NCEE 2012-4013

U. S. DEPARTMENT OF EDUCATION

Effects of curriculum and teacher professional development on the language proficiency of elementary English language learner students in the Central Region

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

Effects of curriculum and teacher professional development on the language proficiency of elementary English language learner students in the Central Region

Final Report

April 2012

Authors:
Sheila A. Arens, Ph.D., Mid-continent Research for Education and Learning
Ginger Stoker, Ph.D., SRI
Jane Barker, M.P.A., Mid-continent Research for Education and Learning
Susan Shebby, M.Ed., Mid-continent Research for Education and Learning
Xin Wang, Ph.D., Mid-continent Research for Education and Learning
Lou F. Cicchinelli, Ph.D., Mid-continent Research for Education and Learning
Jean M. Williams, Ph.D.

Project Officer:
Sandra Garcia
Institute of Education Sciences

NCEE 2012–4013

U.S. Department of Education

U.S. Department of Education
Arne Duncan
Secretary

Institute of Education Sciences
John Q. Easton
Director

National Center for Education Evaluation and Regional Assistance
Rebecca A. Maynard
Commissioner

April 2012

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-06-CO-0023 by Regional Education Laboratory Central administered by Mid-continent Research for Education and Learning. The content of this publication does not necessarily reflect the views or policies of the Department of Education nor does mention of trade names, commercial products, or organizations imply endorsements by the U.S. Government.

Arens, S. A., Stoker, G., Barker, J., Shebby, S., Wang, X., Cicchinelli, L. F., & Williams, J. M. (2012). *Effects of curriculum and teacher professional development on the language proficiency of elementary English language learner students in the Central Region*. (NCEE 2012-4013). Denver, CO: Mid-continent Research for Education and Learning.

This report is available on the Institute of Education Sciences website at http://ncee.ed.gov and the Regional Educational Laboratory Program website at http://edlabs.ed.gov.

Alternate Formats Upon request, this report is available in alternate formats, such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at 202-260-9895 or 202-205-8113.

DISCLOSURE OF POTENTIAL CONFLICT OF INTEREST

This research was conducted by REL Central. No key staff at REL Central or members of the study's technical working group, who provided advice on key aspects of the study, have financial interests that could be affected by the findings.[1]

[1] Contractors carrying out research and evaluation projects for IES frequently need to obtain expert advice and technical assistance from individuals and entities whose other professional work may not be entirely independent of or separable from the tasks they are carrying out for the IES contractor. Contractors endeavor not to put such individuals or entities in positions in which they could bias the analysis and reporting of results, and their potential conflicts of interest are disclosed.

ACKNOWLEDGMENTS

This report reflects the culmination of design, data collection, and data analysis work begun in 2005.

This study and its report were made possible by a collaborative effort of school districts, schools, teachers, students, researchers, reviewers, technical advisors, and editors. The authors wish to thank the following individuals for their participation and assistance.

First, we thank all the school districts, schools, principals, teachers, and students who participated in the study for their commitment to the project, the time and effort they devoted to providing data, and their willingness to support our efforts to provide rigorous evidence of effectiveness. This study would not have been possible without their involvement.

Numerous individuals provided technical assistance and support in preparing this report. At REL Central, the authors thank Sandy Foster, Carrie Germeroth, Andrew Newman, and Deborah Mazzeo for site recruiting, site relations, and data collection; Bruce Douglas, Laurie Moore, and Mariana Enriquez-Olmos for contributions to the study design and instrumentation; Kirsten Miller for editorial review; Robyn Alsop for site relations, data collection, and editing; Jessica Rainey for data collection, editing, and administrative support; and Susie Bachler, Trudy Clemons, and Charles Igel for data collection.

REL Central's technical working group and reviewers at the Analytical and Technical Services contact administered by Mathematica Policy Research, Inc., provided advice and support throughout this study. Technical working group members included Geoffrey Borman, Susan Brookhart, William Clune, Kathy Escamilla, John Golden, Robin Morris, Barbara Plake, Andrew Porter, Robert St. Pierre, and Norman Webb.

Contents

SUMMARY X
Interventions studied x
Research questions xi
Study design xii
Analysis and results xiii
Conclusions and limitations xiii
1: STUDY BACKGROUND 1
Need for the study 2
Descriptions of the interventions 2
Conceptual model of the study 3
Prior research on interventions 4
Need for experimental evidence of effectiveness 6
Study design overview 6
Research questions 7
Organization of this report 8
2: STUDY DESIGN AND METHODOLOGY 9
Study timeline 9
Site recruitment 10
Random assignment 12
Study sample 13
School sample 13
Teacher sample 15
Student sample 19
Data collection 22
Teacher data 22
Teacher background surveys. 23
Online teacher logs. 23
Teacher observations. 24
Teacher interviews. 24
Student data 25
Use of the IDEA Proficiency Test (IPT) as the outcome measure 25
Attrition and response rates 27
Data analysis 29
Methods used to address missing data 30
Intent-to-treat estimates 31
Effect sizes 31
3: IMPLEMENTATION OF THE INTERVENTIONS 32
On Our Way to English 32
Program description 32
Program timing and delivery 33
Program training and implementation 33
Responsive Instruction for Success in English 34
Program description 34
Program training and implementation 34

Fidelity of implementation 35
Findings from teacher interviews 37
Responsive Instruction for Success in English (RISE) professional development 38
On Our Way to English curriculum and materials 38
Effect of the interventions on instruction 39
Continued use of the interventions 39
Findings from first teacher log 39
Differences between control and intervention groups 39
4: IMPACT ON STUDENT ENGLISH LANGUAGE ACQUISITION 43
Impact analyses 43
Sensitivity analyses 44
Case deletion of missing data (Model 3) 46
Analysis with no race/ethnicity and pretest covariates (Model 4) 46
Exclusion of race/ethnicity covariates (Model 5) 46
Subtest analyses (Models 6, 7, & 8) 46
Inclusion of eligibility for free or reduced-price lunch as a covariate (Model 9) 46
Exclusion of schools with low enrollment of ELL students (Model 10) 47
5: EXPLORATORY ANALYSES 48
Measuring intermediate outcomes 48
Teacher reports of student engagement 49
Teacher instructional practices 49
Teacher instructional responsiveness and assessment practices 49
Findings on exploratory research questions 50
6: SUMMARY OF FINDINGS AND STUDY LIMITATIONS 52
Fidelity of implementation 52
Impact of interventions on English language acquisition 52
Impact of interventions on intermediate outcomes 53
Study limitations 53
External validity or generalizability 53
Missing data 54
Self-report bias 54
APPENDIX A: DESCRIPTION OF INTERVENTIONS AND INSTRUCTIONAL MODELS USED IN PARTICIPATING SCHOOLS A-1
Interventions A-1
On Our Way to English A-1
Responsive Instruction for Success in English (RISE) A-2
ELL instructional programs and approaches A-4
Pull-out, small-group instruction A-4
Structured English immersion A-4
Sheltered instruction A-4
Dual-language/bilingual instruction A-4
Other types of instructional programs A-5
APPENDIX B: DESIGN ASSUMPTIONS AND POWER ANALYSIS B-1
APPENDIX C: TEACHER CHARACTERISTICS C-1
APPENDIX D: VALIDITY, RELIABILITY, AND SCORING OF LANGUAGE ACQUISITION OUTCOME MEASURE D-1

Validity D-1
Reliability D-1
Scoring D-1
APPENDIX E: SURVEY INSTRUMENTS E-1
Implementation measures E-1
APPENDIX F: IMPLEMENTATION OF THE INTERVENTIONS F-1
APPENDIX G: SURVEY SCALES G-1
Student engagement G-1
Teacher instructional practices G-2
Teacher instructional responsiveness/modification and assessment practices G-4
APPENDIX H: BASELINE EQUIVALENCE, CONFIRMATORY IMPACT, SENSITIVITY, AND EXPLORATORY ANALYSIS MODEL SPECIFICATIONS H-1
Model 1: specification of baseline equivalence model H-1
Model 2: specification of impact analysis model H-2
Models 3-10: specification of sensitivity analysis models H-3
Model 3: case deletion of missing data H-3
Model 4: analysis with no race/ethnicity and pretest covariates H-3
Model 5: exclusion of race/ethnicity covariates H-3
Models 6, 7 & 8: subtest analyses H-3
Model 9: inclusion of eligibility for free or reduced-price lunch as a covariate H-3
Model 10: exclusion of schools with low enrollment of ELL students H-3
Models 11-19: specification of exploratory analysis models H-4
APPENDIX I: VARIANCE COMPONENTS I-1
APPENDIX J: PRELIMINARY, IMPACT, AND SENSITIVITY ANALYSES J-1
APPENDIX K: EXPLORATORY ANALYSES K-1
REFERENCES REF-1

Tables

Table 2.1. Study timeline 9
Table 2.2. Number of schools in intervention and control groups in Colorado, Kansas, and Nebraska 14
Table 2.3. Pretreatment characteristics of intervention and control group schools 15
Table 2.4. Teacher characteristics at baseline, by treatment group 18
Table 2.5. Hours of professional development during five years before baseline, by treatment group 19
Table 2.6. Student characteristics impact analysis sample by treatment group 21
Table 2.7. Unadjusted IDEA Proficiency Test (IPT) pretest scores, by treatment group 21
Table 2.8. Data collected for study 22
Table 2.9. Response rate for IDEA Proficiency Test (IPT), spring 2010 posttest 27
Table 2.10. Teacher response rates to data collection instruments, by treatment group and sample 28
Table 3.1. Number of teachers falling into low to medium and high categories of Responsive Instruction for Success in English (RISE) modules completed 35
Table 3.2. Log questions used to determine teacher fidelity of implementation of On Our Way to English (OWE) and Responsive Instruction for Success in English (RISE) 36

Table 3.3. Fidelity with which On Our Way to English (OWE) was implemented by intervention teachers 37
Table 3.4. Weekly teacher-reported hours spent on English language development/literacy instruction, by treatment group for exploratory sample 38
Table 3.5. Teacher-reported ELL instructional models in exploratory sample in October of Year 2, by treatment group 40
Table 3.6. Teacher-reported instructional strategies in exploratory analysis sample, by treatment group 41
Table 4.1. Results from the confirmatory impact analysis model for basic composite score 44
Table 4.2. Sensitivity analyses for basic composite score 45
Table 5.1. Analysis results for intermediate outcomes 50
Table A1. RISE unit topics and instructional components A-2
Table C1. Characteristics of teachers in exploratory sample, by treatment condition C-1
Table C2. Characteristics of teachers in intervention group in baseline and exploratory samples C-2
Table C3. Characteristics of teachers in control group in baseline and exploratory samples C-3
Table F1. Training in Responsive Instruction for Success in English, by state location F-1
Table F2. Month materials received by intervention schools, by district F-1
Table F3. Month and year of training in On Our Way to English (OWE) in intervention schools, by district F-2
Table F4. Percentage of baseline study teachers attending Responsive Instruction for Success in English (RISE) training sessions during Year 1, by district and module F-3
Table F5. Teacher-reported fidelity of implementation F-4
Table G1. Difficulty of online teacher log items on student engagement G-1
Table G2. Difficulty of online teacher log items on differentiated instruction G-2
Table G3. Difficulty of online teacher log items on sheltering instruction G-2
Table G4. Difficulty of online teacher log items on receptive and expressive language instruction G-3
Table G5. Difficulty of online teacher log items on reading instruction G-3
Table G6. Difficulty of online teacher log items on writing instruction G-4
Table G7. Difficulty of online teacher log items on modification of instruction/teacher responsiveness G-4
Table G8. Difficulty of online teacher log items on student-centered instruction G-5
Table G9. Difficulty of online teacher log items on assessment practices G-5
Table I1. Variance components and intraclass correlation for student language acquisition I-1
Table I2. Variance components and intraclass correlation for listening, reading, and writing pretests I-2
Table I3. Variance components and intraclass correlation for intermediate outcomes I-3
Table J1. Model 1, Baseline equivalence model, preliminary analysis of unadjusted IDEA Proficiency Test (IPT) pretest data J-1
Table J2. Model 2, Confirmatory impact analysis model, adjusted IDEA Proficiency Test (IPT) posttest data J-2
Table J3. Model 3, Sensitivity analysis of adjusted IDEA Proficiency Test (IPT) posttest data with no imputation J-3
Table J4. Model 4, Sensitivity analysis of unadjusted IDEA Proficiency Test (IPT) posttest data with no race/ethnicity and pretest covariates J-4

Table J5. Model 5, Sensitivity analysis of adjusted IDEA Proficiency Test (IPT) posttest data with no race/ethnicity covariate J-5
Table J6. Model 6, Sensitivity analysis of adjusted IDEA Proficiency Test (IPT) posttest listening data J-6
Table J7. Model 7, Sensitivity analysis of adjusted IDEA Proficiency Test (IPT) posttest reading data J-7
Table J8. Model 8, Sensitivity analysis of adjusted IDEA Proficiency Test (IPT) posttest writing data J-8
Table J9. Model 9, Sensitivity analysis of adjusted IDEA Proficiency Test (IPT) posttest data using eligibility for free or reduced-price lunch as an additional covariate J-9
Table J10. Model 10, Sensitivity analysis of adjusted IDEA Proficiency Test (IPT) posttest data excluding schools with low enrollments of English language learner students J-10
Table K1. Model 11, Exploratory analysis, adjusted results of school-level outcomes on student engagement K-1
Table K2. Model 12, Exploratory analysis, adjusted results of school-level outcomes on differentiated instruction K-2
Table K3. Model 13, Exploratory analysis, adjusted results of school-level outcomes on sheltering instruction K-3
Table K4. Model 14, Exploratory analysis, adjusted results of school-level outcomes on receptive and expressive language instruction K-4
Table K5. Model 15, Exploratory analysis, adjusted results of school-level outcomes on reading instruction K-5
Table K6. Model 16, Exploratory analysis, adjusted results of school-level outcomes on writing instruction K-6
Table K7. Model 17, Exploratory analysis, adjusted results of school-level outcomes on modification of instruction/teacher responsiveness K-7
Table K8. Model 18, Exploratory analysis, adjusted results of school-level outcomes on student-centered instruction K-8
Table K9. Model 19, Exploratory analysis, adjusted results of school-level outcomes on assessment practices K-9

Figures

Figure 1.1. Theory of change 4
Figure 2.1. Teacher flow from Year 1 (training) to Year 2 (implementation) to exploratory sample 16
Figure 2 2. Student impact analysis sample 20

SUMMARY

English language learner (ELL) students who speak Spanish as their primary language represent the fastest-growing student group in the nation's public schools (U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics 2007; Planty et al. 2009; Editorial Projects in Education Research Center 2009) as well as in the Central Region.[2] Between 1997/98 and 2007/08, the number of ELL students rose more than 37.2 percent in Colorado, 171.1 percent in Nebraska and 127.6 percent in Kansas (U.S. Department of Education, Office of English Language Acquisition, Language Enhancement, and Academic Achievement for Limited English Proficient Students 2010a-d).

These demographic changes have led to an increased demand for teachers who are able to effectively address the needs of ELL students in classrooms (Hill and Flynn 2004) and ensure that ELL students have the same opportunity to learn as their native English-speaking peers (Herman and Abedi 2004). However, data from the National Center for Education Statistics (U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics 2002) reveal that the majority of teachers have not been trained in ELL-specific strategies. For example, in 2002 41 percent of teachers in the United States reported teaching ELL students but less than 13 percent reported receiving professional development related to the needs of this student subgroup. Teacher training and professional development in ELL-specific strategies have the potential to influence the skills and knowledge that teachers bring to their work—and ultimately improve student achievement.

Interventions studied

This study responds to regional and national needs by examining the impact on students' English language proficiency of a particular set of ELL-specific classroom materials in combination with a specific teacher professional development program. The classroom materials used in this study, entitled On Our Way to English (OWE), were authored by David Freeman, Yvonne Freeman, Aurora Colon Garcia, Margo Gottlieb, Mary Lou McCloskey, Lydia Stack, and Cecilia Silva and were published in 2003 by Rigby. According to the publisher, OWE is a comprehensive English curriculum for elementary classrooms (grades K–5) developed to provide ELL students with simultaneous access to English oral language development, comprehensive literacy instruction, and standards-based content area information in science and social studies (Freeman et al. 2003).

The professional development program, entitled Responsive Instruction for Success in English (RISE), was written by Clara Amador-Watson and published in 2004 by Harcourt Achieve. RISE is a professional development program designed to meet the needs of K–5 teachers by providing them with sustained adult learning opportunities to acquire the knowledge and skills to support ELL students in language and literacy learning. RISE is intended to be delivered to teachers in eight separate, core modules.

We conceptualized the intervention being examined as a dual intervention in which instructional resources and teacher professional development combine to improve student language

[2] Central Region States include Colorado, Kansas, Missouri, Nebraska, North Dakota, South Dakota, and Wyoming.

proficiency. OWE and RISE were selected as the intervention for this study for several reasons. First, OWE and RISE represent a popular option for districts seeking to meet the needs of ELL students in their classrooms. According to the publisher, approximately 1,400 districts across the United States have purchased OWE, RISE, or both interventions since the year 2000 (Harcourt Achieve, personal communication 2006). In addition, by 2008 Houghton Mifflin Harcourt had subsumed both Rigby and Harcourt Achieve (and therefore both the OWE and RISE products). We surmised schools and districts would be more likely to purchase ELL-specific classroom materials and teacher professional development from the same publisher. Finally, both products were specific to the K–5 level, corresponding with the grades targeted in this study.

Although there has been a proliferation of programs developed to meet the needs of ELL students, very little evidence on their effectiveness has been collected under controlled conditions. Some data suggest that OWE and RISE may be effective when implemented in isolation (Educational Research Institute of America [ERIA] 2004; Harcourt Achieve 2005), but little evidence is available regarding the combined effectiveness of the two programs. Despite this lack of evidence, schools across the country continue to adopt and implement curricula such as OWE and professional development such as RISE, promising improvement in ELL performance. Moreover, programs are often implemented without site-based or curriculum-linked professional development that is ongoing and proximal to practice. Rather, the professional development that typically accompanies a curricular intervention is a traditional workshop, short in duration and distal to teacher practice (Penuel et al. 2007).

This randomized trial is intended to generate unbiased estimates of the combined program effect on the English language proficiency of Spanish-speaking ELL students. To determine the effects of the combined OWE and RISE interventions on the English language achievement of ELL students in listening, reading, and writing, we assessed student progress with the IDEA Proficiency Test (IPT). This English language assessment, published in 2005 by Ballard & Tighe, measures student English language proficiency and progress.

Research questions

The study addresses one confirmatory research question and three exploratory questions. The confirmatory question addressed is:

- Does implementation of OWE in conjunction with the use of RISE have a significant impact on the acquisition of English language skills for ELL students as measured by the IPT composite score (based on subsection scores for listening comprehension, reading/vocabulary comprehension, and writing)?

The exploratory questions are:

- Does the combination of OWE and RISE have a significant impact on teacher-reported student engagement with ELL-specific educational materials?

