treatment condition indicator variable, and for the treatment condition, the level of implementation in terms of the total number of minutes using the *Words Their Way* program.

Student achievement was measured using the MAT8 Sounds and Print and MAT8 Spelling subtests. The results from both subtests were converted into scaled scores in order to combine the results from grades two and four. The following sections address each MAT8 subtest individually.

Metropolitan Achievement Tests 8 Sounds and Print

Performance on Sounds and Print: Comparing Treatment and Control. As shown in Table 11, controlling for various student characteristic measures, we found that on average, there was no statistically significant treatment effect such that the students using *Words Their Way* performed comparably to students using other reading intervention programs (see the coefficient associated with "condition"). However, we observed a statistically significant interaction effect between the condition and grade level (see the coefficient associated with "Grade by condition"). Specifically, holding constant other predictors in the model, the fourth grade students in the treatment condition outperformed students in the other three groups defined by the condition and grade level on the MAT8 Sounds and Print test. These three groups included fourth grade control, second grade treatment, and second grade control groups. Given that we expected fourth grade students to outperform second grade students it is not surprising that the second grade groups were outperformed by the fourth grade *Words Their Way* group; however, the important finding is that the fourth grade *Words Their Way* group did in fact outperform the fourth grade control group based on this analysis.

This significant interaction effect of condition by grade level was observed after controlling for various student characteristics, some of which were significantly related to the outcome measure (see the coefficients associated with "pretest", grade level indicator, "other ethnicity", "subsidized lunch", and "disability" indicators in Table 11). When interpreting the results of the HLM analysis, it is important to realize that each variable is reported on after controlling for all other characteristics in the HLM model. In other words, the results of the variables are reported after considering all other characteristics as equal. Using all of the available data gathered on this sample of students, this HLM model was the best fit to the outcomes

Table 11. HLM Results for MAT8 Sounds and Print Scaled Scores

Fixed Effect	Coefficient	Standard Error	Approximate *t*-Ratio	*p*-value
Pretest MAT8 Sounds and Print*	0.598	0.050	12.060	< 0.001
Grade 4*	15.215	5.546	2.740	0.006
African American	-2.586	4.483	-0.580	0.564
Hispanic/Latino	-0.093	3.647	-0.030	0.980
Other ethnicity*	-14.952	6.543	-2.290	0.022
Subsidized lunch	-5.208	2.789	-1.870	0.062
Disability*	-17.764	5.553	-3.200	0.001
English language learner	-1.194	3.327	-0.360	0.720
Female	2.054	2.637	0.780	0.436
Condition	-3.670	4.393	-0.840	0.404
Grade by condition*	14.306	6.321	2.260	0.024
Intercept*	239.132	26.711	8.950	< 0.001

**Significant predictor of MAT8 Sounds and Print posttest score; n = 252*

measured. Essentially, this HLM model could be used to predict a student's score after identifying the student and teacher characteristics that are contained therein. Although the other variables in the HLM model are not significant, they are included because they provided the best fit for the data (i.e., they are theoretically meaningful and provide more precision in the overall prediction of the MAT8 *Sounds and Print* scaled score).

Treatment Group Performance on Sounds and Print: Comparing Performance by Implementation. In order to examine whether there was an association between the level of implementation (i.e., total number of minutes using *Words Their Way* by a treatment teacher), we ran a model on students and teachers in the treatment condition. The total number of minutes was standardized to have a mean of zero (i.e., the mean total number of minutes in the sample) and a standard deviation of one (i.e., the standard deviation of total number of minutes in the sample). In addition, we also included a proxy measure of teacher quality in this model since it was available for the treatment teachers. This proxy measure was years of experience teaching K-12. Results in Table 12 suggest that neither the levels of implementation nor the proxy measure of teacher quality were statistically significant predictors of the outcome measure (see the coefficients associated with "zmin" and "yrsk12").

Table 12. HLM Results for Treatment Group MAT8 Sounds and Print Scaled Scores

Fixed Effect	Coefficient	Standard Error	Approximate *t*-Ratio	*p*-value
Pretest MAT8 Sounds and Print*	0.616	0.073	8.480	**< 0.001**
Grade 4*	28.288	5.609	5.040	**< 0.001**
African American	-7.040	6.446	-1.090	0.275
Hispanic/Latino	5.484	4.806	1.140	0.254
Other ethnicity*	-25.100	9.711	-2.580	**0.010**
Subsidized lunch	-5.891	3.598	-1.640	0.102
Disability	-9.518	7.389	-1.290	0.198
English language learner	-7.385	4.420	-1.670	0.095
Female	3.932	3.555	1.110	0.269
Zmin (Minutes of implementation)	-0.480	1.999	-0.240	0.810
Yrsk12 (Years of teaching experience)	-0.052	0.199	-0.260	0.794
Intercept*	226.889	39.227	5.780	< 0.001

**Significant predictor of MAT8 Sounds and Print posttest score; n = 135*

Table 13. HLM Results for the MAT8 Spelling Scaled Scores

Fixed Effect	Coefficient	Standard Error	Approximate *t*-Ratio	*p*-value
Pretest MAT8 Spelling*	0.718	0.046	15.480	< 0.001
Grade 4	-8.456	6.448	-1.310	0.190
African American	-2.894	6.022	-0.480	0.631
Hispanic/Latino	-4.508	4.551	-0.990	0.322
Other ethnicity	-1.727	8.512	-0.200	0.839
Subsidized lunch	-2.297	3.680	-0.620	0.533
Disability	-8.513	6.807	-1.250	0.211
English language learner	3.969	4.162	0.950	0.340
Female	4.788	3.473	1.380	0.168
Condition	2.950	5.161	0.570	0.568
Trtgrade	3.101	7.398	0.420	0.675
Intercept*	188.840	24.359	7.750	< 0.001

**Significant predictor of MAT8 Spelling posttest score; n = 255*

Some of the student characteristics were significantly related to the outcome measure (see the coefficients associated with "pretest", grade level indicator, "other ethnicity", and "English language learner" indicators in Table 12). With regard to student characteristics, we found none of the covariates were significantly associated with students' MAT8 Spelling scores, except for the pretest scores (see the coefficient associated with "pretest" in Table 13).

Metropolitan Achievement Tests 8 Spelling

Performance on Spelling: Comparing Treatment and Control. As shown in Table 13, controlling for various student characteristics, we found that students using *Words Their Way* performed comparably to students using other reading intervention programs students on the MAT8 Spelling test (see the coefficient associated with "Condition" in Table 13).