- Does the combination of OWE and RISE have a significant impact on teacher-reported instructional practices (differentiated instruction, sheltering instruction, receptive and expressive language instruction, reading instruction, and writing instruction)?

- Does the combination of OWE and RISE have a significant impact on teacher-reported instructional responsiveness and assessment practices (modification of instruction or teacher responsiveness, student-centered instruction, and use of assessments)?

This study was designed as a randomized controlled trial conducted over two years. Year 1 (school year 2008/09) was conceptualized as a training year, during which OWE and RISE materials would be distributed to intervention group schools, intervention group teachers trained in OWE and RISE, and baseline data collected. Year 2 (school year 2009/10) was conceptualized as an implementation year, during which OWE and RISE would be implemented in intervention schools and fidelity and outcome data collected. Because the focus of this study was on the impact of the combined interventions, the study was not designed to investigate impacts of the two interventions separately.

Study design

Schools in the Central Region states with the largest percentages of Spanish-speaking ELL students in the elementary grades—Colorado, Kansas, and Nebraska—were recruited and randomly assigned using a 2:1 ratio to the intervention group (34 schools) or the control group (18 schools). Teachers in the intervention group were provided with OWE and RISE training and materials. In schools in the control group, English as a second language teachers in grades 1–5 used their existing strategies and materials in teaching ELL students. Schools were blocked before random assignment to ensure that each participating district contained both intervention and control schools and that the distributions of schools across districts were similar.

Conducting this randomized controlled trial over the course of two school years ensured that intervention teachers would be trained in use of materials before the study year. This level of exposure supports extended implementation of OWE and RISE and reduces any effects associated solely with the introduction of a new program. During Year 1, the training year, the publisher provided professional development for teachers in the intervention group on the use of OWE in their classrooms and intervention teachers were trained in the use of RISE strategies. Materials for both programs were distributed during this year. During Year 2, the implementation year, teachers in the intervention group were asked to use OWE each day for at least 30 minutes per class and implement the RISE techniques fully in their classrooms.

To understand the instructional practices used with ELL students at these sites, we examined online teacher logs, site coordinator surveys, and interview data. The results revealed that the instructional practices used with ELLs by intervention and control teachers were similar, although teachers in the control group were significantly more likely than teachers in the intervention group to self-report using graphic organizers with their ELL students. Between-group variance in the number of teacher-reported hours of English language development[3] instruction with ELL students was not statistically significant.

[3] English language development instruction is also known as English as a second language instruction.

Analysis and results

Confirmatory impact analyses estimated the mean difference between intervention and control groups on student acquisition in English, as measured by a composite of the listening, reading, and writing sections of the IPT (*IPT Testing System* 2005). All native Spanish-speaking ELL students receiving instruction from a study teacher on the day of testing in spring 2010 (Year 2) were included in the student sample. This sample included 2,612 students nested within 52 schools. Intervention and control groups were compared to determine whether the groups differed on school-level characteristics, including the percentage of students eligible for free or reduced-price lunch, school size, the percentage of students in different racial/ethnic groups, and location. A statistically significant difference was found between the percentages of White and Hispanic students; this variable was included as a covariate in the impact analysis model.

The impact analysis revealed no statistically significant difference between the composite IPT scores of students in the intervention group (who were taught by teachers trained in RISE and who used OWE materials) and students in the control group. The exploratory analyses revealed no statistically significant differences between the intervention and control groups. The combination of OWE and RISE did not have a significant impact on teacher-reported student engagement in ELL-specific educational materials, teachers' self-reported instructional practices, or teachers' self-reported instructional responsiveness and assessment practices.

Conclusions and limitations

The study found no statistically significant difference between the IPT scores of students in the intervention and control groups. Results of sensitivity analyses revealed that this finding was invariant to the inclusion of covariates in the analytic model, the method used to treat missing data, and the use of a composite score. Several limitations should be considered when interpreting the study findings:

- This study relied on a volunteer sample of schools that may not be representative of schools across the United States.

- Findings apply only to similar implementation scenarios: although a newer version of OWE was released in 2010, the 2004 edition was used in the study; the study employed a train-the-trainer model for RISE (other schools may elect to have their staff trained directly by the publisher); and teachers were instructed to use OWE materials at least 30 minutes a day with their ELL students but were allowed to vary from that recommendation, as might be likely under real-world conditions.

- Study findings are applicable only to the acquisition of English listening, reading, and writing by Spanish-speaking ELL students in grades 2–5. Moreover, only one outcome measure (the IPT) was used to assess student gains. Use of additional student outcome measures might have yielded findings that would have provided additional insight on the effectiveness of the intervention.

- The impact analysis sample was defined as Spanish-speaking ELL students enrolled in study schools on the day of testing in spring 2010. Students defined as ELL students at the start of the 2009/10 school year who tested out of English as a second language services during the school year were excluded from the analysis.

- This study was not designed to have the statistical power to detect effects smaller than 0.35 standard deviation.

- Most teachers in the intervention group received their RISE training from a publisher-trained site coordinator rather than directly from the publisher. Results might have differed had training been provided by the publisher.

- Implementation analyses relied upon self-reported measures of teacher training, professional qualifications, classroom environment, and classroom use of instructional strategies. Responses to these items may have been influenced by self-report biases.

1: STUDY BACKGROUND

English language learner (ELL) students are the fastest-growing student group in the nation's public schools (U. S. Department of Education, Institute of Education Sciences, National Center for Education Statistics 2007). The number of school-age children is projected to increase by 5.4 million between 2005 and 2020, with most of the increase among children of immigrants (Fry 2008). Between 1997/98 and 2007/08, the number of ELL students rose more than 37.2 percent in Colorado, 171.1 percent in Nebraska and 127.6 percent in Kansas (U.S. Department of Education, Office of English Language Acquisition, Language Enhancement, and Academic Achievement for Limited English Proficient Students 2010a-d). The majority of these ELL students speak Spanish as their primary language (Planty et al. 2009; Editorial Projects in Education Research Center 2009) and are enrolled in the poorest districts in the region.

This growth has implications for student achievement, as ELL students consistently score lower than their non–ELL peers on standardized assessments in mathematics and reading. In 2009, for example 43 percent of ELL students scored below the basic level on the grade 4 National Assessment of Educational Progress (NAEP) mathematics assessment compared with 16 percent of non–ELL students (U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics 2010a). On the NAEP reading assessment, 71 percent of ELL students and 31 percent of their non-ELL students scored below the basic level (U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics 2010b). Like their peers across the nation, ELL students in the Central Region states (Colorado, Kansas, Missouri, Nebraska, North Dakota, South Dakota, and Wyoming) perform below their non–ELL counterparts on the National Assessment of Educational Progress (Slavin and Cheung 2003).

The growing number of ELL students in U.S. schools and persistent achievement gaps have led to increased demand for teachers who are able to effectively address the needs of ELL students (Educational Research Institute of America [ERIA] 2004; Herman and Abedi 2004; Hill and Flynn 2004). The Elementary and Secondary Education Act (also known as the No Child Left Behind Act) requires that all subgroups of students, including ELL students, achieve at the same high levels. Part A of the English Language Acquisition, Language Enhancement, and Academic Achievement Act of Title III[4] of the No Child Left Behind Act of 2001 is designed to:

> help ensure that children who are limited English proficient, including immigrant children and youth, attain English proficiency, develop high levels of academic attainment in English, and meet the same challenging State academic content and student academic achievement standards as all children are expected to meet (sec 3102).

[4] Title III of the No Child Left Behind Act provides federal financial support to eligible schools for instruction of both English Language Learners and immigrant students.
http://www.ncela.gwu.edu/files/uploads/5/NCLBTitleIII.pdf

Need for the study

Meeting the needs of ELL students in the Central Region is a growing challenge. The Central Region is experiencing an increase in the population of ELL students and a corresponding shortage of teachers qualified to meet the demands of this student population. ELL students are more likely to be placed in classrooms with teachers who are not qualified or fully credentialed to teach them (Rumberger and Gandara 2004). Competency requirements enacted under the No Child Left Behind Act (Public Law No. 107-110, 115 Stat. 1425 2002) to ensure that qualified teachers are available to meet the increasing enrollment of ELL students have led to challenges for districts in identifying and hiring qualified teachers (Center on Education Policy 2006). In 2002, 41 percent of teachers in the United States reported teaching ELL students, but less than 13 percent reported receiving any training or professional development in doing so (U.S. Department of Education, Institute of Education Sciences, and National Center for Education Statistics 2002). Seventy-three percent of teachers who reported participating in more than eight hours of professional development addressing the needs of ELL students reported having improved their classroom teaching "moderately" or "a lot"; among teachers who reported participating in less than eight hours of professional development, 39 percent reported improvements in classroom teaching (Fry 2008). These findings suggest that professional development may enhance the skills teachers bring to their work—and ultimately affect student achievement.

In terms of the demand for qualified teachers, based on current projections, the region will need an additional 6,273 ESL-certified teachers within the next five years, with the majority needed in Missouri (3,285) and Colorado (2,500) (Editorial Projects in Education [EPE] Research Center 2009). This projected growth in the ELL student population and corresponding need for ESL-certified teachers suggests the need to identify professional development and curricula that will improve the educational outcomes of this student subgroup.

Descriptions of the interventions

The study design was conceptualized to test the effectiveness of a dual intervention, comprising an ELL-specific student curriculum and an aligned teacher professional development program, in improving the English language proficiency of Spanish-speaking ELL students. The classroom materials used in this study, entitled On Our Way to English (OWE), were authored by David Freeman, Yvonne Freeman, Aurora Colon Garcia, Margo Gottlieb, Mary Lou McCloskey, Lydia Stack, and Cecilia Silva and were published in 2003 by Rigby. The professional development program, entitled Responsive Instruction for Success in English (RISE), was written by Clara Amador-Watson and published in 2004 by Harcourt Achieve.

OWE is a curriculum for K–5 ELL students developed to provide simultaneous access to English oral language development, comprehensive literacy instruction, and standards-based content area information in science and social studies (Freeman et al. 2003). OWE curricular materials include books, posters, picture cards, and audio tapes/CDs designed to be interesting and accessible to students with different levels of English ability at each grade level. OWE teacher resources include a teacher's guide and student monitoring and assessment resources (chapter 3 and appendix A include more detailed descriptions of OWE). We surmised that OWE would

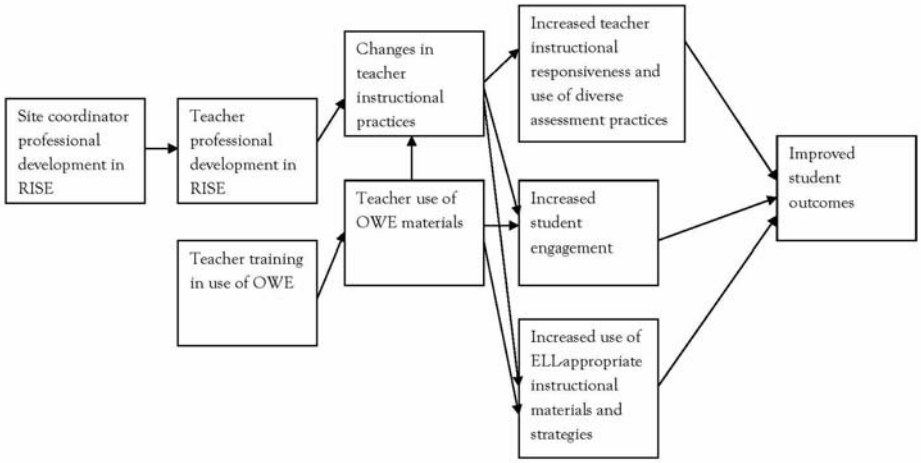
provide intervention teachers with a selection of instructional materials appropriate for students' age and level of English proficiency not available to control condition teachers.

RISE was selected for this study in order to provide teachers with foundations in language acquisition and communicative competence, which form the foundation for ELL instruction (Téllez and Waxman 2005). RISE professional development provides teachers with second language acquisition theory as well as practical information on strategies to incorporate in ELL instruction, such as use of formative assessment practices and small group instructional techniques. RISE teacher materials include resource notebooks for each of eight RISE modules, a data collection booklet, and forms for lesson planning. (For additional discussions of RISE unit topics and teacher training activities, see chapter 3 and appendix A.) We posited that use of RISE, a professional development program for ELL teachers, would provide intervention teachers with the professional knowledge and skills to help them address the needs of ELL students. All intervention group teachers agreed to participate in RISE training to foster school-wide understanding of RISE strategies.

Conceptual model of the study

The study is based on a conceptual model that links improvements in student English language acquisition to teacher professional development in instructional strategies relevant to ELL students and a comprehensive English language development curriculum (figure 1.1). Although the model depicts mediated effects on student outcomes, these mediated effects are not tested in the current study. In this conceptual model, intervention teachers are trained in RISE content and strategies. Through this training, teachers become familiar with the research on second language acquisition and literacy development, as well as strategies to support the instruction of ELL students. Teachers then apply this knowledge to their instructional practice through use of OWE materials, using, for example, leveled readers (books categorized by difficulty level) to individualize instruction. Such changes are related to improved teacher responsiveness and the use of diverse assessment practices, increased student engagement, and increased use of ELL appropriate instructional materials and strategies, ultimately resulting in improved student outcomes. This model aligns with Desimone's (2009) conceptual framework, in which teacher professional development experiences lead to increased teacher knowledge, changes in teacher practice, and ultimately improved student academic outcomes. This model also includes the effect of teacher use of a comprehensive curriculum for students.

Figure 1.1. Theory of change



OWE was selected to support both teacher and student outcomes. Access to OWE materials provides teachers with increased opportunity to use instructional materials and strategies that are appropriate for ELL students. These resources support the adoption of instructional practices learned through RISE, including comprehensive thematic units, lesson plans, and materials to support all levels of language acquisition at each grade level.

OWE was designed to provide students with engaging and challenging learning experiences. We therefore collected data on teacher perceptions of student engagement, exploring the theory that students who are engaged in classroom activities will demonstrate increased achievement (Dweck and Leggett 1988).

Prior research on interventions

Studies of the effectiveness of OWE have yielded mixed results. No studies have been conducted on grades 1, 3, or 5, and none of the research on other grades would be considered rigorous. No causal evidence exists regarding the effectiveness of RISE, and no research has been conducted on the impact of using OWE program materials in conjunction with the RISE professional development course.

In a four-month, one-group study of OWE (ERIA 2004), researchers assessed pretest to posttest achievement gains for students in grade 2 (n = 36) and grade 4 (n = 81–85, depending on the test section examined) using the total score and the scores on the reading, writing, vocabulary, listening comprehension, and retelling subscales of the Language Assessment Scales (*LAS Links K–12 Assessments* n.d.) published by CTB/McGraw Hill. Grade 2 students made statistically significant gains on the total standard score and on the subscales for reading, writing, listening

comprehension, and retelling; they did not make significant gains on the vocabulary subscale. Effect sizes for significant gains ranged between .40 and .59. Results for grade 4 students were mixed, with significant gains found for the total scale and the subscales for writing, listening comprehension, and retelling but not on the reading or vocabulary subscale. Effect sizes for significant gains ranged between .28 and .69. Given that the study design did not include a comparison group, internal validity threats such as maturation are of concern.

A follow-up study (Marketing Works 2005) found similar results for grade 2 students (n = 31–35, depending on the test section examined) in Nevada, with statistically significant gains over four months for all areas assessed (the total score and the subscales for reading, writing, vocabulary, listening comprehension, and retelling). For grade 4 students (n = 8), statistically significant gains were found for writing and vocabulary but not for the total score or the reading, listening comprehension, or retelling subscales. These results may have been influenced by the small sample size as well as by other serious concerns regarding internal validity (maturation) resulting from the lack of an equivalent comparison group.

In a randomized controlled trial in which teachers were randomly assigned to use OWE materials, researchers examined reading achievement gains made by students in grade 2 (n = 89) and grade 4 (n = 80) in bilingual and English immersion ELL programs (Harcourt Achieve 2005). The STAR reading assessment (Renaissance Learning 2003) and the oral portion of the IDEA Proficiency Test (IPT) (Ballard & Tighe 2001) were used as outcome measures. The IPT is an English language assessment that measures student English language proficiency and progress. Researchers found no statistically significant differences in STAR reading achievement between students in either grades 2 or 4 who received the OWE curriculum and those who did not. They found an interaction between study condition and implementation environment—English immersion versus bilingual classrooms—for grade 2 students, suggesting that the structure of instruction delivery may make a difference (Harcourt Achieve 2005). This interaction was not found for grade 4 students.

Findings from the oral (speaking) IPT were provided only in percentages of students falling into various categories of English-speaking ability. Positive trends were found for students in grades 2 and 4. In grade 2, 16 percent of students in intervention classrooms and 19 percent of students in control classrooms were considered fluent in English at pretest. By posttest, this margin had widened, with 44 percent of students in intervention classrooms and 35 percent of students in control classrooms considered fluent in English. In grade 4, 26 percent of students in intervention classrooms and 18 percent of students in control classrooms were considered fluent in English at pretest. By posttest, this margin had widened, with 74 percent of students in intervention classrooms and 44 percent of students in control classrooms considered fluent in English. Positive trends were more pronounced in English immersion classrooms than bilingual classrooms. However, researchers failed to take into account of issues of nesting for their analyses, and student attrition for the study was 43 percent.

There is no experimental evidence on the impact of RISE.

Need for experimental evidence of effectiveness

RISE courses have been offered in every state in the country. Since 2000 an estimated 3,400 schools from more than 1,400 districts across the United States have purchased RISE, OWE, or both (Harcourt Achieve, personal communication 2006).

Although numerous curricula have been developed to meet the needs of ELL students, very little evidence relating to their effectiveness has been collected under controlled conditions. Despite this lack of evidence, schools across the country continue to adopt and implement curricula that promise improvement in ELL performance. Programs are often implemented without site-based or curriculum-linked professional development that is ongoing and proximal to practice; rather, the only professional development that typically accompanies a curricular intervention is a traditional workshop that is short in duration and distal to teacher practice (Penuel et al. 2007).

The use of OWE and RISE represents a popular option for educators seeking to meet the needs of ELL students in their districts, through both a curricular intervention and an aligned, longer-duration professional development intervention. By the publisher's estimates, about 8 percent of U.S. school districts have already adopted these interventions, alone or in combination (Harcourt Achieve, personal communication, September 25, 2006). The widespread adoption of OWE and RISE in conjunction with the incomplete and inconsistent results from studies on the effectiveness of the interventions indicate a need for research on their effectiveness.

Study design overview

To investigate the impact of OWE in combination with RISE, the study employed a randomized-controlled trial design based on voluntary school participation. Fifty-two elementary schools from 13 districts in three Central Region states (Colorado, Kansas, and Nebraska—the states in the REL Central region with the highest enrollments of ELL students) were randomly assigned to the intervention or control group.[5] At each school, a site coordinator was selected and given the responsibility of coordinating all study activities. Thirty of the 52 site coordinators also participated in the study as study teachers.