Treatment Group Performance on Spelling: Comparing Performance by Implementation. Next, we investigated whether the levels of implementation or the teacher quality proxy measure were associated with students' spelling outcome for treatment students only. Results in Table 14 show that the total number of minutes was not a statistically significant predictor of students' MAT8 spelling scores (see the coefficient associated with "zmin"). In contrast,

Table 14. HLM Results for the MAT8 Sounds and Print Scaled Scores

Fixed Effect	Coefficient	Standard Error	Approximate *t*-Ratio	*p*-value
Pretest MAT8 Spelling*	0.664	0.063	10.540	< 0.001
Grade 4	-3.615	6.602	-0.550	0.584
African American	-7.118	9.201	-0.770	0.439
Hispanic/Latino	-11.823	6.240	-1.890	0.058
Other ethnicity	-6.546	13.175	-0.500	0.619
Subsidized lunch	-0.826	5.005	-0.160	0.869
Disability	-12.553	8.941	-1.400	0.160
English language learner	6.929	5.770	1.200	0.230
Female	7.878	4.856	1.620	0.105
Zmin (Minutes)	-0.575	2.424	-0.240	0.812
Yrsk12*	0.556	0.243	2.290	0.022
Intercept*	211.729	33.362	6.350	< 0.001

**Significant predictor of MAT8 Spelling posttest score; n = 138*

Table 15. Illustration of the Significant Interaction Effect

Groups	Grade	Condition	Grade by condition	Coefficient
Grade 2, control	0	0	0	–
Grade 2, treatment	0	1	0	-3.67 *(n.s.)*
Grade 4, control	1	0	0	–
Grade 4, treatment	1	1	1	14.31

Note. The numbers "0" and "1" are the values of indicator variables (grade and treatment). Together, they define the four groups (see the first column).
n.s.= *not statistically significant.*

the proxy measure of a treatment teacher's quality (i.e., years of teaching K12) was positively related to students' performance on the MAT8 spelling test, though the effect was small (see the coefficient associated with "yrsk12"). With a one-year increase in experience, there is a predicted close to two-thirds of a point increase in students' spelling scores, other things being equal. Finally, we found one other significant predictors of students' MAT8 Spelling outcome, which was students' pretest score.

Summary of HLM Results

Results indicate that overall students using *Words Their Way* performed comparably to control group students in Sounds and Print and Spelling subtests; however, fourth grade students using *Words Their Way* outperformed all other groups in the Sounds and Print subtest, including fourth grade control group students. To help understand what this means, we illustrate the interaction effect using Table 15. Again, while focusing on the main predictors of grade level, condition, and grade level by condition interaction, we are holding constant other predictors in our HLM model.

Table 15 indicates that holding constant other predictors in the model, there was no difference in the MAT8 Sound and Print test scores between second grade students in the treatment or control conditions (the coefficient of -3.67 was not statistically significant). In contrast, fourth grade treatment students outperformed fourth grade control students by rough 10 points, calculated as follows: 1 * (-3.67) + 1 * (14.31), which is close to 10. Since our evaluation collected many key student variables, some of which were significantly related to the outcome, we can be confident in the observed positive treatment effect for the fourth graders and focus our effort on understanding why no treatment effect was observed for the second graders, or on the other key outcome measure (i.e., MAT8 Spelling).

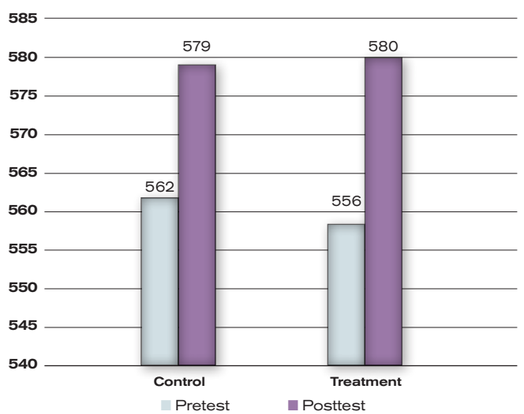
Additional Analyses of Outcome Measures

To further explain and explore the results of the HLM analysis, we performed several additional analyses on the Sounds and Print and Spelling subtests. Similar to the HLM, only the Sounds and Print subtest showed effects in these analyses. We looked at various subgroup analyses (e.g., gender, socio-economic status, grade level) to determine if any effects existed in these subgroups between control and treatment groups. The following section contains the results of these analyses *if they yielded differences between the control and treatment groups*. While most of the subgroup analyses did not yield significant results, the trends in the data favor the treatment group in many analyses performed for both subtests. That is, the rate of change from pretest to posttest tended to be greater for the treatment group although not statistically significant.

Performance: Comparing Treatment and Control

Figure 2 shows the pretest and posttest scaled scores of the MAT8 Sounds and Print separated by control andtreatment groups. While each of these groups increased their scores from pretest to posttest, the control group's increase for the scaled score was 17.2 points compared to the treatment group's increase of 24.0. When analyzing the rate of change from pretest to posttest for all students, the treatment group is shown to outperform the control group, $F(1, 253) = 3.874$, $p = .05$, overall.

Figure 2. Pretest and Posttest MAT8 Sounds and Print: Control vs Treatment



When analyzing the rate of change from pretest to posttest for all students, the treatment group is shown to outperform the control group.

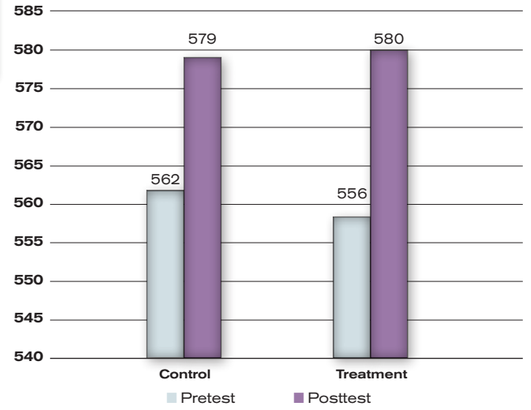
Performance by Grade. As was also shown in the HLM analysis, there was a significant difference between control and treatment students in fourth grade, $F(1, 127) = 10.53$, $p = .002$. This effect was not shown with students in the second grade, $F(1, 124) = .457$, *n.s.* Table 16 shows the breakdown of grade level scores along with percentile ranks and grade equivalents. The grade equivalent scores show a slight increase for second grade students and larger increases for fourth grade students. Percentile ranks showed a decrease from pretest to posttest in second grade where percentile ranks were available.

Performance by Ethnicity. Another significant effect was found when looking only at ethnic minority (non-Caucasian) students in the study sample. Among ethnic minority students, the treatment group (pretest to posttest difference = 24) outperformed the control group (pretest to posttest difference = 15) from pretest to posttest, $F(1, 99) = 5.008$, $p = .027$ (see Table 17 and Figure 3).

AIMSweb R-CBM Reading Fluency Assessment Results

The AIMSweb R-CBM assessment was administered at the beginning of the year (for most students) and was used to determine eligibility for the study as a diagnostic test. While the AIMSweb R-CBM was a measure of reading fluency, results showed that the pretest percentile scores of the AIMSweb R-CBM were significantly correlated with the pretest scaled scores of the MAT8 Sounds and Print, $r = .157$, $p < .05$, and the MAT8 Spelling, $r = .328$, $p < .001$.

Figure 3. Pretest and Posttest MAT8 Sounds and Print Results for Minority Students



Among ethnic minority students, the treatment group (pretest to posttest difference = 24) outperformed the control group (pretest to posttest difference = 15) from pretest to posttest.