To determine the teacher sample, site coordinators were asked to identify all English as a second language teachers who instructed native Spanish-speaking ELL students in grades 1–5. At sites without a designated English as a second language teacher, the research team asked site coordinators to identify all literacy teachers who taught native Spanish-speaking ELL students in grades 1–5. These teachers were subsequently asked to voluntarily participate in the study. The number of teachers identified at each school varied based on school enrollment, the size of the native Spanish-speaking ELL population, and the school's instructional model.[6] (In order to avoid the variability that might have been introduced by studying the impact of the interventions on different languages, the student sample was limited to native Spanish-speaking students, the

[5] Fifty-three schools were randomized, but one intervention school dropped out of the study immediately after randomization because of a change in school leadership (the new principal did not feel the school was equipped to participate in the study). No data were collected from this school.

[6] For example, schools using a pull-out model for ESL identified only one study teacher, whereas schools using a transitional bilingual model reported at least one teacher per grade level.

largest group of ELL students in the United States and the Central Region.[7]) After baseline testing was conducted, we learned that the publisher would provide only one set of OWE curricular materials per grade level to each school.[8] Schools in the intervention group were provided with OWE and RISE training, one set of RISE professional development materials per grade level, one set of OWE materials per grade level, and one teacher's guide per study teacher. In schools assigned to the control group, teachers in grades 1–5 used their existing strategies and materials for teaching ELL students.

This randomized control trial was implemented over the course of two years, with the first year considered a training year. Conducting the experiment over two years allowed teachers in the intervention group to be exposed to the RISE professional development program for a full year before implementation as recommended by Harcourt Achieve. Designing the study with a training year and an implementation year helped ensure that the study estimated the impact of the intervention, not just the introduction of and training on a new program. During Year 1, teachers were also trained in the use of OWE materials and received OWE materials.

The measure of English language acquisition selected for the current study is the same measure that was used in some prior research – the IDEA Proficiency Test (IPT; Ballard & Tighe 2001, 2005). Students took the IPT at the beginning of Year 1 (the 2008/09 school year) of the study and again at the end of Year 2 (2009/10). The student impact analysis sample comprised native Spanish-speaking ELL students in grades 2–5 who were placed with a study teacher in spring 2010 (Year 2) and whose parents consented to their participation. IPT testing was conducted in grades 1–4 during Year 1 to provide a baseline pretest for the student impact analysis sample. The impact of OWE and RISE on students' English language acquisition was estimated using a hierarchical linear model to account for the clustered structure of the data and school-level random assignment. (Details of the study design and student impact estimation are in chapter 2.)

Research questions

This experimental study evaluates the effectiveness of the OWE program for ELL students in conjunction with the RISE program for teachers in developing student listening, reading, and writing skills in English. The study addresses one confirmatory research question and three exploratory questions. The confirmatory question is:

- Does implementation of OWE in conjunction with the use of RISE have a significant impact on the acquisition of English language skills for ELL students as measured by the IPT composite score (based on scores for listening comprehension, reading/vocabulary comprehension, and writing subtests)?

[7] For example, there is a wide variance in writing conventions. ELL students whose native language includes writing that is largely symbolic have different experiences and challenges in learning to form words in written English.

[8] At schools where multiple teachers were identified per grade level, we asked pairs of teachers to share OWE curricular materials. At some schools, site coordinators indicated that it would be a hardship to share materials or implement an instructional intervention in only two classrooms. When requested by intervention site coordinators, we provided one extra strand of the OWE curriculum (the oral language development strand) for each grade level in which more than two teachers instructed native Spanish–speaking ELL students. To address these requests, we purchased five grade 2, four grade 3, seven grade 4, and four grade 5 strands. In other cases, teachers unable to share materials exited the study. We present the results of attrition analyses in chapter 2.

Because the focus of this study was on the impact of the combined interventions, the study was not designed to investigate impacts of the two interventions separately. We collected data from teachers on pedagogical practices and the materials they use in their classrooms. These data provided contextual information about differences in intervention and control group teachers' behaviors, skills, and resources related to teaching ELL students.

The data also provided a means of exploring the hypothesized impact of OWE and RISE on teacher-reported student engagement, instructional practices, and responsiveness and modification of instructional practices and assessment practices. The exploratory research questions are:

- Does the combination of OWE and RISE have a significant impact on teacher-reported student engagement with ELL-specific educational materials?

- Does the combination of OWE and RISE have a significant impact on teacher-reported instructional practices (differentiated instruction, sheltering instruction, receptive and expressive language instruction, reading instruction, and writing instruction)?

- Does the combination of OWE and RISE have a significant impact on teacher-reported instructional responsiveness and assessment practices (modification of instruction and teacher responsiveness, student-centered instruction, and assessment use)?

Organization of this report

Chapter 2 describes the study sample, the study design, data collection, and the impact estimation approach. Chapter 3 describes implementation of both interventions. Chapter 4 presents estimates of the impact of OWE and RISE on primary outcomes. Chapter 5 presents the exploratory analyses. Chapter 6 summarizes the study's key findings and discusses its limitations.

2: STUDY DESIGN AND METHODOLOGY

This chapter describes the study timeline and the recruitment and randomization of participating schools; summarizes the characteristics of participating schools, teachers, and students; and details the measures employed to assess the effectiveness of the intervention. It also provides an overview of the data quality and the methods used to analyze the data.

Study timeline

During Year 1 of the study (the 2008/09 school year), teachers in intervention schools received training in both OWE and RISE and were provided with the materials for both interventions (see study timeline, table 2.1). Implementation of OWE and RISE occurred in Year 2 of the study (the 2009/10 school year) at the intervention schools. Control schools used their existing materials for teaching ELL students and agreed to not purchase or use the study materials during the course of the study.

Table 2.1. Study timeline

Phase	Date	Activity
Recruitment and randomization	November 2007–June 2008	District and school recruitment
	August 2008	Consent secured (district, school, and teacher memoranda of understanding)
	August 2008–September 2008	Random assignment of schools
Year 1 (training)	September 2008–October 2008	RISE training for intervention site coordinators
	October 2008–May 2009	RISE training for teachers in intervention schools
	November 2008–May 2009[a]	OWE training for intervention site coordinators and teachers
	October 2008–February 2009	Administration of student IPT pretest
Year 2 (implementation)	August 2009–May 2010	OWE and RISE training for new teachers provided by site coordinators
	February 2010–May 2010	Administration of student IPT posttest

Note: OWE is On Our Way to English. RISE is Responsive Instruction for Success in English. IPT is a series of language proficiency assessments for English language learners in K–12.

a. Because of publisher delays in shipping materials, several sites did not receive OWE training until the end of Year 1.

Site recruitment

The target sample for this study was determined through a multistage selection process. Because the study focuses on ELL-specific interventions aimed at elementary school-age students, the sample was limited to states in the REL Central region with high percentages of such students.[9] The focus of OWE and RISE on grades K–5 is consistent with the body of research documenting an increased rate of language acquisition and improved level of ultimate attainment when students begin to develop a second language before puberty (Long 1990; Patkowski 1982; Singleton and Ryan 2004). In order to avoid the variability that might have been introduced by studying the impact of the interventions on different languages, analyses were limited to native Spanish-speaking students, the largest group of ELL students in the United States and the Central Region.

For each state in the Central Region, we compiled a list of public schools serving grades 1–5 from publicly available sources, such as the Common Core of Data (U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Common Core of Data 2005-2009) and State education agency websites and staff. Lists of public elementary schools that had expressed interest in the OWE and RISE programs to the publishers but had not yet adopted either program were obtained from the publisher for each target state. By examining these sources, we identified schools in which more than 20 percent of the student population were Spanish-speaking ELL students (the cutoff was originally set at 33 percent but adjusted to 20 percent during recruiting). These schools were considered eligible for participation in the study unless teachers in the schools had already been exposed to RISE or previously used the OWE classroom materials.[10] Schools with transience rates exceeding 50 percent were also eliminated.[11]

Colorado, Kansas, and Nebraska had the largest number of eligible schools in the Central Region. These states were therefore targeted for recruiting. Fifty-five districts and 248 schools were identified through this process. We met with the directors of English language acquisition in each of these states to secure their support for conducting the study and to verify the lists of eligible schools.

---

[9] ELL students are defined differently by different states. Colorado defines ELL students as students whose dominant language is not English and whose academic achievement is "impaired" because of their inability to comprehend or speak English adequately. The Nebraska Department of Education defines an ELL student as a student 3–21 enrolled or preparing to enroll in an elementary or secondary school who meets one or more of the following criteria: was not born in the United States or whose native language is a language other than English; is Native American, Alaskan Native, or a native resident of the outlying areas; comes from an environment in which a language other than English has a significant impact on the level of English language proficiency; or is migratory and whose native language is a language other than English. Given the variations in definitions across states, for purposes of this research, we adopted the definition of an ELL student in use in the state from which the sample was derived.

[10] Although all sites projected an ELL enrollment of at least 20 percent in the 2008/09 school year, in the fall of 2008, eight sites (15.3 percent) indicated ELL enrollment of less than 20 percent. A sensitivity analysis was conducted by dropping these eight schools from the impact analysis data (see chapter 4).

[11] The following equation (from Dobson and Henthorne 1999) was used to estimate transience in cases where a school did not have this information: [(pupils joining school + pupils leaving school) x 100]/total school enrollment.

Once the three states and the pool of eligible schools were identified, we sent letters to the three state commissioners of education to provide details about the study, share the names of the eligible schools, and gain state-level support. We then contacted school administrators to describe the study, explain the benefits of participating in it, verify school eligibility, and gain support and approval for participation. Schools also had to be willing to accept the study requirement that the school could be assigned to either the intervention or control group and agree to allow REL Central to test all of their ELL students, observe participating teachers' classrooms, and interview participating teachers. Schools were also required to provide REL Central with access to student data.

District and school officials from 17 districts with eligible schools met with REL Central staff, who provided presentations for district- and school-level staff members charged with the instruction or oversight of instruction of ELL students. Districts and schools unable to attend informational meetings were contacted by e-mail and phone to gauge their interest in participation. Follow-up e-mails and phone calls were used to verify interest and secure district and school agreement to participate in the study.

During the recruitment process, teachers learned about the study purpose and requirements for participation from district and school staff who had attended informational meetings. After school-level memoranda of understanding (MOUs) had been signed, teachers identified as eligible by their school site coordinator also received study information from REL Central staff.[12] Eligible teachers who were interested in participating in the study were required to agree to the specifics of the study, as reflected in signed MOUs. In addition, teachers in schools randomly assigned to the intervention group had to agree to use OWE materials in their classrooms for at least 30 minutes a day and to be trained in the RISE professional development.

During the recruitment phase, districts and schools were informed about several incentives for participation. Incentives for the intervention schools included a $1,000 participation stipend and OWE and RISE training and materials. Incentives for control schools included a $5,000 resource balancing stipend and OWE and RISE training and materials after the conclusion of the study.[13] Site coordinators received stipends of $250 each from REL Central and training in RISE from the publisher.

[12] Teachers identified as eligible by site coordinators received an introductory e-mail with study information from their school's REL Central study liaison and a binder with study information in early fall 2008.

[13] Intervention and control group schools received resources during the intervention period. Therefore, any intervention effects identified in this study were attributed to the use of OWE and RISE rather than to the increase in resources for intervention schools.

Random assignment

Random assignment of schools to study conditions enabled the research team to draw inferences about the effectiveness of the combination of OWE and RISE. To obtain the needed statistical power to address the primary research question and to allow for potential school attrition of about 25 percent, we determined that at least 48 schools were needed for the study.[14] Appendix B includes detailed power analyses and assumptions.

Fifty-three elementary schools from 13 districts were recruited and randomly assigned to either the intervention or control condition. Prior to random assignment, schools were blocked by district.[15] Random assignment of schools was done within each participating district after district and school administrators formally agreed to participate by signing the memorandum of understanding. In this way, we could be certain that each district block contained both treatment and control schools.[16] We assigned schools using a 2:1 intervention to control ratio (that is, the chance of being assigned to an intervention school was twice that of being assigned to a control school).[17] In cases in which only one school from a district participated in the study, schools were blocked by pseudo-district (a group of similar schools). In each block, schools were assigned a number from 1 to *n*. The research team used the "select cases, random sample of cases" procedure in the Statistical Packages for the Social Sciences (SPSS) software to select 66 percent of each district block. We assigned these schools to the intervention condition; schools not selected were assigned to the control condition.

One school from the District 3 intervention group dropped out of the study shortly after randomization. No data were collected from this school.

All four schools in District 2 withdrew from the study at the beginning of Year 2. The decision to withdraw was made at the district level as a result of turnover in district-level leadership and a resultant change in approach to the instruction of ELL students. We collected student outcome data from these schools in Years 1 and 2, and subsequent analyses and discussions of student outcomes include these data. However, teachers at these schools did not complete the data collection instruments in Year 2. Therefore, these schools were not included in the exploratory analyses of teacher outcomes.

[14] This power analysis assumed an effect size of at least 0.35, a midpoint intraclass correlation coefficient of 0.10, a two-tailed test with $p < .05$, power greater than 0.80, and the inclusion of covariates ($R^2 = 0.50$).

[15] Blocking by district enabled us to statistically take into account potential variations between districts in ELL policies and practices.

[16] Not all schools in a district were eligible or agreed to participate.

[17] Increasing the odds that a school could be randomly assigned to the intervention condition was employed as a recruitment strategy. Given the assumptions detailed in the power analysis, uneven assignment to conditions was assumed to have a minimal effect on the number of schools needed. With an $R^2 = 0.50$ and an expected effect size of 0.35, the difference in the number of schools needed for a balanced versus an unbalanced allocation to the treatment groups was about four. We oversampled by 25 percent to accommodate for the unbalanced assignment of schools to study groups as well as to account for attrition.

Study sample

This section describes the composition of the school sample, the teacher sample, and the student impact analysis sample.

School sample

Random assignment theoretically yields two groups that do not differ at baseline on key characteristics. To check to see whether this was the case, the research team compared the intervention and control groups to determine whether the groups differed on the following school-level characteristics: the percentage of students eligible for free or reduced-price lunch, the number of students enrolled in the school, the percentage of students in different racial/ethnic groups, and the location of the school. Table 2.2 provides details on the assignment of schools to treatment conditions.

Table 2.3 compares the characteristics of the two groups of schools at baseline. A statistically significant difference was found between the two groups of schools on race/ethnicity, specifically on percentage of Hispanic and percentage of White students.[18]

Table 2.2. Number of schools in intervention and control groups in Colorado, Kansas, and Nebraska

State	Intervention group	Control group	Total
Colorado	21	10	31
Kansas	7	4	11
Nebraska	7	4	11
Total	35	18	53

Note: All four schools in District 2 dropped out of the study at the beginning of Year 2. We collected student data from these schools in the form of enrollment records and IPT scores. We also collected teacher background surveys, but were unable to collect data in the form of teacher logs. One intervention school in District 3 withdrew from the study shortly after randomization. No student or teacher data were collected from this school.

[18] Although the difference between groups eligible and ineligible for free or reduced-price lunch was not statistically significant ($p<.27$), the difference between groups was considered important enough to include as a covariate in a sensitivity analysis.

Table 2.3. Pretreatment characteristics of intervention and control group schools

Characteristic	Intervention group (n = 34) Mean	Intervention group (n = 34) Standard deviation	Control group (n = 18) Mean	Control group (n = 18) Standard deviation	Difference	Test statistic	p-value
Eligibility for free or reduced-price lunch (percentage of students)	75.86	16.94	68.91	23.20	6.95	1.12	.27
School size (number of students)	422.24	136.54	433.94	156.58	–11.70	–0.27	.79
Race/ethnicity (percentage of all students)							
African American	5.57	7.33	6.42	9.32	–0.85	0.19	.67
Hispanic	60.04	21.61	50.97	18.29	9.07	239.00	<.0001
White	30.56	19.66	39.02	17.30	–8.46	261.11	<.0001
Other	3.83	3.76	3.60	2.99	0.23	0.10	.76
Location (percentage of all schools)							
Colorado	58.82	49.96	55.56	51.13	3.26	0.05	.82
Kansas	20.59	41.04	22.22	42.78	–1.63	0.02	.89
Nebraska	20.59	41.04	22.22	42.78	–1.63	0.02	.89

Note: Significance tests for eligibility for free or reduced-price lunch and school size are t-tests. Significance tests for race/ethnicity and location are chi-square tests. Numbers may not sum to 100 percent because of rounding.

Source: U.S. Dept. of Education, Institute of Education Sciences, National Center for Education, Common Core of Data 2008–2009.

Teacher sample

In Years 1 and 2 of the study, site coordinators were asked to identify all teachers at their schools who taught English language development to native Spanish-speaking ELL students in grades 2–5. At schools in which ELL students did not receive English language development instruction, all literacy teachers who taught native Spanish-speaking ELL students in grades 2–5 were identified as study teachers.

At the beginning of Year 1, site coordinators identified 203 teachers meeting the above criteria for inclusion in the study who agreed to participate; of these, 67 teachers were at control schools and 136 were at intervention schools. [19] The number of teachers identified at each school varied based on school enrollment, native Spanish-speaking ELL population, and school instructional model.[20]

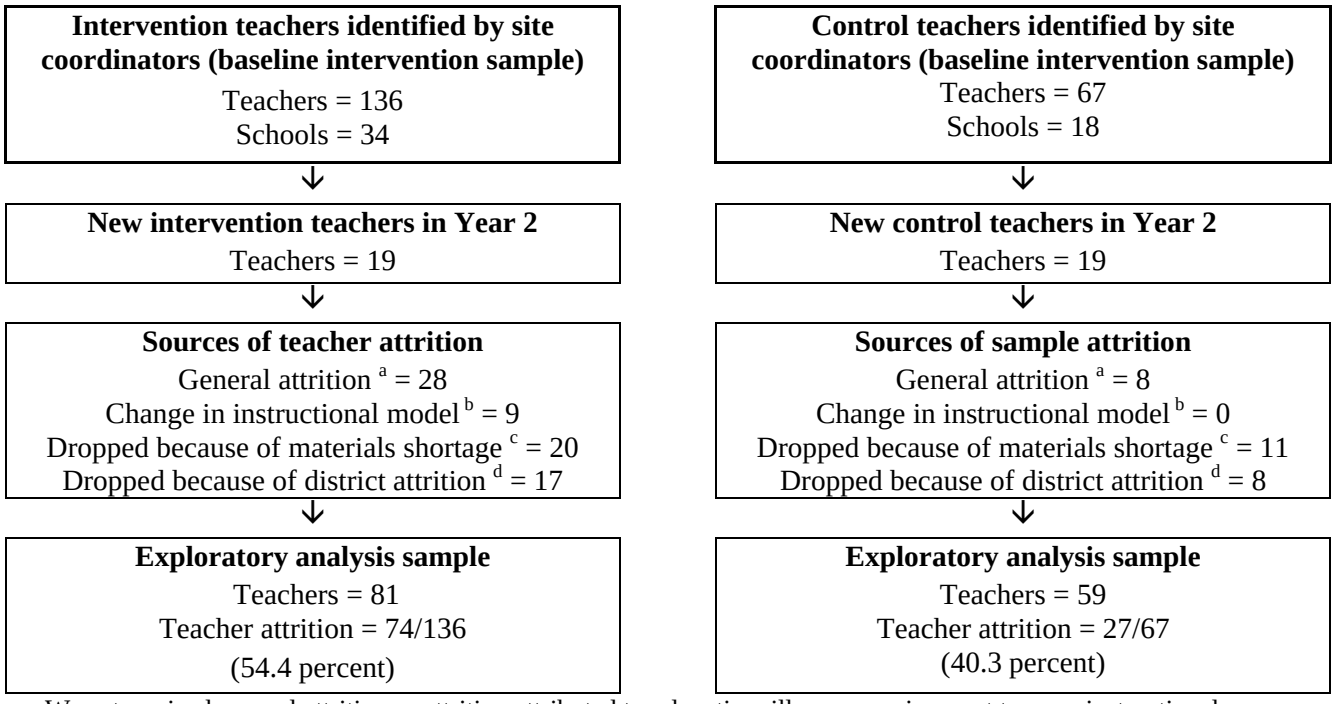
[19] Eligible teachers in a participating school could elect not to participate in the study; data were not collected on the percentage of all eligible teachers who agreed to participate in the study.

[20] For example, schools using a pull-out model might identify one study teacher whereas schools using a transitional bilingual model would report at least one teacher per grade level.

All participating teachers signed memoranda of understanding and were sent teacher background surveys by e-mail. Baseline teacher data were collected after teachers were informed of their school's random assignment status. We requested that all site coordinators clarify schoolwide instructional models, the school's ELL enrollment, and the ELL enrollment for all study teachers to ensure that they accurately identified teachers eligible for study participation.

Site coordinators were asked to use the same criteria to determine the teacher sample in Year 2 as in Year 1. Nineteen intervention teachers and 19 control teachers were identified as new study teachers in Year 2. These teachers were asked to sign memoranda of understanding and complete all data collection instruments. New intervention group teachers were also asked to attend OWE and RISE trainings provided by their site coordinators, use OWE materials for at least 30 minutes per day, and implement RISE strategies in their instruction. The final exploratory teacher sample therefore comprised 81 intervention teachers and 59 control teachers. Figure 2.1 presents the flow of teachers identified as study teachers through Year 1 and Year 2.

Figure 2.1. Teacher flow from Year 1 (training) to Year 2 (implementation) to exploratory sample



a. We categorized general attrition as attrition attributed to relocation, illness, reassignment to a noninstructional position within the school, or reassignment to a classroom which did not include any eligible study students (e.g., a first grade classroom).

b. A change in instructional model resulted in the reduction in the number of study teachers at two intervention sites. In Year 2, both schools implemented a model in which English language development (and therefore use of OWE and RISE) was provided by an English as a second language specialist in a pull-out setting rather than by classroom teachers, as had been the case in the Year 1.

c. Intervention teachers representing three sites dropped from the study because of a materials shortage. Although lack of resources was not an issue for control schools, we asked control site coordinators to identify a maximum of two study teachers per grade level in an attempt to maintain parity with intervention sites.

d. Because of a change in district-level leadership, all schools from one district dropped from the study. The teachers representing these schools were dropped from all exploratory analyses.

Both intervention and control schools experienced high levels of teacher attrition. Of the 67 control and 136 intervention teachers initially identified as study participants by site coordinators during Year 1, 74 intervention and 27 control teachers attrited prior to student IPT post-testing in spring 2010 and were therefore excluded from exploratory analyses. Teacher attrition fell into four major categories, as described below:

(1) Of those teachers initially identified for study participation, 8 control and 28 intervention teachers withdrew from the study prior to IPT post-testing because of retirement, relocation, illness, reassignment to a noninstructional position within the school, or reassignment to a classroom which did not include any eligible study students (e.g., a first grade classroom).

(2) Two intervention schools implemented an instructional model in Year 2 that differed from the one they had used in Year 1.[21]

(3) In the winter of Year 1, we learned that the publisher would provide only one set of OWE curricular materials per grade level to each school. In intervention sites where this limitation prohibited each enrolled study teacher from obtaining a full set of OWE materials, we provided additional OWE resources.[22] Twenty intervention teachers representing three sites dropped from the study because of this shortage. Although lack of resources was not an issue for control schools, we asked control site coordinators to identify a maximum of two study teachers per grade level in an attempt to maintain parity with intervention sites (which had been asked to identify a maximum of two study teachers per grade level because of a shortage of materials). No additional guidance was given to site coordinators in identifying teachers who were retained as study teachers. As a result, another 11 control teachers representing two sites withdrew.

(4) One district dropped from the study prior to the Year 2. The 8 teachers representing the control school and the 17 teachers representing the three intervention schools were excluded from subsequent analyses.

Teacher attrition in study schools was higher in intervention (54.4 percent) than in control sites (40.3 percent). Control teachers were provided the intervention after Year 2, and there was no cross-over of teachers between the intervention and control groups. Control sites were provided with a resource-balancing stipend at the beginning of Year 1, reducing the likelihood that qualified applicants would elect to apply to intervention schools to receive study-related professional development and materials.

Teachers in the 34 schools randomly assigned to the intervention group were trained in the OWE curriculum and RISE strategies during Year 1; study teachers in the 18 control group schools used strategies and curricula other than OWE and RISE with their ELL students.

Delivery of the intervention was coordinated to minimize conflict with baseline student IPT testing conducted in Year 1 of the study. All site coordinators in the intervention group were trained in RISE by the publisher on or before October 1, 2008 (see appendix F, table F1). Using a train-the-trainer model, between October 2008 and May 2009 these site coordinators trained teachers in the eight RISE modules, using timelines appropriate for their schools. OWE materials

[21] See appendix A for definitions of the instructional models employed at various study schools.

were delivered to schools between November 2008 and February 2009 (see appendix F, table F2); OWE training occurred between November 2008 and May 2009. Teachers' exposure to the RISE modules before IPT baseline testing varied, depending on the training schedules implemented at their sites. However, neither teachers nor site coordinators were exposed to OWE materials before baseline testing.

The sample of Year 1 teachers in the intervention and control groups was compared on the following baseline characteristics: gender, education level, type of teacher credential, certification to teach English as a second language, and self-reported Spanish fluency (table 2.4). Results from the comparisons indicate statistically significant differences between intervention and control groups on teachers' teaching certificate and their Spanish proficiency at baseline (see appendix C for additional analyses of teacher characteristics).

Table 2.4. Teacher characteristics at baseline, by treatment group

	Intervention (n = 136)		Control (n = 67)				
Characteristic	**Percent**	**Number**	**Percent**	**Number**	**Difference**	**Test statistic**	***p*-value**
Gender							
Male	13.24	18	10.45	7	2.79	0.32	.57
Female	86.76	118	89.55	60	–2.79		
Highest degree earned							
Bachelor	33.09	45	46.27	31	–13.18	3.13	.08
Postgraduate	54.41	74	43.28	29	11.13		
Teaching certificate							
Regular/state standard	76.47	104	65.67	44	10.80	5.51	.02*
Other	11.03	15	23.88	16	–12.85		
English as a second language certificate							
Yes	41.91	57	41.79	28	0.12	0.02	.88
No	45.59	62	47.76	32	–2.17		
Spanish proficiency							
None	5.15	7	14.93	10	–9.78	10.84	.01*
Minimal	36.03	49	47.76	32	–11.73		
Intermediate	26.47	36	14.93	10	11.54		
Advanced/fluent	19.85	27	11.94	8	7.91		

Note: Significance tests are chi-square tests; * $p < .05$. Numbers may not sum to 100 percent because of missing data.
Source: Teacher background survey and study records.

Study teachers were also compared at baseline on their previous related professional development experience. We administered a teacher background survey in fall 2008, asking respondents to describe any teaching English as a second language-related professional development they had received in the previous five years, including the name of the program, the number of hours, and the year of participation (see appendix E). Teachers were allowed to provide up to seven related professional development experiences.

We examined the teacher-reported number of professional development hours to determine whether there was a statistically significant difference between teachers in intervention and control group schools. No statistically significant difference was found (table 2.5).

Table 2.5. Hours of professional development during five years before baseline, by treatment group

Intervention group teachers (*n* = 76)		Control group teachers (*n* = 43)				
Mean	Standard deviation	Mean	Standard deviation	Difference	Test statistic	*p*-value
18.82	40.21	9.35	22.33	9.47	1.65	.10

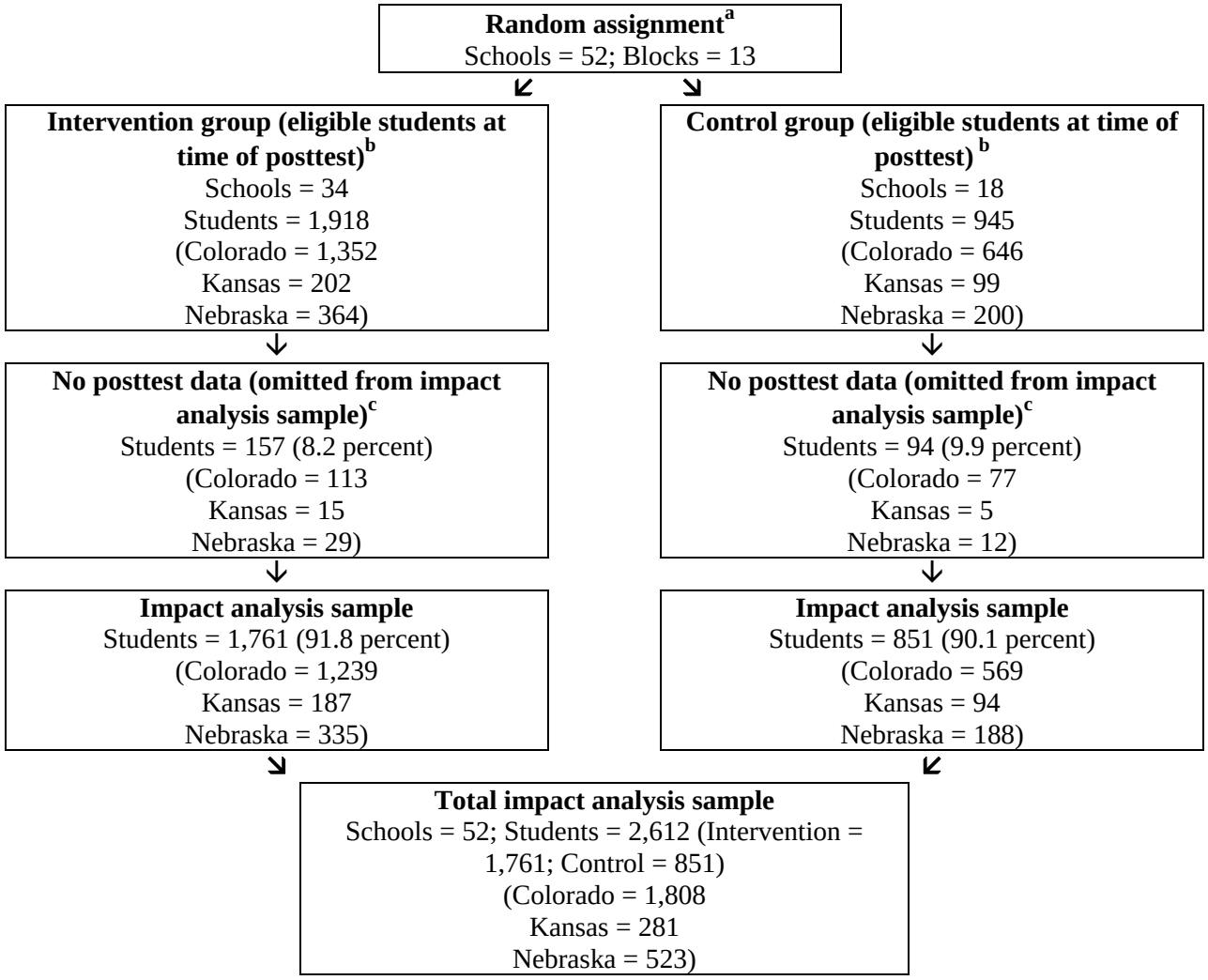
Note: p-values were derived from *t*-tests between group means.
Source: Teacher background survey, baseline sample.

Student sample

The student impact analysis sample comprised native Spanish-speaking ELL students in grades 2–5 who were placed with a study teacher in spring 2010 and whose parents did not decline to allow their child to participate (figure 2.2).[23] The parents of six students declined to allow their children to participate. The pretest was administered in spring of Year 1, when eligible study students were in grades 1–4. The impact analysis sample included 2,612 students (1,761 students in the intervention group and 851 students in the control group), and was comprised of students who took the pretest and remained in the study sample as well as students who entered the study after the pretest had been administered. No statistically significant differences between the two groups were found for gender or state; however, a statistically significant difference was found between intervention and control schools on the percent of grade 4 students (table 2.6; student-level data on eligibility for free- or reduced-price lunch, family income, parent education were not systematically available).

[23] Researchers requested that site coordinators distribute English and Spanish-language passive consent forms to the parents of children identified for participation in the study one week prior to the administration of the IPT pretest at each site. Students not present for the IPT pretest received an identical passive consent form prior to the administration of the IPT posttest. The distributed consent forms included a study overview, a discussion of methods used to ensure data confidentiality, and assurance that parents could withdraw their child from the study at any time without penalty to grade or status. To prevent data collection from students without consent, we verified that no parents of students identified for testing had declined to provide consent.

Figure 2.2. Student impact analysis sample



a. Schools were randomized in blocks. A 2:1 intervention to control ratio was used for assignment; the number of intervention schools is 1.9 times the number of control schools.
b. Includes all ELL students enrolled at a study school in spring 2010, one to three weeks before administration of the IPT posttest, as indicated by class rosters.
c. The IPT posttest was administered only at the time scheduled with the site coordinator. If students did not take the posttest at this time, data on them were not collected. In addition, no IPT pretest data (administered in fall 2008) were available for 560 students (29 percent of the intervention group sample) randomized into the intervention condition (405 in Colorado, 54 in Kansas, and 101 in Nebraska). Within the control condition, 225 students (24 percent of the control group sample) were missing pretest scores (174 in Colorado, 10 in Kansas, and 41 in Nebraska). Pretest data are missing because students were eligible to join the study until the posttest was administered in Year 2.

Table 2.6. Student characteristics impact analysis sample by treatment group

Characteristic	Intervention group (schools = 34, students = 1,761)			Control group (schools = 18, students = 851)			Difference	Test statistic	*p*-value
	Percent	Standard deviation	Number	Percent	Standard deviation	Number			
Gender									
Male	50.82	.50	895	51.94	.50	442	–1.12	0.29	.59
Female	49.18		866	48.06		409	1.12		
Grade									
2	30.04	.46	529	28.44	.45	242	1.60	0.71	.40
3	27.77	.45	489	24.21	.43	206	3.56	3.73	.05
4	22.43	.42	395	27.14	.45	231	–4.71	7.01	.01*
5	19.76	.40	348	20.21	.40	172	–0.45	0.07	.79
Location									
Colorado	70.36	.46	1,239	66.86	.47	569	3.50	3.29	.07
Kansas	10.62	.31	187	11.05	.31	94	–0.43	0.11	.74
Nebraska	19.02	.39	335	22.09	.42	188	–3.07	3.38	.07

Note: Significance tests were conducted using a one-way ANOVA; * $p < .05$. Numbers may not sum to 100 percent because of rounding. Total student population excludes one control group student and five intervention students (<0.01 percent of the study sample) whose parents refused to consent to their child's participation in the study.
Source: School records provided by site coordinators.

Table 2.7 displays the unadjusted pretest means (with missing pretest data imputed[24]) on the basic IPT composite score (standardized as *z*-scores) for the intervention and control groups, taking into account the clustered data structure (that is, a random-intercepts model with a fixed intervention effect and control variables for block randomization and student grade). The unadjusted basic IPT composite scores on the pretest did not differ by a statistically significant margin ($\gamma_{01} = –0.11$, standard error = 0.08, $p = .17$) for schools in the intervention and control groups (see appendix H, Model 1, specifications of baseline equivalence model and appendix J, table J1 for complete results).

Table 2.7. Unadjusted IDEA Proficiency Test (IPT) pretest scores, by treatment group

Outcome measure	Intervention group schools (n =34)	Control group schools (n = 18)	Standardized intervention/ control difference	*p*-value	95 percent confidence interval
Unadjusted IPT pretest (see appendix table J1)	–0.04 (–0.05) [0.88]	0.07 (0.11) [0.91]	–0.11 (0.08)	.17	–0.26 to 0.05

Note: Missing data were imputed. Numbers in parentheses are unadjusted pretest means for intervention and control groups and standard errors for estimated impact. Numbers in square brackets are standard deviations for the unadjusted means. The unadjusted means listed in table 2.7 were unadjusted for race/ethnicity and posttest covariates.
Source: IPT pretest data.

[24] Missing pretest data were imputed by using the expectation-maximization algorithm (see chapter 2).

Data collection

Data collection began in fall 2008 and continued through spring 2010 (table 2.8). The instruments used to collect data from teachers are in appendix E.

Table 2.8. Data collected for study

	Year 1		Year 2		
Group	**Fall 2008**	**Spring 2009**	**Fall 2009**	**Winter 2009/10**	**Spring 2010**
Intervention and control	IPT [a]				IPT
	Student demographics		Student demographics		Student demographics
	Student home language				Confirmation of student home language
	Teacher background	Teacher background	Teacher background		Teacher background
			Teacher log	Teacher log	Teacher log
	Teacher observation and interview (subsample) [b]				Teacher interview
Intervention only	OWE and RISE training records	OWE and RISE training records	OWE and RISE training records	OWE and RISE training records	OWE and RISE training records

Note: OWE is On Our Way to English. RISE is Responsive Instruction for Success in English. IPT is a series of language proficiency assessments for language learners in K–12.
a. We administered the IPT to all native Spanish-speaking ELL students in grades 1–4 who were present on testing day in fall 2008 and to all native Spanish-speaking ELL students in grades 2–5 present on testing day in spring 2010.
b. Teacher observations and interviews were conducted in fall 2008. We determined that 29 percent of the observed teachers were no longer participating in the study during Year 2; other teachers had transferred to a different grade level. Teacher observations and interviews were discontinued in spring 2010 and replaced with attempts to interview the full sample of participating teachers.

Teacher data

To contextualize findings, we collected data on teacher practices related to the teaching of ELL students. Four data collection instruments were used to gather contextual information: teacher background surveys, online logs, teacher observations, and interviews (see appendix E). In addition, OWE and RISE training records were used to determine completion of study requirements.

The teacher log and background survey were developed and pilot tested throughout 2006–08. The interview protocol was based on the observation protocol. Modifications were made to all

instruments as appropriate. Pilot testing with three teachers from the target population helped ensure the clarity of the instrument and estimate the time required to complete it.