Table 16. MAT8 Sounds and Print: Treatment and Control Scaled Scores by Grade

Grade Level	Condition	Test	*n*	Scaled Score	Scaled Score Increase	Grade Equivalent	Percentile Rank*
2	Control	Pretest	59	528	23	1.3	26
		Posttest	59	551		1.7	20
	Treatment	Pretest	71	531	20	1.3	28
		Posttest	71	551		1.7	20
4	Control	Pretest	61	594	12	3.0	40
		Posttest	61	606		3.9	N/A
	Treatment	Pretest	68	581	26	2.3	29
		Posttest	68	607		4.0	N/A

**The percentile rank was not available for the MAT8 Sounds & Print for spring of 4th grade*

Table 17. Pretest and Posttest MAT8 Sounds and Print Results for Minority Students

Condition	Test	*n*	Scaled Score	Increase
Control	Pretest	44	561	15
	Posttest	44	576	
Treatment	Pretest	57	553	24
	Posttest	57	577	

Figure 4. AIMSweb R-CBM Median Words Read Correct: Raw Score Results

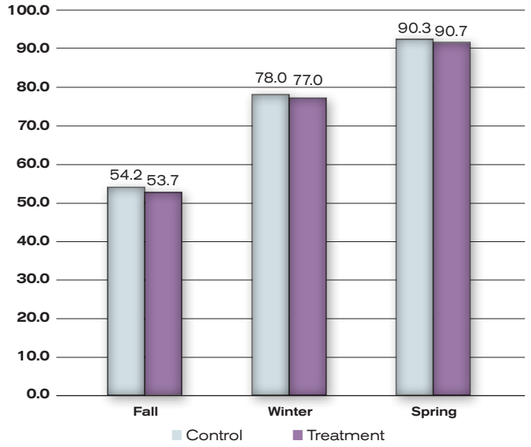
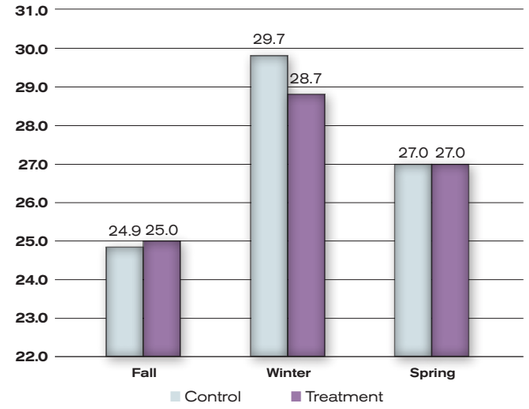


Figure 5. AIMSweb R-CBM Median Words Read Correct: Percentile Score Results



Students were retested on the AIMSweb R-CBM in winter and spring of the 2010-11 school year. Figure 4 displays the results of raw scores from all three administrations. As can be seen, both the control and treatment groups increased their raw scores on each administration. However, after converting the raw scores to percentile scores, both groups show an increase on the second administration followed by a decrease in score for the third administration (see Figure 5). Given that the increase in raw scores between the second and third administration was almost half of the increase between the first and second administration, the percentile scores were not surprising. Statistical analyses showed that there were not any difference between the control and treatment groups overall, $F(2, 492) = .152$, *ns*.

Research Question 3:

How does student reading achievement differ from pretest to posttest for those Tier II students using Words Their Way?

An overall examination showed that the treatment group had significant growth from pretest to posttest for the MAT8 Sounds and Print and MAT8 Spelling but not for the AIMSweb R-CBM percentile score (see Table 18). Additional analyses on subgroups (i.e., grade, ethnicity, socioeconomic status, ELL status, gender) showed similar results in that each subgroup showed significant improvement from pretest to posttest in the treatment group for only the MAT8 subtests.

Research Question 4:

How do Tier II students using Words Their Way *compare to those Tier II students using another intervention program from pretest to posttest on attitudes about reading?*

As mentioned previously, the student survey was administered as a pre-posttest and measured students preference for academic and recreational reading. Figure 6 shows the pretest and posttest results of the student survey. The control group had a slight upward trend from pretest to posttest as opposed to the treatment group's slight downward trend. However, there were not any significant changes for either the control group or treatment group from pretest to posttest (see Table 19).

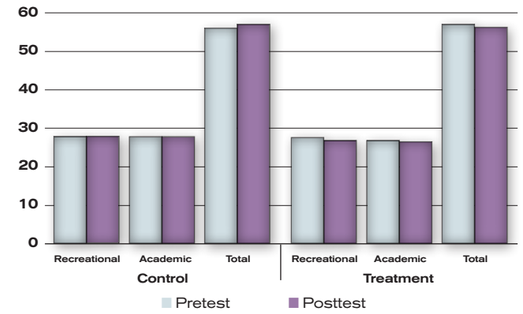
Table 18. Pretest vs. Posttest Scores for Students Using *Words Their Way*

Test	Pretest Mean	Posttest Mean	*n*	SD	*t*	*df*	*p*-value
Sounds and Print Scaled Scores	556	580	135	23.63	11.344	134	< .001
Spelling Scaled Scores	561	592	138	30.53	10.581	137	< .001
AIMSweb R-CBM Percentile	25	27	138	13.79	1.723	137	.087

Table 19. Student Survey Pretest and Posttest Results

	Survey	Pretest Mean	Posttest Mean	*n*	*df*	*SD*	*t*	*p*-value
Control	Recreational	27.98	28.18	120	119	7.33	0.299	0.766
	Academic	27.75	28.35	120	119	6.5	1.01	0.314
	Total	55.73	56.53	120	119	12.43	0.705	0.482
Treatment	Recreational	28.11	27.46	138	137	6.85	-1.118	0.265
	Academic	27.41	27.07	138	137	6.4	-0.638	0.524
	Total	55.52	54.52	138	137	11.67	-1.007	0.316

Figure 6. Student Survey Pretest and Posttest Results



Results Q1
What analyses were used to determine whether the treatment group learned more than the control group?

Results Q2
What are the proximal and distal results?

Summary of Research Findings

A combination of hierarchical linear modeling (HLM), repeated measures ANOVA, and independent *t* tests were used to determine the difference between the control group and treatment group. HLM was used to determine which variables influenced outcome measures including if the study condition was a significant predictor in posttest scores. Repeated measures ANOVA determined if the rate of change from pretest to posttest for students was statistically different depending on their study conditions ignoring other variables. The independent *t* tests were used to determine if there was a difference between control and treatment at pretest and posttest scores.

Research Question 1: Intervention teachers were able to implement many of the *Words Their Way* program elements in intervention groups. On average, the program was used for about 100 minutes each week and groups covered an average of 21 sorts during the year.

Table 20. Results for Proximal and Distal Outcome Measures

Measure		Posttreatment data	
		Treatment	Control
Metropolitan Achievement Tests 8: Sounds and Print	Unadjusted Mean	579.50	578.90
	Adjusted Mean	581.52	576.63
	Unadjusted Standard Deviation	38.97	39.44
	n	135	120
Metropolitan Achievement Tests 8: Spelling	Unadjusted Mean	591.88	584.75
	Adjusted Mean	590.48	586.36
	Unadjusted Standard Deviation	45.00	44.40
	n	138	120
AIMSweb R-CBM	Unadjusted Mean	27.25	27.13
	Adjusted Mean	27.23	27.15
	Unadjusted Standard Deviation	16.58	15.96
	n	139	120

Research Question 2: The most notable finding was that fourth grade treatment students outperformed all other study groups (i.e., fourth grade control, second grade treatment & control). This finding was supported in multiple analyses that were performed on student data. Overall, when ignoring other classroom and group effects, the treatment group outperformed the control from pretest to posttest. However, this overall effect was not seen in the more conservative HLM analysis. Finally, minority students in the treatment group outperformed minority students in the control group when ignoring other classroom and group variables.

Research Question 3: Results showed that students made significant gains from pretest to posttest on the MAT8 Sounds and Print and MAT Spelling subtests. However, students did not make significant gains from pretest to posttest on the AIMSweb R-CBM percentile score.

Research Question 4: Both treatment and control groups did not show gains from pretest to posttest on the student survey when measuring attitudes about reading. Results also showed that there were no differences between treatment and control at the pretest and posttest administrations of the student survey.