The research team collected data on teacher pedagogical practices to provide context for the primary research questions. These data were collected primarily through teacher logs. Post-intervention interview data were also used to contextualize findings related to the primary research question.

During the study orientation session, teachers and site coordinators were provided with participation guidelines for the research study, including a short training on the completion of data collection instruments. Teachers were also provided with the contact information of research team members and encouraged to contact us with any questions about the completion of the data collection instruments. We aimed to minimize the burden on teachers by administering surveys and facilitating the completion of teacher logs online. Actions to improve response rates included e-mail announcements of upcoming surveys and student testing dates as well as follow-up e-mail and telephone reminders.

Although we relied on self-report data to collect teacher-level information, the content and administration of instruments did not vary between the intervention and control groups. It is therefore likely that limitations imposed by self-report data would impact the two groups to the same manner and extent.

Teacher background surveys. Teachers were asked to complete an online background survey that solicited general demographic information such as gender and race/ethnicity as well information on teaching experience, certification status, experience teaching ELL students, knowledge of other languages, and prior exposure to professional development for teaching ELL students.

Online teacher logs. The research team created an online teacher log that included checklists and rating scales derived from research findings on best practices for teaching ELL students. Teachers were asked to detail classroom practices, activities, and any accommodations or strategies used when working with ELL students.

Log items measured self-reported frequency of use of specific constructs identified in the literature as characteristic of high-quality ELL instruction (Darder 1993; Freeman and Freeman 2000; Godina 1998; Krashen 1996; National Council of Teachers of English 2006; Smith 1983). A section of the log administered only to teachers in the intervention group asked about the use of OWE and RISE. Each section of the teacher log included four to seven items. Rasch modeling techniques were employed to assess whether the items held together and to create a summary score for each section.[25]

[25] Rasch analysis is a one-parameter item response latent trait model. It produces an interval scale that estimates item difficulties and individuals' latent attitude scores on the same interval scale. In Rasch analysis, items are first arranged on the scale according to how difficult they are to endorse; each item is assigned a difficulty value. The scale units for item difficulties are logits (log odds units); 0.0 indicates an item of average difficulty. Next, person-level measures are generated for each individual, showing the degree to which a person endorses the attitude represented by the scale. This item-level conditioning of the raw scores allows for calibration into common units and enables the creation of an interval scale. We used Rasch analysis because (1) our response set for the items had

The online teacher logs were designed to be completed in approximately 25 minutes; teachers were asked to complete the logs three times during the study year.[26] We asked all participating teachers (in both the intervention and control conditions) to complete the logs.

Data obtained from the teacher logs were used to describe the ELL teaching strategies and materials used by study teachers in their classrooms throughout the year. Differences in the teaching strategies and classroom practices used by intervention and control teachers were analyzed (the results are described in chapter 5).

Teacher observations. We developed a classroom observation and interview protocol to capture baseline and end-of-study use of best practices in control and intervention sites.[27] However, data from the teacher log and post-observation interviews conducted in spring 2009 led us to conclude that classroom observations conducted at only two time points (the number of observations was limited by cost considerations) would be insufficient to capture variation in school practice, program structure, and implementation over the study period. Moreover, 29 percent of the teacher observation sample withdrew from the study between Years 1 and 2. Finally, teacher practices were not seen as central to this study; they were considered useful insofar as these data provided context for interpreting impacts on student achievement. Consequently, we discontinued the observations and instead supplemented teacher log data with data derived from telephone interviews.

Teacher interviews. The research team attempted to interview all study teachers at the conclusion of the study, in spring 2010. The response rate was 70 percent for intervention group teachers and 71 percent for control group teachers. The telephone interviews covered instructional models, delivery of English language development, perceptions of support at the district and school level, and perceptions of instructional practices in use at the school to support ELL students. Teachers in the intervention group were also asked a set of questions regarding their use of the OWE and RISE materials, including the extent to which they implemented OWE and RISE during the study, their perceptions of positive and negative elements of the two programs, their perceptions of whether their instructional practices had been significantly affected by the use of OWE and RISE, and whether they thought they would continue to use the materials in the subsequent school year. Interviews with teachers in the control group lasted an average of 5 minutes; interviews with teachers in the intervention group lasted an average of 10 minutes.

---

restricted variability—Rasch allowed for maximizing differentiation among teachers; (2) it allowed for the detection of items that did not fit with the particular constructs; and (3) it enabled us to create comparable person-level measures across the three quarterly teacher logs, with item difficulties and raw-to-scale-score conversions based on teacher responses on the first log. Rasch modeling provided us with an accurate way to interpret the likelihood of a respondent indicating that his/her students were engaged in the classroom, that he/she employed student-centered instruction, or that he/she modified instruction such that it would be responsive to student needs. The summary score measures used in this study were not rescaled from their original Rasch-produced logits; item difficulties are reported in logits (see appendix G). Internal consistency estimates were also calculated.

[26] Toward the end of each quarter, an e-mail containing a link to the study teacher log was sent to participating teachers to remind them to complete the online teacher log.

[27] Observations and follow-up interviews were initially proposed as a means of triangulating self-report data intended to help address a second research question about changes in teacher practices.

Student data

Three types of data were collected on students: demographic information, information on students' home language, and outcome data. Site coordinators provided us with student demographic information from school and district records, including information on student gender and race/ethnicity. We developed a home language confirmation form for completion by site coordinators during the spring 2010 testing timeframe to verify that students' primary home language was Spanish. Student outcomes were measured with the IPT.

Baseline testing began in October 2008 and continued through February 2009 of Year 1.[28] It was conducted for all eligible study students in grades 1–4, to provide a pretest for the student impact analysis sample (students in grades 2–5 in Year 2 of the study). All native Spanish-speaking ELL students enrolled in a classroom with a study teacher on the day of testing in spring 2010 were included in the student impact analysis sample (see figure 2.2). Student absences on testing dates resulted in missing pretest data, missing posttest data, or missing pretest and posttest data.[29] For students in the impact analysis sample, pretest IPT scores were unavailable for 31.7 percent of the intervention group and 26.4 percent of the control group. Missing pretest data were imputed by using the expectation-maximization algorithm.

Because students who enrolled in the study at any time following the pretest administration were eligible for the posttest administration, pretest data were not available for all students. Pretest scores were available for 68.2 percent of students in the intervention group (n = 1,201) and 73.6 percent of students in the control group (n = 626). Rates of pretest completion within each group demonstrated statistical significance at the .05 level.[30] A pretest completion covariate was included in the student impact analysis to account for differential rates of completion.

Use of the IDEA Proficiency Test (IPT) as the outcome measure

To determine the effects of the combined OWE and RISE interventions on the English language achievement of ELL students in listening, reading, and writing, it was imperative to choose an outcome measure that would indicate student English language proficiency and progress. A number of English language assessment instruments for ELL students were reviewed. Many were developed for use in placement decisions rather than for assessment of student proficiency.

[28] Some teachers were exposed to RISE modules before IPT baseline testing; neither teachers nor site coordinators were exposed to OWE materials before baseline testing.

[29] Researchers also considered whether students exited ELL programs between the time the student sample was identified and the posttest was administered, in spring 2010 (approximately a two-week period at each study site). Site coordinators reported no such instances. Of the 2,009 students pretested during Year 1, 74 (3.7 percent) exited ELL programs before test administration in spring 2010. A chi-square was performed to test for significance in ELL program exits between the treatment (n = 50) and control (n = 24) conditions [χ^2 (1, N = 2,009) = 0.297, p = .59]. Students who exited ELL programs prior to the posttest administration were not part of the study sample and were therefore excluded from analyses.

[30] A chi-square test of distribution was performed to test for significance in rates of pretest completion [χ^2 (1, N = 2,621 = 7.84, p = 0.0051].

We assessed student progress with the IPT. According to its publisher, the test was developed to respond to No Child Left Behind requirements for measuring English language ability.[31] IPTs are available for each of the following grade spans: kindergarten, 1–2, 3–5, 6–8, and 9–12. The test yields separate scores for listening, reading, writing, and speaking. It provides standardized scales for reporting scores across all levels. The speaking section is administered individually; all other sections are group administered. Previous versions of the test have been used by 30 states with a combined ELL enrollment of more than 1 million students (*IPT: The New Title III Testing System 2004)*. Because the test is vertically equated (scores are comparable from test to test), use of this instrument to assess school-level outcomes was considered appropriate. Appendix D provides detail on the test's reliability, validity, and scoring procedures.

To eliminate possible bias when teachers assess their own students and to improve data quality, we administered the IPT to Spanish-speaking ELL students in grades 1–4 (pretest) in fall 2008 and to Spanish-speaking ELL students in grades 2–5 (posttest) in spring 2010. The research team was trained to administer the test before both administrations. The first training was conducted by a representative of Ballard & Tighe, the test developer; the second training, considered a refresher training, was conducted internally.

Because of time and cost constraints, we assessed student oral language proficiency by administering the speaking section of the IPT to only a random sample of students stratified by school and grade. Using classroom rosters and a random number generator, we generated a 5 percent sample (rounding up) to identify students who would take the speaking subtest. At least one student from each grade was included in the sample. The oral language sample included 8.5 percent of all students in the study.

We constructed two composite scores, a basic composite score and an overall score capturing student acquisition of English as a second language.[32] The basic composite score is based on scores from the three group-administered sections of the IPT (listening, reading, and writing). The overall English as a second language proficiency score (ESL proficiency score) is based on the basic composite score and the score on the speaking section; this score was calculated for the 8.5 percent subsample of students who completed the speaking section. The basic composite score was used for all impact and exploratory analyses; the overall ESL proficiency score was used to determine the suitability of the basic composite score as a proxy for the overall ESL proficiency score. Using a composite score is preferable to using the score from each of the individual test sections, because composite scores assessing performance on both productive and receptive language provide a better, and more stable, indicator of overall English language ability (Bachman 2000). Moreover, creating a composite score reduces the number of outcomes and eliminates the need to correct for multiple comparisons (Schochet 2008).

[31] Although there is no relationship between state testing for No Child Left Behind and the IPT per se, the publisher claims that the test is aligned with No Child Left Behind requirements (http://www.nclb.ballard-tighe.com/system.html). State assessment requirements for ELL students differ. This study required a controlled data collection effort, with all students tested with the same assessment instrument, regardless of state testing requirements.

[32] Researchers considered conducting multilevel MANOVAs/MANCOVAs for the impact analyses. However, the high correlations between the various sections of the IPT (0.65–0.80) suggested that creating a composite score fit the data better.

Although the test publisher, Ballard and Tighe, has developed a conversion chart enabling the conversion of raw scores from the listening, reading, writing, and speaking sections of the IPT into a standardized overall composite score, the basic composite score used in this study includes only three sections of the test (listening, reading, and writing). We therefore created a basic composite score by averaging students' *z*-scores on the listening, reading, and writing test sections. To carry out this procedure, we first computed a standard score for each of the test sections by converting the number of raw score points to a standard score using the test publisher's conversion chart. According to Ballard & Tighe (2009), standard scores on the listening, reading, and writing sections for grades 1–2 are vertically equated with standard scores on the three parallel sections of the IPT for grades 3–5. However, the standard score scale for each of the test sections is not the same; for example, the standard score scale for the writing test section is 404–1,057, while the standard score scale for the reading test section is 462–1,066. To put the standard scores for each test section on the same scale, we converted standard scores for each of the test sections to *z*-scores, which were averaged to create the basic composite score.[33]

A similar process was used to create overall composite scores for the subsample of students to whom the speaking section of the IPT was administered. The overall ESL proficiency score is an average of students' *z*-scores on the listening, reading, writing, and speaking test sections.

Attrition and response rates

One intervention school withdrew from the study before any data were collected. Teacher- and student-level data were collected from study participants at the remaining 52 schools. We aimed to achieve high student response rates through the use of a standardized, group-administered student assessment and by working with site coordinators to determine optimal testing dates and times. We requested student enrollment lists one to three weeks before the scheduled test date at each site. The student impact analysis sample was compiled from these enrollment lists. Response rates for the student data collection efforts were similar across study groups [91.8 percent for intervention group students and 90.1 percent for control group students (table 2.9)].[34]

Table 2.9. Response rate for IDEA Proficiency Test (IPT), spring 2010 posttest

Group	Number of students enrolled in study schools	Number of students completing IPT	Percentage of students completing IPT
Intervention	1,918	1,761	91.8
Control	945	851	90.1
Total	2,863	2,612	91.2

Note: IPT is a series of language proficiency assessments for English language learners in K–12.
Source: School records provided by site coordinators; study records.

[33] Basic composite score *z*-scores were generated using the mean and standard deviation of the full study sample by test section. Overall composite score *z*-scores for the overall composite score were generated using the mean and standard deviation of students randomly selected to take the IPT speaking section, by test section.

[34] A chi-square test of distribution comparing rates of posttest completion in the control and intervention conditions was conducted, χ^2 (1, N = 2,863) = 2.00, p = .16). Results were not significant.

Teacher response rates to data collection instruments were tracked across the Year 1 baseline sample and the exploratory teacher sample (see table 2.10; figure 2.1 provides additional detail on the composition of the exploratory sample). Over the course of the study, 54.4 percent of intervention teachers and 40.3 percent of control teachers left the study.

Table 2.10. Teacher response rates to data collection instruments, by treatment group and sample

	Intervention		Control		Total	
Instrument	**Number**	**Percent**	**Number**	**Percent**	**Number**	**Percent**
Teacher background survey						
Baseline sample [a]	119/136	87.50	60/67	89.55	179/203	88.18
Exploratory sample	75/81	92.59	52/59	88.14	127/140	90.71
First teacher log						
Baseline sample	68/136	50.00	40/67	59.70	108/203	53.20
Exploratory sample	67/81	82.72	46/59	77.97	113/140	80.71
Second teacher log						
Baseline sample	66/136	48.53	44/67	65.67	110/203	54.19
Exploratory sample	66/81	81.48	49/59	83.05	115/140	82.14
Third teacher log						
Baseline sample	67/136	49.26	45/67	67.16	112/203	55.17
Exploratory sample	68/81	83.95	50/59	84.75	118/140	84.29
At least one teacher log						
Baseline sample	80/136	58.82	46/67	68.66	126/203	62.07
Exploratory sample	78/81	96.30	53/59	89.83	131/140	93.57
Teacher interview						
Baseline sample	na	na	na	na	na	na
Exploratory sample	57/81	70.37	42/59	71.19	99/140	70.71

a. Includes teachers who did not continue on to Year 2 of the study. Baseline sample does not include teachers who began the study during the implementation year (Year 2).
Source: Teacher background surveys; online teacher logs.

Data analysis

The primary purpose of the impact analysis was to provide an unbiased estimate of the impact of the combination of OWE and RISE on student English language acquisition. Consistent with the random assignment of schools to either the intervention or the control group, impact was estimated at the school level, using multilevel modeling to account for the sources of variability inherent in the nested structure of the data. Variance components at the student and school level were estimated to confirm the assumption of the nested structure of the data.[35]

A two-level model with students nested within schools was used to estimate the impact of the combination of OWE and RISE on students' English language acquisition. Because of the nature of the ELL program models instituted across study schools, students could be linked to schools and districts but not individual teachers.

At the student level (level 1), the model includes the prior English language ability of the student (the IPT pretest score) and dummy-coded grade level variables. All variables in the level 1 model were grand-mean centered, with grade 2 the omitted grade level. Level 1 of the model was specified as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(IPT\ Pretest)_{ij} + \beta_{2j}(Grade\ 3)_{ij} + \beta_{3j}(Grade\ 4)_{ij} + \beta_{4j}(Grade\ 5)_{ij} + r_{ij}$$

where

- Y_{ij} is the IPT posttest score for student *i* in school *j*.
- β_{0j} is the regression-adjusted mean IPT posttest score for students in school *j*.
- β_{1j} is the effect of previous English language ability on current English language ability.
- $\beta_{2j...}\ \beta_{4j,}$ are the fixed level 1 covariate effects for grades 3–5.
- r_{ij} is the random error for student *i* in school *j*.

At the school level (level 2), the model includes a dummy variable to indicate group assignment (intervention = 1, control = 0), as well as the district blocking variables used in randomization. The intervention indicator variable remained in the original dummy metric; the district blocking variables were grand-mean centered, with Block 1 the omitted district blocking variable.

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(Intervention)_j + \gamma_{02}(\%\ White)_j + \gamma_{03}(\%\ Hispanic)_j + \gamma_{04}(Block2)_j + \gamma_{05}(Block3)_j + \gamma_{06}(Block4)_j + \gamma_{07}(Block5)_j + \gamma_{08}(Block6)_j + \gamma_{09}(Block7)_j + \gamma_{10}(Block8)_j + \gamma_{011}(Block9)_j + \gamma_{012}(Block10)_j + \gamma_{013}(Block11)_j + \gamma_{014}(Block12)_j + \gamma_{015}(Block13)_j + u_{0j}$$

[35] Preintervention equivalence was estimated using a multilevel model in order to account for the clustered nature of the data. For the multilevel model used for the analyses, see appendix H. For the multilevel model used for the variance components, see appendix I. For the results of the preintervention equivalence analysis, see chapter 4.

$$\beta_{1j} = \gamma_{10} \ldots \beta_{4j} = \gamma_{40}$$

where

- γ_{00} is the regression-adjusted mean IPT basic composite score for control schools.
- γ_{01} is the difference between intervention and control schools in the regression-adjusted mean IPT basic composite score (intervention effect).
- γ_{02} is the effect of the percentage of White students in the school on the adjusted mean IPT basic composite score.
- γ_{03} is the effect of the percentage of Hispanic students in the school on the adjusted mean IPT basic composite score.
- γ_{04}–γ_{01} are the additive effects of each district block used in the random assignment of schools, with Block 1 the omitted block.
- γ_{10}–γ_{40} are the average regression slopes for IPT pretest and grade fixed across schools.
- u_{0j} is the random error for school *j*.

In both the level 1 and level 2 models, all variables except the treatment dummy were grand-mean centered to enhance interpretability of the intercept (Enders and Tofighi 2007; Raudenbush and Bryk 2002). As modeled above, grade dummies are fixed. Only the school-adjusted average on the outcome (that is, the school intercept) was allowed to vary randomly across schools; the regression slopes for the student grade-level IPT pretest were fixed across schools, making this a random intercept model (Raudenbush and Bryk 2002). Given these specifications, the primary parameter of interest in this study was γ_{01}, the estimated impact of the combination of OWE and RISE on whole school English language acquisition (that is, the adjusted school mean difference between intervention and control groups).

Methods used to address missing data

We used the multiple imputation expectation-maximization algorithm to handle missing pretest data, as recommended by Puma et al. (2009). To impute missing pretest data, we developed a model that incorporated posttest data and the student-level covariates included in the impact model. Missing posttest data were not imputed, as the student sample was defined as all native Spanish-speaking ELL students receiving instruction from a study teacher on the day of testing in spring 2010 (Year 2). Imputation of missing data using the expectation-maximization algorithm was conducted using the PROC MI feature in SAS, which was used to create a set of 10 complete data sets.[36] Impact analyses were then conducted on each of the 10 data sets using the PROC MIXED procedure. The 10 sets of findings were then combined using the PROC MIANALYZE procedure to produce an overall impact estimate with a correct standard error. As

[36] Ten imputed data sets were constructed, because Schafer (1999) suggests that no more than 10 are usually required. Other experts in missing data have suggested that as few as five are needed (Graham 2009).

it is not possible to specify the multilevel data structure in the implementation of the expectation-maximization algorithm in SAS PROC MI (Puma et al., 2009), we used fixed-effects dummy variables for schools in place of the random intercept terms included in the impact model, as recommended by Puma and his colleagues. Following standard practice for randomized controlled trials, missing data values were imputed separately for the intervention and control groups. (Imputations were not conducted for the school that dropped out of the study, as no teacher or student data were collected.) The combined data sets were used in the impact analyses.

Intent-to-treat estimates

Impacts were estimated using intent-to-treat analyses from the 52 study schools. Outcome data were collected at the student level and transformed into a composite score for the IPT using the previously described procedures.

Consistent with the random assignment of schools to treatment groups, confirmatory impact analyses were estimated at the school level using multilevel modeling to account for the sources of variability in the data that result from the nested structure of the school environment. Two-tailed *t*-tests ($p < .05$) were used to assess the significance of the impact estimates. A two-level hierarchical model was estimated to address the primary research question regarding student achievement. Estimation of an unconditional model was used to determine the grand mean for student English acquisition and to partition the variance of the outcomes around the grand mean into its components (within and between schools). The planned confirmatory analyses used a single composite score for the outcome. Thus, no corrections for multiple comparisons were required.

Effect sizes

Effect sizes are provided in order to help indicate the magnitude of the impact of the intervention on student and teacher outcomes. Glass's Δ (Glass, McGaw, and Smith 1981) was calculated to present the difference between the intervention and control groups in standard deviation units of the control group. It was chosen because the intervention may change the standard deviation of the intervention group such that the pooled standard deviation (as used in Hedges' *g*) may not represent the population on which the results would be most interesting, namely, schools that are considering using the intervention training and materials to increase student achievement. Ninety-five percent confidence intervals for the difference in group means are also provided. Confidence intervals, although not a typical effect size measure, provide information complementary to standard effect size measures (Grissom and Kim 2005).

3: IMPLEMENTATION OF THE INTERVENTIONS

This chapter describes the OWE and RISE interventions (see appendix A for detailed descriptions of OWE and RISE) and their implementation. It discusses training and implementation as intended by the developer and as defined within the study design. It also examines the fidelity with which the interventions were implemented.

Both the OWE and RISE programs were designed to allow teachers flexibility in selecting and implementing components. Over the course of the study, we collected data on training in and implementation of OWE and RISE from teachers (through training logs, quarterly teacher logs, surveys, and telephone interviews) and from site coordinators (through surveys and informal communications). Teacher-reported implementation data were obtained through online logs[37] and telephone interviews, with teachers self-reporting moderate implementation levels. We found variation in the instructional models used both across and within schools (see appendix A for detailed descriptions). However, neither the number of minutes of English language development instruction nor reported professional development plans differed between intervention and control schools.

On Our Way to English

Program description

The OWE curriculum (Harcourt Achieve 2004a) uses thematic units to integrate English oral language development, literacy learning, and standards-based content area instruction. The program is designed to be used in a variety of ELL instructional settings, including bilingual and dual-language classrooms, mainstream classrooms, self-contained ELL classrooms, and pull-out or push-in programs delivered by an English-as-a-second language specialist. Regardless of the setting, the developer recommends that instructors organize daily classroom instruction around OWE thematic units, which are intended for use with all children in the classroom. Each thematic unit is designed to be completed within four weeks. Assessment opportunities are embedded throughout each unit, as well as at the end of each unit. Teachers are encouraged to tailor the units to meet the needs of their students.

Teachers' guides suggest classroom activities and processes for each day over the four weeks. For example, on day one of week one in grade 1, teachers are encouraged to use a manipulative chart to "set the scene." Teachers are instructed on how to make each component understandable to students, by pointing to items and gesturing while speaking about the scene-setting activities and materials. Teachers are also asked to use an audio CD with spoken songs related to the scene-setting activities and materials. A script is provided for each day of the week, with instructions and responses (possible phrases, sentences) as well as additional ideas for deepening a lesson. Options for different stages of language development are also provided. For example, students in the early language development stages of grade 1 might be asked to provide the names of the items in their backpacks and act out how to use them; students at later language

[37] Collected quarterly; respondents were explicitly instructed to respond based on a "typical week" of teaching.

development stages might be asked to use phrases or complete sentences to describe the use of the items in their backpacks.

Teachers are encouraged to set up "centers" (classroom areas in which small groups of students work on learning tasks without direct teacher support), where students can review concepts from OWE activities previously introduced to the whole class. Such centers augment unit activities and allow teachers time to meet with small groups and individual students.

OWE also includes word study and phonics instruction, in which teachers incorporate props to develop phonological awareness through listening and speaking activities, with the goal of helping students recognize sounds and explore sound-letter relationships. In addition, OWE includes guided reading (an instructional strategy in which a teacher works with a small group of students on target skills) and comprehension instruction based on an approach that blends listening, reading, writing, and speaking. OWE materials include 40 guided reading book sets for use in instruction and four short benchmark books for use in assessing literacy skills, all of which are aligned with a variety of student proficiency levels to ensure that students with a wide range of literacy skills can be accommodated in the mainstream classroom.

Program timing and delivery. The pacing chart included in OWE kits suggests incorporating comprehensive thematic units for at least 30 minutes a day. Developer-recommended OWE delivery varies according to the ELL instructional setting within a school. When programs for ELL students are delivered by an English as a second language specialist (for example, a pull-out or push-in instructional model), the developer recommends that the specialist use thematic units and phonics during ELL-specific instructional time and that regular classroom teachers assign OWE guided reading titles to ELL students during small-group time. In instructional settings in which ELL-specific instruction is incorporated into a mainstream or dual-language classroom setting, the developers suggest that OWE thematic units form the basis of classroom instruction and that all subcomponents (word study, guided comprehension, and so forth) are delivered by the classroom instructor. (See appendix A for a detailed description of OWE).

Program training and implementation

Schools assigned to the intervention group received OWE materials and training over the course of Year 1 (see appendix F). Each intervention school teacher attended a half-day OWE training session, facilitated by a trained Harcourt representative.[38] Training sessions were designed to introduce study teachers to the OWE curriculum and materials. The Harcourt trainer reviewed materials included in a standard OWE package (thematic units, leveled guided readers, song charts, word wall components, and assessment materials); guided teachers through a sample OWE lesson plan; discussed incorporation of OWE materials in a typical lesson; and offered guidance on how teachers might incorporate OWE thematic units and materials into their existing curricula. For instance, instructors working with ELL students might choose to enhance a standard district science curriculum using complementary exercises from an OWE thematic unit. Per the standard approach of the publisher, districts or schools adopting the OWE curriculum were allowed to contact Harcourt for curricular guidance. Intervention teachers were trained in OWE implementation but had latitude to implement the program as they saw fit.

[38] New teachers were trained by site coordinators.

Responsive Instruction for Success in English

Program description

The RISE professional development program (Amador-Watson 2007) was designed for use by teachers in a variety of instructional settings. It aims to familiarize teachers with research-based practices for use in instruction with ELL students through participation in a series of eight modules, each designed to build teacher skills and knowledge in a specific area. For example the goal of the first module is to extend the range of approaches used to teach and assess ELL students; the goal of the second module is for teachers to demonstrate understanding of assessment and how it guides instruction. Teacher materials include resource notebooks for each of eight RISE modules, a data collection booklet that includes materials teachers can use to assess the progress of a case-study student (an application activity using RISE approaches with an ELL student), worksheets teachers use to track their personal goals, and forms for lesson planning. (See appendix A for a detailed description of RISE.)

Program training and implementation

RISE training is designed to be delivered by Rigby consultants[39] or teachers who have been trained by a Rigby consultant using a train-the-trainer model. In both the consultant-led and trained-trainer models, facilitators deliver eight RISE instructional modules to teachers in the form of interactive workshops (see appendix A, table A1). During their training, teachers discuss teaching practices with their peers, view videotaped lessons, and read articles about curriculum. Workshop modules also include a CD-ROM that illustrates sample classroom vignettes modeling crucial RISE instructional concepts and provides interactive activities. Additional classroom support is provided by a facilitator who performs instructional demonstrations and coaching (Knox and Amador-Watson 2002).

Because of cost considerations, this study used the RISE train-the-trainer rather than the consultant model. During the facilitator training, a Rigby professional development consultant trained site coordinators to deliver RISE training to teachers at their schools.[40] All intervention group site coordinators attended a three-day train-the-trainer session facilitated by a Harcourt representative. Four training sessions were held during fall 2008 (see appendix F, table F1). During each session, the facilitator discussed the application of RISE in a real-world classroom scenario, identified websites relevant to teaching English language learners, and explored the interactive RISE CD-ROM/DVD (Amador-Watson 2007) provided to each participant. In addition to the CD-ROM/DVD, participants received facilitator and teacher resource manuals for each RISE module, transparencies, and copies of the Rigby ELL chart for assessing ELL literacy and language development. Following each train-the-trainer session, site coordinators evaluated RISE training content and instruction; 63.5 percent of site coordinators indicated that they felt "very prepared" or "prepared" to instruct teachers across each of the eight RISE modules.

[39] A subsidiary of Houghton Mifflin Harcourt.

[40] Researchers asked each school to identify a site coordinator during site recruitment. Following consultation with the publisher, researchers created a list of characteristics necessary to fulfill the role of site coordinator/trainer for the RISE program. Individuals considered for this role were asked to provide information that allowed researchers to ensure that they possessed characteristics that enabled them to perform this role.

Site coordinators served as the RISE trainers for intervention schools. Each teacher in the intervention group was provided with a RISE professional development kit. Although we requested that site coordinators complete each of the eight RISE training modules before the end of the 2008/09 school year, site delivery and timing of module instruction varied. Some sites completed multiple RISE modules in one day; other sites paced RISE instruction to coincide with regular departmental meetings or in-service days. Some site coordinators assigned homework or conducted classroom observations to reinforce training. Allowing flexibility in delivery of the professional development was necessary to accommodate the varied school schedules for teacher professional development. Such differences in implementation were expected as a part of a study intended to assess impacts in real-world implementation settings.

Fidelity of implementation

Given the flexibility to implement OWE and RISE as determined appropriate, implementation varied across and within implementation group schools. For example, some teachers chose to organize all classroom instruction around OWE thematic units (as recommended by the developer); others chose to incorporate specific lessons or components of OWE instruction (for example, a particular daily lesson or a regular guided comprehension section during small-group ELL reading time). OWE implementation also varied by instructional setting. In some pull-out/push-in settings, both ELL specialists and mainstream classroom teachers delivered OWE instruction to ELL students at various times during the week. In other settings, OWE instruction was delivered only by English as a second language specialists. Classroom teachers in immersion settings reported incorporating OWE into mainstream classroom instruction as well as targeting only ELL students.

As an early implementation measure, we assessed the number of RISE modules intervention site coordinators delivered to participating study teachers during Year 1 (table 3.1; see also appendix F, table F4). We also tracked whether teachers entering the study during Year 2 received training in RISE.

Table 3.1. Number of teachers falling into low to medium and high categories of Responsive Instruction for Success in English (RISE) modules completed

Teacher Sample	Number of teachers completing fewer than 7 modules	Number of teachers completing 7–8 modules	Number of teachers not reporting
Year 1; n = 136	25	90	21
Year 2; *n* = 81	9	72	0

Source: Study records.

In addition to monitoring the reported number of RISE modules completed, we assessed fidelity of implementation during the 2009/10 school year using items from the online teacher logs. Intervention group teachers were asked to respond to questions regarding whether they used OWE with ELL students for at least 30 minutes a day, in keeping with the developer's recommendations. They were also asked whether they incorporated general OWE and RISE

teaching strategies in their ELL instruction. These log items were included in the composite fidelity measure (described below).

Table 3.2 shows the teacher log questions, the categorical response options, and the number of points allocated to each response. For the first three questions, teachers were awarded a specific number of points for each response. For the last set of questions, the average score (rounded to the next highest integer) across the eight individual OWE and RISE strategies included on the online teacher logs was used to award points (see appendix F, table F5, for specific items and average frequencies for all items used in calculating implementation fidelity). The total number of possible points was 12. Based on their responses on each of the three online teacher logs, intervention teachers earned a fidelity score of 1–12 on each log. An overall fidelity of implementation score was created by averaging teachers' fidelity scores across the three logs; average scores were rounded to the nearest integer. Missing log data were not figured into the average implementation fidelity score: if data were missing the implementation fidelity score was calculated as the average of data from the existing log or logs.

Table 3.2. Log questions used to determine teacher fidelity of implementation of On Our Way to English (OWE) and Responsive Instruction for Success in English (RISE)

Question	Points
1. Did you implement OWE this quarter?	
Yes	1
No	0
2. How many days do you use OWE with your ELL students in an average week?	
0	0
1	1
2	2
3	3
4	4
5	5
3. On average how many minutes a day do you use OWE with ELL students at each grade level?	
0	0
1–29 minutes	1
30+ minutes	2
4. In a typical week, how often do you incorporate the following OWE and RISE strategies during ELL instruction? (average score over the 8 listed strategies)	
Never	1
Rarely	2
Sometimes	3
Always	4
Total points possible	12

Note: OWE is On Our Way to English. RISE is Responsive Instruction for Success in English.
Source: Online teacher logs.

Based on these average fidelity scores, teachers were designated as low, medium, or high implementers. Cut-off points for determining fidelity categories were created before the analysis of the data. We established cut scores based on standards established for intervention fidelity outlined in the study analysis plan and communicated to study participants through memoranda of understanding and other study guidance. All teachers self-reporting that they used OWE for at least 30 minutes every school day with ELL students received a fidelity of implementation score of 9 or higher (1 point awarded for item 1; 5 points awarded for item 2; 2 points awarded for item 3; and 1+ points awarded for item 4). Respondents reporting that they never used OWE could receive a maximum score of 2 if they regularly incorporated RISE strategies (2 points on item 4). These standards were used to establish high (9–12), medium (5–8), and low (1–4) scores for implementation fidelity.

Table 3.3 shows the number and percentage of teachers categorized in each of the three fidelity categories. It indicates that 16 percent of all intervention teachers self-reported low levels of implementation, 51 percent self-reported moderate levels of implementation, and 30 percent self-reported high levels of implementation.

Table 3.3. Fidelity with which On Our Way to English (OWE) was implemented by intervention teachers

Category	Number of points	Number of teachers	Percent of teachers
Low	1–4	13	16.05
Moderate	5–8	41	50.62
High	9–12	24	29.63

Note: Three intervention teachers did not provide data to enable the calculation of an implementation score; n = 78. The total intervention sample size (n = 81) was used as denominator; therefore percentages do not sum to 100.
Source: Online teacher logs.

Findings from teacher interviews

At the end of Year 2, we conducted telephone interviews with teachers in the intervention group to help us ascertain teachers' opinions of the OWE and RISE programs, their perceptions of whether using OWE and RISE had affected their ELL instruction, and whether they planned to use OWE and RISE the following year. Ten-minute telephone interviews were conducted with 57 intervention school teachers, a response rate of 70 percent[41]

We first examined the average weekly number of minutes of English language development/literacy instruction reported by teachers (table 3.4). There was no statistically significant difference between teachers in intervention schools and teachers in control schools.

[41] We also conducted brief telephone interviews with 42 control group teachers, a response rate of 71%. These data allowed for the comparison of hours spent on English language development / literacy instruction (table 3.4).

Table 3.4. Weekly teacher-reported hours spent on English language development/literacy instruction, by treatment group for exploratory sample

Intervention group (n = 53)		Control group (n = 34)				
Mean	Standard deviation	Mean	Standard deviation	Difference	Test statistic	p-value
6.98	7.34	8.43	9.68	–1.45	.75	.46

Note: The *p*-value was derived from *t*-tests of the means of each group.

Source: Teacher interviews.

Responsive Instruction for Success in English (RISE) professional development

Teachers were first asked about the extent to which they implemented the RISE professional development strategies in their English language development/literacy instruction. Of the 57 teachers who responded to this question, 15 (26 percent) reported implementing RISE "to a great extent," 26 (46 percent) reported implementing RISE "somewhat," and 16 (28 percent) reported implementing RISE "a little" or "not at all." Of respondents who reported implementing RISE "a little," three indicated that they did not feel they had received sufficient training and three stated that they did not have sufficient time to implement RISE. Other explanations provided by respondents included that they had already been using RISE strategies in their instruction and that the school focus was not on implementing RISE.

Teachers were also asked if there was anything in particular that they liked or disliked about the RISE professional development program. Fifty teachers (88 percent) responded to this question. Among teachers who responded, 17 (34 percent) indicated that they liked the RISE materials, the training, or both; 6 (12 percent) criticized the RISE professional development program. One area of dissatisfaction was the perceived misalignment between the professional development training and actual conditions at the school sites. Other areas of dissatisfaction included the amount of time devoted to RISE training and the redundancy of the training with strategies with which teachers were already familiar. Of the remaining respondents, 5 (10 percent) provided neutral feedback, 5 (10 percent) reported that they never received RISE training, and 8 (16 percent) were unable to distinguish RISE from OWE, as their responses referred to elements of the OWE program.

On Our Way to English (OWE) curriculum and materials

Teachers were next asked about the extent to which they implemented the OWE curriculum. Of the 57 teachers who responded to this question, 29 (51 percent) reported implementing OWE "to a great extent," 22 (39 percent) reported implementing OWE "somewhat," and 6 (11 percent) reported implementing OWE "a little." Teachers who reported implementing OWE "a little" reported that the materials were not challenging enough for students, that they did not have sufficient time to implement the curriculum, or that the materials were not convenient.

Teachers were asked if there was anything in particular that they liked or disliked about the OWE curriculum and materials. Fifty-two teachers (91 percent) responded to this question.[42] Forty-five teachers (87 percent) provided positive comments on the OWE curriculum or materials, expressing their satisfaction with the OWE leveled guided readers, the chant books, the big books, and the thematic units; 20 respondents (38 percent) provided negative comments on the OWE curriculum and materials. The most common concern was that the OWE materials were too easy or not appropriate for the grade level to which they were assigned. Nine respondents (17 percent) indicated that they had to supplement the OWE curriculum and materials in order to flesh out the units or fill in gaps in the curriculum.

Effect of the interventions on instruction

The majority of respondents reported being influenced by at least one of the interventions, with 43 of 57 respondents (75 percent) indicating that OWE, RISE, or both had had an effect on the way they taught English language development/literacy. Among those who indicated the intervention had influenced their teaching practice, twenty-seven teachers (63 percent) indicated that OWE had affected their English language development/literacy instruction. Nine teachers (21 percent) indicated that OWE and/or RISE had affected their teaching practice, but did not specify which intervention had the effect.