Section Six: Product Satisfaction

We asked students and teachers to provide feedback about how much they liked the *Words Their Way* program and also asked teachers to comment on how much this met students' literacy needs and provide any other recommendations for product improvement. Product satisfaction of the *Words Their Way* program was assessed from multiple sources including student surveys, teacher logs, and formal and informal teacher interviews.

Student Satisfaction

Students rated their satisfaction with various components of the *Words Their Way* program using a 4-point scale. Students were asked to rate how each feature of the program made them feel from 1 = Very upset to 4 = Very happy, and were instructed to circle of picture of Garfield that reflected this emotion. Program components rated included the *Words Their Way* books, sorting, book of rhymes, games, and the draw and label activities. Figure 7 displays the questions from the student survey and the percentage of students that felt happy about the various components. Results indicate that most students (73%) were at least happy about these specific program components, and many were "very happy". Not surprisingly, about 87% of all the students rated Games in the WTW books as making them happy or very happy. The *Big Book of Rhymes* and the general *Words Their Way* books had roughly equally numbers of students (approximately 27%) not happy with the program.

"They feel very successful."
"They love it. They don't want it to end."
"…they like something different each day."
"They look forward to coming. They really like the activities."
"They really enjoy manipulating their own learning…"
"…they really liked the sorts. They had fun with them."

–Teacher observations of students using *Words Their Way*

Table 21 includes a summary of the means and standard deviations for each program component for second grade students, fourth grade students, and ratings overall. Fourth grade students consistently provided lower satisfaction ratings in comparison to second grade students on all but one item. Games were rated the highest with a mean of 3.46 for both grade levels. These data correspond to what many teachers reported in which they often expressed fourth grade students felt that the WTW books were "beneath them" and some of the resources, such as the student libraries were "babyish" for fourth grade students and students made comments to teachers such as 'These look like kindergarten books'. A few teachers, however, specifically mentioned that the students thoroughly liked the fact that the *Words Their Way* program was "interactive," "hands-on," and an opportunity to

Figure 7. Students' Favorability of the *Words Their Way* Program

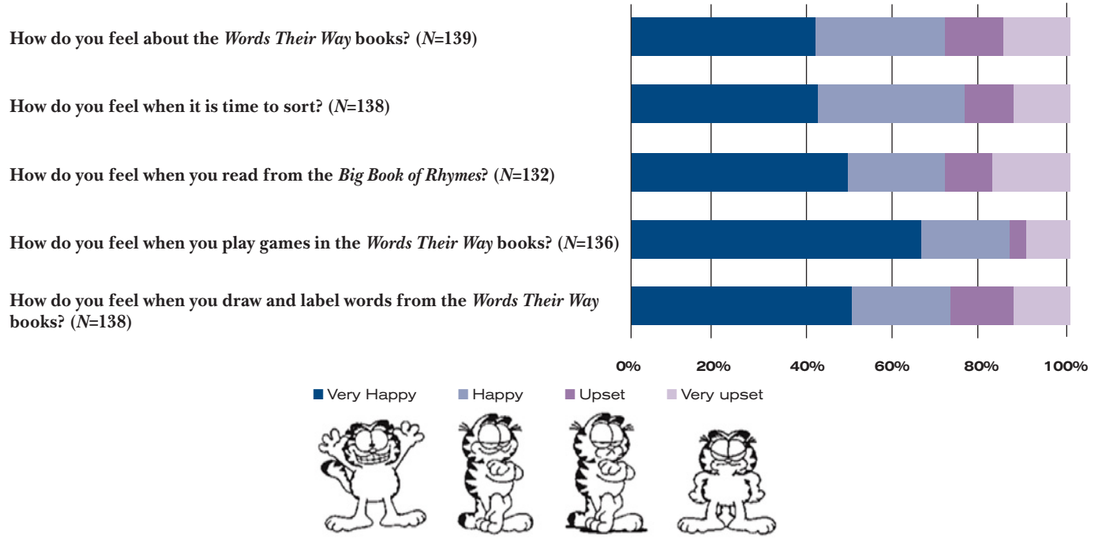


Table 21. Product Satisfaction Ratings of *Words Their Way* Program Components by Grade Level

WTW Components		Mean	SD
WTW Books	Overall	3.00	1.06
	2nd Grade	3.07	1.14
	4th Grade	2.93	.98
Sorting	Overall	3.07	1.02
	2nd Grade	3.19	1.03
	4th Grade	2.96	.99
Book of Rhymes	Overall	3.06	1.13
	2nd Grade	3.18	1.06
	4th Grade	2.94	1.19
Games	Overall	3.46	.93
	2nd Grade	3.46	1.03
	4th Grade	3.46	.84
Draw & Label	Overall	3.12	1.06
	2nd Grade	3.17	1.06
	4th Grade	3.07	1.08

"manipulate materials." Overall, as also revealed through teacher comments during the interviews, students generally enjoyed the program.

Teacher Satisfaction

In general, most teachers provided positive feedback about the *Words Their Way* program. During their interviews, teachers stated that having students sort words was the best program component. Many also discussed that the structure of the program was helpful, often allowing students to engage in different activities and giving the teachers the chance to the option to incorporate a variety of activities. Conversely, the majority of teachers found the organization of the *Teacher Resource Guide* and CD unsatisfactory, primarily when sorts in the books were not aligned to the CD. One teacher said, "… it just got very confusing," while another teacher stated it was "not very user friendly." Teachers were informed that changes in the Developmental model were responsible for this lack of alignment. Nevertheless, this is one area where materials should be modified to increase usability and reduce confusion.

We conducted an analysis of feedback provided by teacher interviews to understand which element of the program best contributed to students' learning. Several teachers determined that it was the act of sorting as well as the repetitive nature of the program. Most teachers emphasized the fact that repetition was the key to the program's success. The following Figure 8 is a graphic representation of teacher feedback about the *Words Their Way* program.

Figure 8. Teacher Feedback for *Words Their Way* Components that Enhance Student Learning

Instructional Components

Teachers commonly provided positive comments about the instructional components of the program, including the pace and flow of lessons, the subject matter addressed, and the influence on teacher instruction. The dominant perspective about the pacing and flow of the lessons was that it was "appropriate" or "great." Along with this, teachers recognized and appreciated the opportunity to adjust lessons, often condensing easier sorts or taking a day longer on more challenging sorts, depending on the students' needs. When asked which areas of language arts were covered by the program, several teachers indicated that the students' vocabulary and spelling deficits were addressed. Additionally, a handful of teachers believed that reading was addressed in the program, but even more of them thought that a focus on reading, reading fluency, and writing were missing components. Lastly, virtually all teachers agreed that implementing the program influenced their instruction. Their comments suggest that even though not all teachers would utilize the entire *Words Their Way* program in the future, they acquired valuable methods of teaching that they can apply in the classroom.