Continued use of the interventions

Forty-four respondents (77 percent) reported that they planned to use at least one of the interventions the following school year. Of the 13 respondents (23 percent) who reported that they did not plan to use either intervention the following year, seven (54 percent) reported that they were retiring or would not be working in the same position the following school year.

Findings from first teacher log

In October of Year 2, we deployed the first teacher log. Data from this log were used to describe differences between the intervention and control groups during the implementation year.

Differences between control and intervention groups

Teachers and site coordinators in both the intervention and control groups reported variation in ELL instructional models used within districts, schools, and, in some instances, grades within schools (table 3.5). In October of the implementation year, there were no significant teacher-reported differences in instructional models used in intervention and control classrooms.

[42] Teacher responses to this question often spanned more than one category. Therefore, percentages do not add up to 100 percent.

Table 3.5. Teacher-reported ELL instructional models in exploratory sample in October of Year 2, by treatment group

	Intervention (n = 67)		Control (n = 46)				
Instructional model	**Percent**	**Number**	**Percent**	**Number**	**Difference**	**Test statistic**	***p*-value**
Pull-out/small group (ELL students only)	50.75	34	45.65	21	5.10	0.28	.59
Modified pull-out/small group (can include non–ELL students)	23.88	16	41.30	19	–17.42	3.87	.05
Dual language (two-way, developmental bilingual)	7.46	5	a	a	a	a	a
Maintenance (late-exit bilingual)	10.45	7	a	a	a	a	a
Transitional (early-exit bilingual)	20.90	14	17.39	8	3.51	0.21	.64
Push-in	a	a	a	a	a	a	a
Newcomer	8.96	6	21.74	10	–12.78	3.67	.06
Other	5.97	4	a	a	a	a	a

Note: Because some sites reported midyear changes in instructional models, we limited this comparison to teacher responses obtained in October of Year 2. Teachers were given the option of reporting more than one instructional model.

a. Data in cells with fewer than three entries are not displayed in order to protect anonymity.

Source: Online teacher logs.

To understand the context of study classrooms, the research team collected information about instructional practices used with ELL students in treatment and control classrooms. These practices were identified in a search of the literature and are intended to describe the breadth of practices employed in study classrooms during the intervention. No data were collected on teacher pre-intervention use of instructional strategies. Because use of particular instructional strategies at particular times during the school year might have varied, we pooled all logs submitted over the course of the study and coded teacher-reported usage using a dichotomous "yes/no" criterion for each possible instructional strategy (table 3.6). Of the 30 tests of statistical significance conducted to examine differences between teachers in the intervention and control groups in the instructional strategies presented in table 3.6, one yielded a statistically significant difference (at $p < .05$): teachers in the intervention group were more likely than teachers in the control group to use graphic organizers.

Table 3.6. Teacher-reported instructional strategies in exploratory analysis sample, by treatment group

	Intervention group (n = 81)		Control group (n = 59)				
Instructional strategy	**Number**	**Percent**	**Number**	**Percent**	**Difference**	**Test statistic**	***p*-value**
Small-group instruction (mixed ability groups)	65	80.25	50	84.75	–4.50	0.47	.49
Small-group instruction (same ability groups)	70	86.42	50	84.75	1.67	0.08	.78
Whole group lecture	64	79.01	45	76.27	2.74	0.15	.70
Whole group demonstration	68	83.95	50	84.75	–0.08	0.02	.90
Whole group direct instruction (for example, introduction, instruction, guided practice, independent practice)	67	82.72	51	86.44	–3.72	0.04	.84
Dyads (paired instruction)	44	54.32	35	59.32	–5.00	0.35	.56
Peer instruction	53	65.43	44	74.58	–9.15	1.34	.25
Computer-based instruction	0	0.00	0	0.00	0.00	na	na
Other multimedia-based instruction	39	48.15	23	38.98	9.17	1.16	.28
One-on-one instruction	69	85.19	49	83.05	2.14	0.12	.73
Sentence strips	38	46.91	32	54.24	–7.33	0.73	.39
Games	63	77.78	47	79.66	–1.88	0.07	.79
Realia/manipulatives	66	81.48	46	77.97	3.51	0.26	.61
Graphic organizers	75	92.59	48	81.36	11.23	4.04	.04*
Guided/leveled readers	69	85.19	50	84.75	0.44	0.01	.94
Shared readers/big books	60	74.07	37	62.71	11.36	2.07	.15
Songs/chants	56	69.14	37	62.71	6.43	0.63	.43
Cooperative learning	64	79.01	46	77.97	1.04	0.02	.88
Drama/guided role play/simulations	33	40.74	33	55.93	–15.19	3.16	.08
Student presentations/report back sessions	46	56.79	39	66.10	–9.31	1.24	.27
Total Physical Response[a]	56	69.14	32	54.24	14.90	3.25	.07
Discovery learning/inquiry-based learning	35	43.21	22	37.29	5.92	0.50	.48
Journals	58	71.60	44	74.58	–2.98	0.15	.70
K-W-L (what we know, what we want to know, and what we learned) graphic	58	71.60	35	59.32	12.28	2.31	.13

Instructional strategy	Intervention group (n = 81) Number	Intervention group (n = 81) Percent	Control group (n = 59) Number	Control group (n = 59) Percent	Difference	Test statistic	p-value
organizer							
Seat work/worksheets	59	72.84	50	84.75	–11.91	2.81	.09
Word/vocabulary walls	64	79.01	45	76.27	2.74	0.15	.70
Flash/vocabulary cards	61	75.31	46	77.97	–2.66	0.13	.71
Native language support	42	51.85	27	45.76	6.09	0.51	.48
Listening centers	38	46.91	34	57.63	–10.72	1.57	.21
Silent reading centers	53	65.43	39	66.10	–0.67	0.01	.93

* $p < .05$.

Note: The *p*-values were derived from *chi-square* tests among frequencies of yes/no responses between treatment and control groups for each of the strategies.

a. Total Physical Response is an instructional method in which students are expected to respond to teacher spoken language with physical movements.

Source: Online teacher logs.

4: IMPACT ON STUDENT ENGLISH LANGUAGE ACQUISITION

This chapter presents the results of the analyses that estimated the impact of the combination of OWE and RISE on students' English language acquisition, as measured by the IPT composite score (based on the scores on the listening, reading, and writing sections). This chapter discusses the impact analyses, the impact estimates generated by the multilevel models, and the sensitivity analyses.

To summarize the findings presented in this chapter, no statistically significant difference was found between the basic IPT composite scores of students who were exposed to OWE and RISE and students in the control group, who were not. Results of the sensitivity analyses reveal that the statistical significance of the impact estimate was invariant to the use of covariates in the analytic model, the method used to treat missing data, and the use of a composite score.

Impact analyses

This study's primary outcome measure was students' basic composite score on the IPT. Because of the nature of the ELL program models instituted across study schools, students were linked to schools and districts but not to individual teachers. Thus, the student impact models nested students within schools and schools within districts (blocks). The analytic sample for estimating the impacts of OWE and RISE included 2,612 students nested within 52 schools. The average number of students nested within a school was 50.23.

As described in chapter 2, two composite scores were created using student test scores on the IPT. For students in the impact sample, the correlation between the two scores was 0.95. This strong, positive correlation indicates that the basic composite score serves as a good proxy for the overall ESL proficiency score, providing justification for the decision to use the basic composite score in all further analyses.

Before estimating the conditional multilevel models (model including at least one covariate), we analyzed an unconditional multilevel model without covariates to assess clustering at the student and school levels. The estimated intraclass correlation between any two students attending the same schools was 0.12 (see appendix I, table I1). The presence of clustering supported the use of multilevel modeling to assess the impact of OWE and RISE on students' English language acquisition.

Results from the adjusted posttest model (that is, the random intercept model including pretest covariates and controls for the percentage of White students and the percentage of Hispanic students, in addition to controls for block randomization and student grade) are presented in table 4.1. The combination of OWE and RISE did not have a statistically significant impact on students' basic IPT composite scores ($\gamma_{01} = -0.03$, standard error = 0.06, $p = .66$) (see appendix H, Model 2 for the model specifications for the confirmatory impact analysis and appendix J, table J2 for complete results).

Another way to interpret the adjusted average posttest difference between intervention and control schools is to standardize the difference as an effect size. Glass's Δ (Glass et al. 1981) was used to estimate the effect of OWE and RISE on students' English language acquisition. This calculation results in a measure of the difference between the two group means expressed in units of the control group standard deviation. The effect size on the adjusted posttest was –0.03.

Table 4.1. Results from the confirmatory impact analysis model for basic composite score

Item	Intervention group mean	Control group mean	Estimated impact	*p*-value	95 percent confidence interval	Effect size[a]
Adjusted IPT posttest score	0.00 (–0.04) [0.84]	0.03 (0.10) [0.86]	–0.03 (0.06)	.66	–0.14 to 0.09	–0.03
Sample size (schools)	34	18				
Sample size (students)	1761	851				

Note: IPT is a series of language proficiency assessments for English language learners in K–12. Numbers in parentheses are unadjusted posttest means for intervention and control groups and standard errors for estimated impact. Numbers in square brackets are standard deviations for the unadjusted means.

a. Effect size is calculated as Glass's Δ (Glass, McGaw, and Smith 1981), in order to present the difference between intervention and control groups in standard deviation units of the control group.

Source: IPT data.

Sensitivity analyses

The research team conducted eight sensitivity analyses to demonstrate the robustness of the impact results for the confirmatory research question (see table 4.1 for the impact results from the confirmatory impact analysis model). None of the impact estimates for these analyses was significant. Models estimated in the sensitivity analyses are shown in appendix H as Models 3–10 (Model 1 is the model used to examine baseline equivalence of the intervention and control groups, while Model 2 is the confirmatory impact analysis model). Complete results for the sensitivity analyses are shown in appendix J.

Table 4.2. Sensitivity analyses for basic composite score

Outcome measure	Adjusted intervention mean	Adjusted control mean	Estimated difference	*p*-value	95 percent confidence Interval	Effect size[a]
Model 3: Adjusted IPT posttest, cases with missing data deleted (see appendix table J3)	–0.01 (–0.04) [0.84]	0.05 (0.10) [0.86]	–0.06 (0.05)	.21	–0.16 to .04	–0.07
Model 4: Unadjusted IPT posttest score, no race/ethnicity and pretests covariates (see appendix table J4)	–0.03 (–0.04) [0.84]	0.08 (0.10) [0.86]	–0.10 (0.07)	.15	–0.25 to .04	–0.12
Model 5: Adjusted IPT posttest, no race/ethnicity covariate (see appendix table J5)	0.00 (–0.04) [0.84]	0.03 (0.10) [0.86]	–0.02 (0.05)	.65	–0.14 to 0.09	–0.03
Model 6: IPT listening posttest subscore (see appendix table J6)	0.00 (–0.04) [0.98]	0.02 (0.09) [1.02]	–0.02 (0.06)	.77	–0.14 to 0.11	–0.02
Model 7: IPT reading posttest subscore (see appendix table J7)	–0.01 (–0.04) [0.99]	0.05 (0.11) [1.00]	–0.06 (0.07)	.35	–0.20 to 0.08	–0.06
Model 8: IPT writing posttest subscore (see appendix table J8)	–0.00 (–0.04) [0.99]	0.03 (0.09) [1.00]	–0.03 (0.07)	.61	–0.17 to 0.10	–0.03
Model 9: Adjusted IPT posttest, eligible for free or reduced-price lunch as covariate (see appendix table J9)	0.00 (–0.04) [0.84]	0.02 (0.10) [0.86]	–0.02 (0.05)	.75	–0.13 to 0.09	–0.02
Model 10: Adjusted IPT posttest, schools with low ELL enrollment excluded (see appendix table J10)	0.02 (–0.04) [0.84]	0.04 (0.10) [0.86]	–0.01 (0.05)	.81	–0.12 to 0.09	–0.01
Sample size (schools)	34	18				

Note: IPT is a series of language proficiency assessments for English language learners in K–12. Numbers in parentheses are unadjusted posttest means for intervention and control groups and standard errors for estimated impact. Numbers in square brackets are standard deviations for the unadjusted means. Estimated difference may not be equal to the difference between adjusted intervention mean and adjusted control mean due to rounding.

a. Effect size is calculated as Glass's Δ (Glass, McGaw, and Smith 1981), in order to present the difference between intervention and control groups in standard deviation units of the control group.

Source: IPT data.

Case deletion of missing data (Model 3)

A total of 1,036 cases (36 percent) were excluded from the analysis because of missing data. Of these, 785 cases (27 percent) were dropped because of missing pretest scores, 172 cases (6 percent) were dropped because of missing posttest scores, and 79 cases (3 percent) were dropped because of missing pretest and posttest scores.

Modeling reveals that the confirmatory impact analysis was not sensitive to the method of treating missing data. The estimated difference between intervention and control group schools on IPT posttest scores for the model that included imputed values for missing pretest data (missing posttest data were not imputed for the impact analysis) ($\gamma_{01} = -0.03$, standard error = 0.06, $p = .66$) (see appendix J, table J2) was about half the estimate for the model that excluded missing data ($\gamma_{01} = -0.06$, standard error = 0.05, $p = .21$) (see appendix J, table J3). The difference was not statistically significant.

Analysis with no race/ethnicity and pretest covariates (Model 4)

To test the robustness of the impact results to the exclusion of covariates, we modeled differences in IPT posttest scores between intervention and control group schools excluding the IPT pretest covariate and the control for the percent of White students and the percent of Hispanic students in the school. The estimated impact from this model was not statistically significantly different from the estimated impact from the confirmatory impact analysis.

Exclusion of race/ethnicity covariates (Model 5)

To test the robustness of the impact results to the inclusion of race/ethnicity covariates, we ran analyses modeling differences in IPT posttest scores between intervention and control group schools excluding the percentage of White students and the percentage of Hispanic students in the school. The estimated impact from this model was not statistically significantly different from the estimated impact from the confirmatory impact analysis.

Subtest analyses (Models 6, 7, & 8)

To test the robustness of the impact results to the use of a composite score, we modeled the differences between schools in the intervention and control groups on the IPT posttest listening, reading, and writing scores. In these analyses, the section scores were substituted for the basic composite scores in the models. In all other aspects, the models estimated were identical to the confirmatory impact model. The estimated impact from these models was not statistically significantly different from the estimated impact from the confirmatory impact analysis.

Inclusion of eligibility for free or reduced-price lunch as a covariate (Model 9)

To test the robustness of the impact results to the inclusion of the percentage of students eligible for free or reduced-price lunch as a covariate, the research team modeled differences in IPT posttest scores between intervention and control group schools including the IPT pretest covariate and the control for the percentage of White students, the percentage of Hispanic students, and the percentage of students eligible for free or reduced-price lunch. The estimated impact from this model was not statistically significantly different from the estimated impact from the confirmatory impact analysis.

Exclusion of schools with low enrollment of ELL students (Model 10)

To test the robustness of the impact results to the ELL enrollment criteria for schools that were eligible for participation, we ran analyses modeling differences in IPT posttest scores between intervention and control group schools excluding schools that did not meet the 20 percent ELL enrollment criteria. The estimated impact from this model was not statistically significantly different from the estimated impact from the confirmatory impact analysis.

5: EXPLORATORY ANALYSES

This chapter presents the results of exploratory analyses that estimated the impact of the combination of OWE and RISE on three sets of intermediate outcomes: teacher-reported student engagement, teacher instructional practices, and teacher instructional responsiveness and assessment practices. No statistically significant differences between teachers in intervention schools and teachers in control schools were found on any of the intermediate outcomes examined. No corrections for multiple hypothesis testing were used for these exploratory analyses, because these analyses were not intended to examine the effectiveness of the intervention (see Schochet, 2008).

Measuring intermediate outcomes

Student engagement has been considered an indicator of successful classroom instruction and a driver of school success. It is theorized that students who are engaged in their learning will be more willing to participate in school activities, attend class, submit required work, and follow teachers' directions in class (see, for instance, Chapman 2003).

One way teachers can improve student engagement is by using curricula students find interesting and challenging. The OWE curriculum is designed to be engaging and to focus on topics relevant to students' lives and experiences in school. RISE was developed to improve teacher communicative competence in delivering effective material for ELL students. We therefore hypothesized that teachers in the intervention group would report that their students had higher levels of engagement with classroom materials than would teachers in the control group. However, the combination of OWE and RISE is unlikely to affect student engagement or English language acquisition without first affecting teachers' classroom behaviors.

We hypothesized that dedicated use of OWE and RISE materials would lead to a significant change in instructional practice for teachers in the intervention group, including increases in the use of differentiated instruction, sheltered instruction, and receptive and expressive language instruction. We also hypothesized that these changes in teachers' classroom instructional practices would lead to improvement in teacher responsiveness and modification of assessment practices, with intervention teachers demonstrating more frequent use of practices such as student-centered instruction and formative assessment.

The following exploratory research questions address the hypothesized impact of OWE and RISE on teacher-reported levels of student engagement, teachers' instructional practices, and teacher responsiveness/modification of instructional practices and assessment practices:

- Does the combination of OWE and RISE have a significant impact on teacher-reported student engagement with ELL-specific educational materials?

- Does the combination of OWE and RISE have a significant impact on teacher-reported instructional practices (differentiated instruction, sheltering instruction, receptive and expressive language instruction, reading instruction, and writing instruction)?

- Does the combination of OWE and RISE have a significant impact on teacher-reported instructional responsiveness and assessment practices (modification of instruction or teacher responsiveness, student-centered instruction, and use of assessments)?

The scale scores used in assessing the impact of OWE and RISE on intermediate outcomes are composites of items designed to measure a specific construct. As described in chapter 2, Rasch modeling techniques were used to assess whether the items held together as scales and to create a score for each teacher on each scale. The scores for intermediate outcomes remain in their original Rasch-produced logits; item difficulties are reported as logits. Although the online teacher log is grounded in research on best practices for ELL instruction and was pilot tested, internal consistency estimates were calculated (see appendix G). Estimates were α = 0.66–0.86.

Teachers were asked to complete online teacher logs three times during Year 2. Analysis of the logs showed that means on each of the intermediate outcomes were the same across all three logs. Because of this, the outcome scores used in the analyses are the teachers' average score across all three logs (see appendix G for scale items for all outcomes).[43]

Teacher reports of student engagement

We asked teachers to report on perceived levels of student engagement using a five-item scale. The items asked teachers to rate student's interest and involvement with ELL-specific materials. Teachers were asked to report on student engagement for students in general rather than for specific students in the class. This scale was used during each wave of data collection. The items were administered using a Likert scale, with score categories of 1–4, where 1 indicates "Never" and 4 indicates "Always."

Teacher instructional practices

Teacher instructional practices were measured on five scales: differentiated instruction, sheltering instruction, receptive and expressive language instruction, reading instruction, and writing instruction. The scale items covered teachers' use of generally accepted principles and practices of ELL classroom instruction. Survey items were administered using a Likert scale, with score categories of 1–4, where 1 indicates "Never" and 4 indicates "Always."