> *"I am more aware of students' needs. I definitely use the techniques in other instruction."*
> *"It motivated me more."*
> *"[I] try to get them to think a little bit more and take the lead, where I'm more like a facilitator, trying to get them to use some creative thinking."*
> *"Keeps me more mindful of keeping them [the students] engaged and mixing things up."*
> *"I think I have become a better teacher with word study…"*
>
> **–Teacher comments feedback on using *Words Their Way***

Response to Intervention Components

A Response to Intervention (RTI) framework encourages teachers to recognize a student's learning capacity and to adjust the intervention according to a student's skills. Therefore, teachers were asked to address RTI-related topics, such as changes in student motivation, meeting students' needs, suggestions for appropriate implementation time, and diversity between students' response to the intervention. Although there was mixed feedback regarding whether the *Words Their Way* program met the needs of all the participating students, most teachers agreed that their students had something to gain from the intervention. Of the few that reported that the program did not fully meet students' needs (typically fourth grade intervention teachers), teachers indicated that the intervention may just not be the right fit for some students or that the program was only helpful in some aspects, but lacking in others. A few fourth grade teachers suggested that perhaps the program would be best suited for younger students, as the bulk of the program's concentration was narrowed in on topics many fourth grade students should have previously experienced. Some fourth grade teachers reported that their students lost interest or got bored with the program. However, some teachers mentioned their students got more enthusiastic as time progressed. In particular, one teacher had an important insight considering this intervention teacher taught both 2nd and 4th grade. This teacher said that the 4th grade students got bored after a while, but that the 2nd graders' motivation increased because they saw "more success with the program" and "became more confident."

> *"I think they have always been pretty engaged."*
> *"They are motivated every day as far as I can see."*
> *"…they really looked forward to it…"*
>
> **–Second grade teachers**

> *"It got worse… It has become mundane and boring."*
> *"…a steady decline of interest."*
> *"They are losing interest now. Becoming monotonous doing same things over and over again."*
>
> **–Fourth grade teachers**

Some teachers noticed that girls were more focused than boys and that ELL students' generally had more difficulty with the material. A majority of teachers stated that students' engagement and motivation did not necessarily change during the intervention time, but that most students were motivated or engaged consistently throughout the program. Nevertheless, teachers believed that the program functioned well over an academic school year, especially for younger students. Teachers typically implemented 20 to 30 minutes of the program daily and stated that this was sufficient time for the lessons. Some teachers would have preferred more time, about 40 to 45 minutes, while a smaller group of teachers would have liked to implement only 10 to 15 minutes as one teacher claimed, "that is all I can hold their attention with." Regardless, all teachers attempted to implement the program with fidelity throughout the year.

Product Satisfaction Summary

Overall, student and teacher users of the *Words Their Way* program were satisfied with the program. As indicated by student surveys and teacher interviews, most students were happy with the various components of the program, especially the games. Teachers were also satisfied with the program and found the method of sorting met students' needs, reported that the repetitive nature of the program was useful, and provided positive feedback about their own practice as a result of using the program. However, contrary to results where fourth grade students using *Words Their Way* performed well (see Section 5), many fourth grade teachers reported that the program was less effective for fourth grade students who found some aspects of the intervention boring and below their level. Still, the program appeared to be appropriate for multiple students and worked well as an RTI intervention.

Section Seven: Discussion

This study investigated how the *Words Their Way* program impacted students' reading skills in comparison to other students not using the program. Specifically, we investigated using *Words Their Way* as an intervention program for Tier II students who were identified as needing help in reading skills in second and fourth grades. The study included complete tracking of product use and satisfaction with the program as well as a range of other implementation and outcome measures. The following is a brief discussion of key themes from the intervention efficacy study as well as study limitations.

Efficacy Study Key Findings

An analysis of study data from logs, observations, and interviews indicates that intervention teachers implemented the *Words Their Way* program with fidelity in their intervention sessions throughout the year. The measures we used were objective tests of reading achievement and the student survey was also an appropriate measure of student attitudes towards academic and recreational reading. Therefore, we are confident that this efficacy study was a fair test of the *Words Their Way* program as a reading intervention for second and fourth grade students. Given this, there are a few important findings worth highlighting in interpreting study data.

Intervention teachers and Students Liked Using the *Words Their Way* Program

Through our conversations with intervention teachers during observations/interviews as well as results from the student surveys, it was clear that students and intervention teachers liked the various components of *Words Their Way* and enjoyed using it as an intervention program. The satisfaction of the program was particularly pronounced in second grade. Intervention teachers liked program structure such as the repetition, liked the "hands on" nature of the materials and suggested that this structure aided in student understanding. Overall, all students and intervention teachers provided favorable reports about *Words Their Way*, but these reports appear more enthusiastic for the second grade group which can be seen in the student survey favorability ratings as well as teacher reports.

Likeability Was Not Correlated with Achievement

In contrast to favorable ratings about using the *Words Their Way* program, likability did not always translate into higher student scores. For example, second grade students using *Words Their Way* consistently (but not significantly) rated the program higher than fourth grade students yet this is contrasted with students' reading achievement scores in which fourth grade students using *Words Their Way* outperformed all other groups (all second grade students and control group fourth grade students) on the MAT8 Sounds and Print subtest. Teacher likability of the *Words Their Way* program was also aligned to student ratings, but not to student scores. Overall, second grade intervention teachers expressed more enthusiasm about the program for second grade students in comparison to using the program with fourth students. Teachers reported that some fourth grade students became "bored" with the act of sorting in contrast to younger children did not find the same types of activities mundane. We understand that fourth grade students thought that student libraries were simplistic and "babyish" and hence less favorable to these older students. However, fourth grade students had greater gains using *Words Their Way* materials in contrast to the other groups and therefore the program 'worked' for them as an intervention program despite their lower ratings of likability.

The instances when intervention teachers reported not liking the program were more related to coordination of ancillary materials (i.e., CD activities were linked to specific sorts but linkages were not always obvious; additional materials were not available for every sort) rather than the core program components. Where intervention teachers reported dissatisfaction on the part of students, this was primarily related to the student libraries, which were less favored than other program components, particularly for the fourth grade group.

Achievement Was Not Related to Student Attitudes about Reading

Our findings also indicated that student attitudes were related to likability of the *Words Their Way* program, however, these attitudes were not correlated with their achievement scores. The survey instrument we used was a published scale that measured students' attitudes towards academic and recreational reading for students in the elementary grades. What is clear about examining these results is that all students (treatment and control) were very consistent in their ratings from pretest to posttest, and consequently scores did not change much from the beginning to the end of the school year. We hypothesized that if students liked the *Words Their Way* (or even control program), that students might demonstrate more positive feelings about either academic or recreational reading over the course of the school year as they increased their skills. However, despite the fact that there were gains from pretest to posttest on assessments (some of which may be attributed to maturation), students did not show any changed in attitudes associated with the survey scales measuring academic and recreational reading interest. This is noteworthy in that many studies attempt to link attitudes (soft indicators) and achievement (hard indicators) in reading, yet this may not be true of intervention students who have most likely experienced many years of low achievement and discouraged attitudes about reading even in the early elementary grades. It is noteworthy that the study sample included an extremely truncated range of students (e.g., mean below the 25th percentile in reading fluency) and therefore the normal range of student attitudes that correspond to achievement is not observed in the present case as it would be by including a wider range of students. Future studies should continue to investigate these relationships with the understanding that the nature of the Tier II student sample might always demonstrate a different pattern of attitudes when contrasted with all elementary students.

Study Summary and Limitations

The current study was conducted to examine how *Words Their Way* could be used as a small-group pull-out intervention program for students who struggle with reading. We have investigated some of the ways in which this program would be implemented as an intervention, and found that *Words Their Way* can be used as a small group intervention study for Tier II students. We also found that despite the fact that fourth grade students and intervention teachers found the program less favorable than second grade intervention teachers and students, it appears objectively more effective for Tier II fourth grade students as compared to other groups. We also found that neither intervention group impacted student attitudes about academic or recreational reading, which remained flat over the course of the study. One caveat worth remembering is that the study does not consider using *Words Their Way* as a whole-group instruction model with a regular classroom teacher. We cannot speak to the efficacy of using the program in this format given these study results, specifically because of the nature of the groups. Future studies should examine how *Words Their Way* functions as both a whole-group intervention as well as a pull-out intervention to see whether this factor accounts for performance.

PEARSON *WORDS THEIR WAY: WORD STUDY IN ACTION*
INTERVENTION EFFICACY STUDY
FINAL REPORT

References

Apel, K. (2009). The acquisition of mental orthographic representations for reading and spelling development. *Communication Disorders Quarterly, 31,* 42-52.

Bear, D. R., Invernizzi, M., Templeton, S., & Johnston, F. (2008). *Words Their Way: Word Study for Phonics, Vocabulary, and Spelling Instruction (4th edition).* Allyn & Bacon.

Bowey, J. A., & Muller, D. (2005). Phonological recoding and rapid orthographic learning in third-grade children's silent reading: A critical test of the self-teaching hypothesis. *Journal of Experimental Child Psychology, 92*, 203–219.

Burns, M. K. (March 2008). Response to intervention at the secondary level. *Principal Leadership,* 12-15

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analytic review of responsiveness-to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment, 23,* 381-394.

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36,* 250-287.

Ehri, L., & Rosenthal, J. (2007). Spelling of words: A neglected facilitator of vocabulary learning. *Journal of Literacy Research. 39,* 389-409.

McKenna, M. C., & Kear, D. J. (1990). Measuring attitude toward reading: A new tool for teachers. *The Reading Teacher, 43,* 626-639.

Nagy, W., & Scott, J. (2000). Vocabulary processes. In M. Kamil, P. Mossenthal, P. Pearson, & R. Barr (Eds.), *Handbook of reading research: Volume III* (pp. 269–284).Mahwah, NJ: Erlbaum.

National Center for Education Statistics (2009). *The Nation's Report Card: Reading 2009* (NCES 2010–458). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

National Center on Response to Intervention (Dec. 2009). *Standard protocol for evaluating intervention programs within a response to intervention framework.* Retrieved [5/18/10] from: www.rti4success.org

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Washington DC: National Institute of Child Health and Human Development.

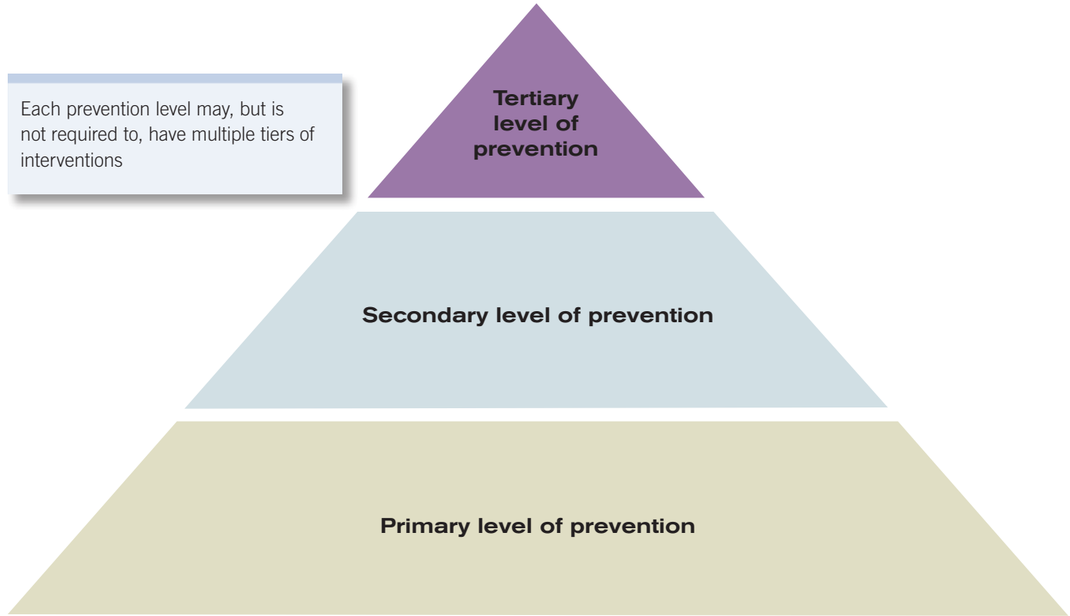
Perfetti, C. A. (1985). *Reading ability.* New York, NY, US: Oxford University Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods. 2nd edition.* Newbury Park, CA: Sage.

Savolainene, H., Ahonen, T., & Aro, M. (2008). Reading comprehension, word reading and spelling as predictors of school achievement and choice of secondary education. *Learning and Instruction, 18,* 201-210.

U.S. Department of Education. (2004). *Building the legacy: IDEA 2004.* Washington, DC: Author

Appendix A: Response-to-Intervention Tiers



	Student Population	Description	Assessment Data
Tier 1	All students	Universal: quality research-based core curriculum and instruction	Benchmark assessments conducted at least three times per year
Tier II	Approximately 15%	Targeted: small-group (three to six students) interventions delivered as part of general education for 30 minutes each day in addition to core reading instruction	Frequent measurement of the skill deficit and at least twice-monthly progress monitoring of general outcome skill
Tier III	Approximately 5%	Intensive: individualized interventions that are based on problem-solving models; could include special education services	At least weekly progress monitoring and frequent informal classroom-based assessments

Appendix B: NCRTI Rubric

Participants
Are the students in the study at risk, and are the program instructors in the study similar to what the vendors state is necessary?
Full Bubble: Evidence is convincing that participants were at risk (i.e., below 30th percentile on local or national norm; or sample mean below 25th percentile on local or national test; or all students below a well justified benchmark; or students with identified disability), and the program instructors were similar to what the vendor states is necessary.
Empty Bubble: Fails full bubble.
Design
Does the study design allow us to conclude that the intervention program, rather than extraneous variables, was responsible for the results?
Full Bubble: Students were randomly assigned. At pretreatment, program and control groups were not statistically significantly different; and were within 0.25 SD on locally or nationally normed achievement measures/well justified benchmark. There was not differential attrition for the program and control group. Unit of analysis matched random assignment.
Half Bubble: Students were not randomly assigned but a tenable quasi-experimental design was used. At pretreatment, program and control groups were not statistically significantly different and were within 0.50 SD on locally or nationally normed achievement measures/well justified benchmark, and outcomes were analyzed to adjust for pretreatment differences. Program and control groups were demographically comparable at pretreatment.
Empty Bubble: Fails full and half bubble.
Fidelity of Implementation
Was it clear that the intervention program was implemented as it is designed to be used?
Full Bubble: Measurement of fidelity of implementation was conducted adequately and observed with adequate intercoder agreement, and levels of fidelity indicate that the intervention program was implemented as intended (at 75% or above).
Half Bubble: Levels of fidelity indicate that the intervention program was implemented as intended (at 75% or above), but measurement of fidelity of implementation either was not conducted adequately or was not observed with adequate intercoder agreement.
Empty Bubble: Fails full and half bubble.
Measures
Were the study measures accurate and important?
Full Bubble: Measures represented a range of proximal and distal outcomes in relation to the program's instructional content. All measures were psychometrically reliable (i.e., all coefficients > 0.59; interscorer agreement not accepted for measures other than writing).
Half Bubble: Measures represented a range of proximal and distal outcomes in relation to the program's instructional content. Most measures were psychometrically reliable (i.e., most coefficients > 0.59; interscorer agreement not accepted for measures other than writing).
Empty Bubble: Fails full and half bubble.

Appendix C: Program Qualifications for NCRTI Review

1) Is your program available for dissemination?

2) Can you provide direct evidence (i.e., refers to data from one or more studies on the program submitted for evaluation) on the effects of your program with students at risk for poor academic outcomes?

3) Does the direct evidence come from a published or unpublished study or technical report that may be obtained?
- **a)** Does the direct evidence address the effects of the overall program rather than individual components of the program?

4) Does your intervention meet the following criteria?
- **a)** Intervention program was delivered in small group or individually
- **b)** Intervention program occurred over a minimum of 60 minutes a week for 8 weeks
- **c)** Characteristics and training of the instructors are described in a users' manual
- **d)** Intervention program is described in sufficient detail in a users' manual so that others can use as conducted in the study

5) Does the study include the following elements of a rigorous design?
- **a)** Random assignment or high quality quasi-experimental assignment methods used
- **b)** Psychometric properties (e.g., reliability) of the dependent measures are described
- **c)** Outcome is a quantitative index of students' academic performance
- **d)** Treatment and control groups are adequately defined and demographically comparable
- **e)** The treatment group is compared to a "business-as-usual" control group

Appendix D: NCRTI Questions and Answers

Study Design		
NCRTI Question	**Answer**	**Page Number**
Q1) Was random assignment used? Please describe the study design.	Yes, the study was designed as a Randomized Controlled Trial (RCT) in which qualified students were randomly assigned to either the treatment group (using *Words Their Way*) or a control group (using the existing reading intervention program at their school or no intervention program).	**12**
Q2) Was the program group comparable to the control group on pretest performance measures? **Q2) Was the program group comparable to the control group on demographics?**	Pretest performance measures: Yes, there were non-significant *p*-values for all pretest academic performance measures (i.e. AIMSweb R-CBM, MAT8: Sounds & Print, MAT8: Spelling), therefore comparability between program and control groups may be assumed. Please see Table 3 for more details. Demographics: Yes, there were non-significant *p*-values for all demographic characteristics (i.e. Race-ethnicity, SES, Disability status, ELL status), therefore comparability between program and control groups may be assumed. Please see Table 4 for more details.	**14-15**
Participants		
NCRTI Question	**Answer**	**Page Number**
Q1) How were students selected to participate in the study?	Students were initially screened with scores on state standardized tests. Of those students screened, students were selected to participate if they scored at or below the 30th national percentile on the AIMSweb R-CBM screening diagnostic reading fluency assessment.	**12-13**
Q2) How were students identified as at-risk for academic failure?	Students were identified through standardized test scores and teacher recommendations; students qualified for participation in the study based on their initial AIMSweb R-CBM reading fluency scores.	**13**
Q2a) What is the treatment program?	*Words Their Way* is the submitted treatment program.	**13**
Q2b) What is the control condition?	Participating students who did not use *Words Their Way* continued using the existing reading intervention program (control condition).	**13**
Q3) What were the sample sizes for the study (for all types of participants and relevant conditions)?	Sample Sizes • Schools: $N = 15$ • Intervention teachers: treatment ($N = 23$); control ($N = 0$) • Classrooms: treatment ($N = 29$); control ($N = 25$) • Students: treatment ($N = 138$); control ($N = 119$)	**13-14**
Q4) How many program students were pre/post tested? **How many control students tested?**	Program students: • Pretest ($N = 144$) • Posttest ($N = 138$) Control students: • Pretest ($N = 127$) • Posttest ($N = 119$)	**14-15**
Q5) What was randomly assigned? **What unit was used for data analysis?**	Students were randomly assigned to a treatment group or a control group. Individual students and intervention groups were both used as units of analysis in the final data analysis.	**12-13**

Fidelity of Implementation		
NCRTI Question	**Answer**	**Page Number**
Q1) How was the program delivered?	*Words Their Way* was implemented in a small group format; Average group size: 5 students; Range of group size: 2 to 8 students	**21**
Q2) What was the duration of the intervention?	Average number of implementation weeks: 18.2; Average number of sessions per week: 3.8; Average duration of sessions: 20 minutes	**21**
Q3) What was the background, experience, training, and ongoing support of the instructors?	16 out of 23 teachers had received their Master's degree. Intervention teachers taught reading for an average of 11.8 years. They engaged in an initial and follow-up training sessions covering the *Words Their Way* program. (See Appendix F. Teachers were provided with trainers' contact information and asked to reach them if they ever had questions about the program.	**14**
Q4) How was the fidelity of treatment information obtained?	The fidelity of treatment was achieved by observing each classroom twice during the study as well as teachers completion of online weekly logs.	**21**
Q5) Provide documentation of fidelity of treatment implementation	The average number of minutes teachers were able to implement every week ranged from 74 to 138 minutes. Teachers were able to implement, on average, 104% of the recommend time each week. See Appendices H & I for more details	**21-22**
Measures		
NCRTI Question	**Answer**	**Page Number**
Q1) What is the proximal outcome measure?	See Table 8	**18-19**
Q2) What is the distal outcome measure?	AIMSweb R-CBM: A norm-referenced measure in which students read three passages aloud and a trained administrator recorded the number of words read correctly and the number of errors. The median score of the three passages was used as the overall score and the percentile was calculated from the overall score. This measure was administered at the beginning, middle and end of the year. Score type and range of measure: Raw Score: 0 – 199; Percentile Rank: 1 – 99 Reliability statistics: Test-retest: Grade 2 Fall-Winter (.93); Grade 2 Winter-Spring (.94); Grade 4 Fall-Winter (.95), Grade 4 Winter-Spring (.95)	**18-19**
Results		
NCRTI Question	**Answer**	**Page Number**
Q1) What analyses were used to determine whether the treatment group learned more than the control group?	A combination of hierarchical linear modeling (HLM), repeated measures ANOVA, and independent *t* tests were used to determine the difference between the control group and the treatment group.	**29**
Q2) What are the proximal and distal results?	See Table 19	**29**

Appendix E: Teacher Training Description

Teacher training was comprised of two distinct sections: research study orientation and product training. All participating sites participated in training at their own school sites prior to the start of study participation. Most training sessions occurred in early fall 2010, while a few sites were trained later in the fall.

Research Study Orientation: A representative from either the Cobblestone research team or a representative from the Pearson Academic Research team provided the study overview training to all participating treatment intervention teachers/ study liaisons. The research study orientation included a review of study activities, including timelines and procedures for pre/post testing and shipping back testing materials. The orientation also included collecting specific teacher information such as contact information, demographic information and signed teacher consent forms. Most study orientation sessions were held immediately prior to the product training sessions.

Product training: A Pearson representative (most with prior expertise in teaching language arts) conducted the product overview training for one full day during the first few weeks of the school year. Trainers were also previously trained on how to conduct teacher training in summer 2010 so training sessions would be consistent across study sites. Trainers used a power point presentation to review the program components and word study pedagogy. All trainers were familiar with product components and referred to the study implementation guidelines (see Appendix F) to ensure that intervention teachers were aware of the most critical components of the program to implement during the study. A follow up training was held with all study sites in which trainers visited individual schools a few weeks after the school year began to reinforce usage of program components and to identify any problems that teachers were having using the new program. During follow up sessions trainers also reviewed additional possible sorts or other classroom activities to use with *Words Their Way*. Trainers also provided their individual contact information for teachers to follow up with them directly if they had any questions about the program or specific components.

Appendix F: Implementation Guidelines

(Minimum of 20 minutes per day/5 days per week)

Day	Teacher Activity (Page #)	Student Activity (Page #)	Materials needed	Homework *(strongly recommended)*
1	**Teacher-Modeled Sort (38)**[1]	**Daily Student Sorting (39)**[2] Write the Sort (34)	Student Book Student Spiral Notebook Teacher Resource CD	Daily Student Sorting (34)
2		Daily Student Sorting (39) Draw and Label (51)	Student Book Student Spiral Notebook	Daily Student Sorting (39) Sentences (48)
3		Daily Student Sorting (39) Writing Sort (40) Buddy Sort (44)	Student Book Student Spiral Notebook	Daily Student Sorting (39) Memory (55)
4	Teacher-Modeled Word Hunt	Daily Student Sorting (39) Word Hunt (41)	Student Book Student Spiral Notebook Student Library Big Book of Rhymes	Daily Student Sorting (39) Word Hunt (41)
5	Informal Assessment[3]	Daily Student Sorting (39) **Glue the Sort (49)** Games (55)[4]	Student Book Student Spiral Notebook	None

Condense if necessary

1 The number in parentheses refers to the page number in the Teacher Resource Guide that provides instructions for each activity.

2 Bolded items are required.

3 An informal assessment can be conducted with a spelling check (see pg. 166 of the Teacher Resource Guide).

4 Games are optional and can be played when applicable.

Appendix G: Dosage of Treatment Implementation

Intervention Teacher	Intervention Logs		
	Average # of minutes/ week	% of Average minutes/ week	Total % of Implementation
1	100.6	101%	84%
2	120.3	120%	102%
3	137.5	138%	125%
	137.5	138%	125%
4	110.4	110%	84%
5	110.2	110%	118%
6	76.3	76%	85%
7	91.3	91%	95%
8	130.7	131%	140%
9	91.4	91%	77%
10	128.7	129%	136%
11	74.4	74%	79%
12	117.5	118%	113%
13	105.7	106%	122%
14	128.3	128%	141%
	120.8	121%	133%
15	118.0	118%	127%
16	75.0	75%	95%
17	90.0	90%	72%
18	115.0	115%	55%
	108.8	109%	60%
19	112.5	113%	125%
20	90.0	90%	96%
21	73.7	74%	78%
22	100.9	101%	98%
23	75.0	75%	66%
	76.5	76%	68%

Appendix H: Program Components Completed

Inter.	Grade	Total # of minutes	Degree of Implem. 1-9	Total % of Implem.	Emergent Early Letter Name	Letter Name	Within Word Pattern	Syllables and Affixes	Total # of Sorts
1	4	1610	7.2	84%	-	-	-	25-39	15
2	2	2165	5.3	102%	-	28, 34-41, 43-48	1-10	-	25
3	2	1650	6.6	125%	-	-	10-21	-	12
	4	1650	6.6	125%	-	-	10-21	-	12
4	2	1325	8.3	84%	-	-	11-21	-	11
5	2	3085	7.5	118%	-	1-44	8	-	45
6	4	1830	5.4	85%	-	-	8-31	-	24
7	4	1825	6.7	95%	-	-	1-4, 6, 8, 10-21	-	18
8	2	3005	7.5	140%	-	-	1-24	-	23
9	2 & 4	2010	7.9	77%	-	18	1-27	-	28
10	4	2445	6.7	136%	-	-	12-28	-	17
11	4	1340	5.4	79%	-	-	20-39	-	20
12	2	2350	7.5	113%	-	8-9, 20-21, 24-46	-	-	27
13	4	2008	6.0	122%	-	-	4, 6, 8-14, 16, 18-24	-	17
14	2	2565	5.4	141%	-	12, 13, 15-37	-	-	25
	4	2415	5.4	133%	-	-	3-26		24
15	4	2595	7.1	127%	-	-	2, 12-35	-	25
16	2	1575	7.2	95%	-	22-47	-	-	26
17	2	1260	7.3	72%	-	6-30	-	-	25
18	2	690	-	55%	-	12, 13, 19, 21, 23	-	-	5
	4	870	-	60%	-	-	2, 25-27, 30, 32, 34	-	7
19	4	2025	6.4	125%	-	-	11, 12, 14, 15, 17, 18, 20, 22, 24, 26-39	-	23
20	2	1890	8.1	96%	32-40, 43-45	25-34, 37	-	-	23
21	4	1695	5.5	78%	-	10-27	-	-	18
22	2	2320	6.3	98%	28-31	6-20, 22-25	-	-	23
23	2	1200	7.2	66%	32, 33	21, 31-36, 38-42, 45	-	-	13
	4	1300	7.2	68%		-	13-19, 21-25, 27, 28	-	14

Appendix I: Random-Intercept Models with Covariates

To estimate the program effect, we ran a series of parallel random-intercept models with covariates using STATA, which falls under two-level linear models in our case since we have students nested within classes (Rabe-Hesketh & Skrondal, 2008). All HLM models were ran in STATA (--xtmixed procedure).

A general linear random-intercept model with covariates can be represented as follows:

$$y_{ij} = \beta_1 + \beta_2 x_{2ij} + \ldots + \beta_p x_{pij} + \zeta_j + \varepsilon_i$$
$$= (\beta_1 + \zeta_j) + \beta_2 x_{2ij} + \ldots + \beta_p x_{pij} + \varepsilon_i$$

In the above model, y_{ij} refers to the outcome of student i in cluster (or class) j; x's refer to various students, teacher/class variables (i.e., covariates). The random intercept term (i.e., ζ) signals the linear model is of multilevel (two-level in our study) rather than simple OLS (ordinary least square) regression. We ran parallel models for each of the outcomes.

The following is a list of variables and their operational definitions associated with student background characteristics and teacher/ classroom/ school characteristics that were used in the HLM models.

Outcome variables:

1. MAT8 sound and print scaled scores posttest
2. MAT8 spelling scaled scores posttest

Student background characteristics variables:

1. Proxies for prior academic achievement
   - Prior MAT8 (sound and print, and spelling respectively) scaled scores pretest
2. Condition
   - Treatment (1)
   - Control (reference group)
3. Gender: female (1) and male (0)
4. Race indicators
   - African American (1)
   - Hispanic or Latino (1)
   - Other ethnicity (1)
   - White (reference group)
5. English language learner indicator
   - English language learner (1)
   - Not English language learner (reference group)
6. Grade level indicators
   - 4th (1)
   - 2nd (reference group)
7. Social economic status proxy measure
   - Subsidized lunch (1)
   - No subsidized lunch (reference group)
8. Disability indicator
   - Various disabilities (1)
   - No disability (reference group)

Teacher/classroom characteristics variables:

1. Grade by condition interaction
2. Levels of implementation (total number of minutes standardized using the sample mean and standard deviation)
3. Years of teaching experience (Number of years teaching)

Notes

For additional information about the *Words Their Way* Efficacy Study, please contact:
Dr. Rebecca Eddy
rebecca.eddy@cobblestoneeval.com
(909) 657-0518

2120 Foothill Blvd., Ste. 202 | La Verne, CA 91750 | (800) 971-3891