Teacher instructional responsiveness and assessment practices

Teacher instructional responsiveness and assessment practices were measured using three scales: modification of instruction/teacher responsiveness, student-centered instruction, and assessment. The scale items covered teachers' use of instructional modification to meet the needs of students and assessment practices that focus on the use of student-centered instruction and address ELL students' instructional needs. The teacher instructional responsiveness/modification practices and assessment practices measures were intended to measure how teachers modify their lessons to address student needs. Survey items were administered using a Likert scale, with score categories of 1–4, where 1 indicates "Never" and 4 indicates "Always."

[43] The number of scores averaged to create the outcomes score depended on the number of logs completed by each teacher (see chapter 2, table 2.10 for online teacher log response rates).

Findings on exploratory research questions

The exploratory research questions were addressed using models similar to those used to estimate the impact of OWE and RISE on student language acquisition (see appendix H). These models nested teachers within schools. For the analyses of the impact of OWE and RISE on the intermediate outcomes, an intervention indicator at level 2 was used to estimate the impact of the intervention on the outcome as the difference between the adjusted mean outcome for intervention and control schools.

The results indicate that the differences between intervention and control groups were not statistically significant for any of the intermediate outcomes (table 5.1). Complete results are presented in appendix K.

Table 5.1. Analysis results for intermediate outcomes

	Adjusted teacher mean					
Outcome	**Intervention group**	**Control group**	**Estimated difference**	***p*–value**	**95 percent confidence interval**	**Effect size[a]**
Student engagement						
Model 11: Teacher reports of student engagement (see appendix table K1)	3.80	4.59	–0.79 (0.71)	.27	–2.24 to 0.66	–0.32
Instructional practice						
Model 12: Differentiated instruction (see appendix table K2)	3.76	3.47	0.29 (0.32)	.38	–0.37 to 0.94	0.18
Model 13: Sheltering instruction (see appendix table K3)	3.71	3.28	0.43 (0.24)	.08	–0.05 to 0.92	0.36
Model 14: Receptive and expressive language instruction (see appendix table K4)	3.47	3.58	–0.11 (0.33)	.74	–0.79 to 0.57	–0.07
Model 15: Reading instruction (see appendix table K5)[b]	3.57	3.47	0.10 (0.42)	.81	–0.75 to 0.95	0.06
Model 16: Writing instruction (see appendix table K6)[b]	4.20	4.02	0.17 (0.37)	0.64	–0.58 to 0.92	0.09
Teacher instructional responsiveness and						

assessment practice						
Model 17: Modification of instruction/teacher responsiveness (see appendix table K7)	5.46	5.15	0.31 (0.41)	.46	–0.53 to 1.15	0.18
Model 18: Student-centered instruction (see appendix table K8)	2.87	2.51	0.36 (0.29)	.23	–0.24 to 0.96	0.24
Model 19: Assessment practices (see appendix table K9)	3.37	3.44	–0.07 (0.34)	.83	–0.77 to 0.62	–0.04

Note: Sample included 31 intervention group schools and 17 control group schools. Three intervention schools and one control school were excluded from the exploratory analyses because teachers in these schools did not complete any of the online teacher logs. Numbers in parentheses are standard errors. Estimated difference may not be equal to the difference between adjusted intervention mean and adjusted control mean due to rounding.

a. Effect size is calculated as Glass's Δ (Glass, McGaw and Smith 1981), in order to present the difference between intervention and control groups in standard deviation units of the control group.

b. Only teachers who reported teaching reading or writing to their ELL students responded to items on this measure.

Source: Online teacher logs.

6: SUMMARY OF FINDINGS AND STUDY LIMITATIONS

This chapter reports on the fidelity with which the interventions were implemented and summarizes the study's findings on the effectiveness of the combination of OWE and RISE on English language acquisition for Spanish-speaking ELL students in grades 2–5. It also discusses the effect on intermediate outcomes and describes the study's limitations.

Fidelity of implementation

To understand the instructional practices used with ELL students at study sites, the research team examined online teacher logs for teachers in both the intervention and control group, surveys of site coordinators, and interview data from intervention and control teachers. The data revealed that teachers in the intervention group were significantly more likely than teachers in the control group to report using graphic organizers. The difference between intervention and control group teachers in the number of teacher-reported hours of English language development/literacy instruction with ELL students was not statistically significant.

Overall levels of implementation fidelity, calculated using the teacher logs, revealed that 16 percent of intervention teachers self-reported low levels, 51 percent reported moderate levels, and 30 percent reported high levels of implementation. The fact that 80 percent of intervention teachers were moderate or high implementers is consistent with teacher interview data showing that 75 percent of respondents reported that OWE and RISE had influenced their ELL instruction. Of those teachers who reported that OWE and RISE had significantly influenced their instruction, the OWE curriculum and materials had the greatest influence on their ELL instruction. One-quarter of respondents indicated that OWE and RISE had provided them with useful strategies for teaching their ELL students.

Impact of interventions on English language acquisition

The combination of OWE and RISE did not have a statistically significant effect on Spanish-speaking students' acquisition of English, as measure by the IPT. The average impact was –0.03 standard deviation, with an effect size of –0.03, indicating that ELL students in schools using OWE and RISE performed no differently from ELL students in control schools.

Sensitivity analyses were conducted to test the robustness of the impact estimates to methodological decisions regarding missing data imputation, the inclusion of covariates and control variables (IPT basic composite pretest scores and the percentage of Spanish-speaking teachers in the school), and the use of a composite score. The results of these sensitivity analyses suggest that the estimates regarding the impact of OWE and RISE on student English language acquisition are robust to our methodological decisions: neither decisions made by the research team about the analytic model nor the treatment of missing data affected the main findings of the study.

Impact of interventions on intermediate outcomes

The combination of OWE and RISE did not have a statistically significant impact on teacher-reported student engagement, instructional practices, or instructional responsiveness and assessment practices.

Study limitations

Three types of limitations should be considered when interpreting the study findings. The first is external validity or generalizability. The second is potential bias as a result of missing data issues, including non-response and attrition. The third relates to self-report bias.

External validity or generalizability

The following issues may have affected the external validity or generalizability of this study:

- The 2004 edition of OWE was used for this study. An updated edition of OWE was released by Houghton Mifflin Harcourt (formerly Harcourt) in 2010. The extent to which the study findings are applicable to the 2010 edition of OWE is unknown. (The RISE materials were not updated during the course of the study.)

- Most teachers in the intervention group received RISE training from a publisher-trained site coordinator rather than directly from the publisher. Results might have differed had training been provided by the publisher.

- We attempted to test all students in fall 2008 before teacher training in RISE, but delivery of the professional development depended on school- and district-scheduled time for professional development. Moreover, because this study employed a train-the-trainer model, site coordinators—some of whom were also teachers—were trained in RISE before baseline testing. No other teachers completed RISE professional development training before student testing. Teachers did not use the OWE materials before testing students in fall 2008.

- Teachers were instructed to use OWE materials at least 30 minutes a day with their ELL students but were allowed to vary from that recommendation, as would likely be the case under real-world conditions. Results presented in this report provide evidence only of the impact of OWE and RISE under the implementation conditions observed in this study. It should not be inferred from this study that comparable results would be produced under differing conditions, including higher or lower levels of implementation.

- This study used a convenience sample: all participating schools, teachers, and students were volunteers. We did not collect data from eligible teachers who did not participate in the study, nor did we collect data on students whose parents indicated that they did not want their children included in the study. And because the sample was one of convenience, the study's findings can be generalized only to this voluntary sample.

- This study was not designed to have the statistical power to detect effects smaller than 0.35 standard deviations.

- Only one outcome measure (the IPT) was used to assess student gains. Use of additional student outcome measures might have yielded findings that would have provided additional insight on the effectiveness of the intervention.

Missing data

- Three intervention group schools and one control group school were excluded from the exploratory analyses. Because the teachers in these schools did not complete any of the online teacher logs, these schools could not be included in the exploratory analyses.

- The student impact analysis sample was defined as Spanish-speaking ELL students who were present on the day of testing in spring 2010. Students defined as ELL students at the start of the 2009/10 school year but who tested out of English as a second language programming were excluded from this analysis. Records collected from site coordinators in fall 2009 indicated that 74 of 2,009 study students pretested in fall 2008 (3.7 percent) either exited English as a second language services or were placed in maintenance programs.[44]

Self-report bias

Implementation analyses relied upon self-reported frequency measures of teacher training, professional qualifications, classroom environment, and classroom use of instructional strategies.

- Self-report data collected through the teacher background survey and online logs are limited by teachers' accuracy in recalling specific training and professional qualifications, perceptions of classroom environment, and teaching strategies. Teachers may have responded differently to items based on their understanding of various teaching strategies.

- Self-report responses may also be impacted by perceptions of social bias. Survey and log items that may have been subject to social bias include professional qualifications, use of specific teaching strategies, and student engagement. The content and administration of both instruments did not vary between the intervention and control groups. It is therefore unlikely that limitations imposed by self-report data would differentially impact the two groups.

- Exploratory outcomes were based on teacher-reported perceptions of student engagement, teacher instructional practices, and teacher instructional responsiveness and assessment practices. Teachers were asked to report on student engagement for students

[44] Maintenance models varied across sites but typically incorporated ongoing review of student educational progress so that ELL and classroom instructors could assess whether students would benefit from additional English language assistance. If additional instruction was deemed necessary, students could have been offered additional language support (that is, special literacy instruction) or transferred back into the regular ELL program, depending on the ELL instructional model employed at the school.

in general rather than for specific students, and individual teachers may have interpreted this question differently.

APPENDIX A: DESCRIPTION OF INTERVENTIONS AND INSTRUCTIONAL MODELS USED IN PARTICIPATING SCHOOLS

This appendix describes the two interventions studied as well as several programs and approaches for teaching ELL students used in the schools that participated in this study.

Interventions

On Our Way to English (OWE)

OWE, written by David Freeman, Yvonne Freeman, Aurora Colon Garcia, Margo Gottlieb, Mary Lou McCloskey, Lydia Stack, and Cecilia Silva and published in 2003 by Rigby, focuses on the development of academic language, using concepts and vocabulary from state content-area standards. The program design is based on research (Harcourt Achieve 2004b) suggesting that language development for English language learner (ELL) students is enhanced by meaningful contexts, bridges between spoken and written language, and connections between English and students' primary language. The design emphasizes the integration of literacy and oral language during a student's earliest experiences in learning English (Freeman et al. 2004).

OWE incorporates the following principles of language and literacy development for ELL students (Harcourt Achieve 2004a):

1. A focus on language functions, different means of using language to communicate in both social and academic contexts, and English grammar.
2. Culturally sensitive instruction that builds on students' background knowledge.
3. Instruction that emphasizes means of instruction comprehensible to ELL students.
4. Instruction that "recognizes that language and literacy development are two separate developmental processes, and children proceed along each process in phases" (p. 14).
5. Explicit instruction techniques, including explicit instruction in comprehension.
6. Language skills taught in context.
7. Background building experiences.
8. Direct, explicit instruction that builds phonemic awareness skills at a developmentally and instructionally appropriate pace using oral activities (that is, chanting, reading aloud).
9. Vocabulary-to-writing links.
10. Instruction that teaches "the alphabetic principle, phonics skills for decoding, and word study skills at the appropriate times within the context of reading" (p. 28).

Responsive Instruction for Success in English (RISE)

RISE, a professional development program written by Clara Amador-Watson and published in 2004 by Harcourt Achieve, familiarizes teachers with research-based best practices for use in instruction with ELL students. The fundamental principles of teaching ELL students as reflected in RISE are also reflected in the content and format of the OWE curriculum. RISE is guided by several fundamental principles (Knox and Amador-Watson 2002, p. 2):

1. Listening, speaking, and writing develop in integrated environments simultaneously.
2. Oral language development, literacy learning, and academic content should be integrated.
3. Knowledge promotes English academic success.
4. Oral language development and English literacy take place:
    - In communicative environments.
    - In situations in which children are engaged in meaning making.
    - In environments in which modeling, coaching, and opportunities for practice are consistently provided.
    - In anxiety-free settings.
    - As a complex process over time with great individual variation.
5. Primary language development has a positive effect on second language learning.
6. Culture and language are intimately connected.

RISE course content includes information on theories of second language acquisition as well as practical information on strategies to incorporate in ELL instruction, such as use of formative assessment practices and small group instructional techniques. RISE consists of six core and two expansion units, which are presented by trained facilitators (either through a trained-trainer or professional consultant model) during eight, two-hour workshops over the course of one school year (table A1).

Table A1. RISE unit topics and instructional components

Unit	Goal	Instructional components
1 Teacher as Decision-Maker	Extend range of approaches used to teach and assess students.	• Examine research on language, language acquisition, and literacy learning. • Introduce integrated model of English language and literacy development. • Consider role of standards. • Introduce decision-making model (teacher as decision maker, a key design feature of RISE professional development).
2 Assessment of Oral Academic Language Use	Demonstrate understanding of assessment and how it	• Focus on assessment framework; consider extent to which students are meeting or approaching local standards. • Learn to observe student responses to instruction and adjust

Unit	Goal	Instructional components
	guides instruction.	(responsive instruction). • Assess students "on the run" and adjust levels of support and challenge (consistent and purposeful responses to students' performance).
3 Oral and Written Shared Experiences	Connect principles of learning theory to second language learning and literacy development.	• Examine instructional practices in oral and written shared experiences through the development of a cognitive map that includes context building, modeling, scaffolding/validating, utilizing, and extending language in oral and written academic settings. • Consider various ways of assessing students in oral and academic language, literacy, and content through team teaching.
4 Assessment of Reading And Writing	Demonstrate understanding of assessment and how it guides instruction.	• Investigate reading and writing processes for ELL students. • Learn how to teach ELL students to use and apply essential reading strategies while accessing, maintaining, and extending meaning for comprehension and fluency. • Learn to assess ELL students' writing.
5 Small-Group Instruction with Leveled Texts	Use leveled texts to improve comprehension.	• Observe, understand, and develop shared to guided reading lesson design and small-group instruction with leveled texts. • Develop strategic instruction for writing with ELL students.
6 Assessment for Learning and Teaching	Use a practitioner framework for making sound instructional decisions.	• Consider and understand importance of assessment for both learning and teaching. • Reflect upon teacher as decision-maker framework. • Set goals for future professional growth.
7 Coaching English Language Learners in Small Groups	Articulate knowledge of small-group strategic instruction for reading and writing.	• Gain working knowledge of the structure of and criteria for leveled texts for ELL students. • Understand guiding principles for forming dynamic, needs-based groups. • Investigate how to extend academic language through coached reading instruction. • Reflect on role of guided texts in strategic reading instruction for ELL students.
8 Using Data to Improve Student Performance	Demonstrate use of student data to inform teaching.	• Consider importance of consistent, systematic data collection to reflective practice. • Identify expectations for use of academic language across proficiency levels and grade levels based on local standards for ELL students. • Observe model for gathering, interpreting, and using assessment data on oral, reading, and writing skills. • Set professional development goals following analysis of responsive instruction and role of self-assessment.

Source: Amador-Watson 2007.

ELL instructional programs and approaches

Participating schools used various ELL instructional models. Some schools employed multiple models across and within grade levels (for example, small-group instruction at one grade level and whole-class sheltered instruction at another). OWE and RISE were designed for use across a range of instructional models and settings.

Pull-out, small-group instruction

The goal of pull-out, small-group instruction is to teach ELL students English language skills, which may include listening, speaking, reading, writing, study skills, content vocabulary, and cultural orientation. Many strategies are used. Instruction is usually in English, with little use of native language, and targets a subsection of a class or grade level. Instruction is not conducted by a classroom teacher but rather by an English as a second language teacher or literacy coach, often in a specially designated room. Teachers see different groups of students throughout the day.

Structured English immersion

The goal of structured English immersion is acquisition of language skills so that students can succeed in English-only mainstream classrooms. English and English language skills are taught through content areas. The bulk of instruction is in English, although students' primary language can be used to clarify instruction. Instruction is conducted by the classroom teacher in the mainstream classroom with a mixed group of ELL and non–ELL students.

Sheltered instruction

Sheltered instruction is an approach in which teachers use strategies such as physical activities, gestures, and visual aids to make academic instruction in English understandable to ELL students.

Dual-language/bilingual instruction

Dual-language/bilingual instruction includes the following programs:

- *Dual-language (two-way, developmental)*: The goal of a dual language program is for students to develop language proficiency in two languages by receiving instruction in English and another language. Programs can differ in the percentage of primary and second language instruction (full immersion, partial immersion) and in the classroom composition of ELL and non–ELL students. Instruction is conducted by classroom teachers in mainstream classrooms.

- *Maintenance (late-exit)*: The goal of a maintenance program is for students to develop language proficiency in two languages by receiving instruction in English and another language. Instruction builds on the ELL student's primary language skills and develops and expands the English language skills of each student to enable the student to achieve proficiency in both languages while providing access to content areas.

- *Transitional bilingual (early-exit)*: The goal of a transitional bilingual program is to facilitate students' transition to an all-English instructional program while providing academic subject instruction in the native language to the extent necessary. Bilingual programs can differ in the percentage of primary and second language instruction and in the classroom composition of ELL and non–ELL students. Instruction is conducted by the classroom teacher in the mainstream classroom.

Other types of instructional programs

Other instructional programs include the following:

- *Modified pull-out, small-group*: This program is similar to a conventional pull-out, small-group program, except that the pull-out group comprises both ELL and non–ELL students.

- *Large ELL group*: Grade-level students are sorted into groups (for example, non–English proficient, limited English proficient, fluent English proficient, and native speakers) for instruction by a grade-level classroom teacher.

- *Newcomer program*: Newcomer programs are separate, relatively self-contained educational interventions designed to meet the academic and transitional needs of new immigrants. Newcomers attend separate schools with other ELL students before they enter more traditional interventions (for example, schools with English language development programs or mainstream classrooms with supplemental English as a second language instruction).

- *Push-in*: The goal of push-in programs is to teach ELL students English language skills, which may include listening, speaking, reading, writing, study skills, content vocabulary, and cultural orientation. Many strategies are used, although English as a second language instruction is usually in English, with little use of native language. In contrast to a pull-out program, instruction is provided by an English as a second language teacher or literacy coach, who works with ELL students inside the mainstream classroom. Teachers see different groups of students throughout the day.

APPENDIX B: DESIGN ASSUMPTIONS AND POWER ANALYSIS

Consistent with the random assignment of schools to either the intervention or control group, effects were analyzed at the school level, with outcome data collected at the student level. Intervention effects were estimated at the school level to account for the sources of variability in the data that result from the nested structure of the school environment. Moreover, because of the clustered nature of the data, with students nested within schools, we used multilevel modeling (with level 1 reflecting the individual level and level 2 reflecting the school level). Power analyses were conducted for fixed effects. They were augmented by blocking by district because of the random assignment of schools to intervention groups and adjusted to reflect the use of covariates to increase precision.

In conducting the power analysis for the main impact analysis, schools were considered to be the unit of assignment and student achievement the dependent variable. In the absence of research suggesting an appropriate effect size for the combined intervention of On Our Way to English (OWE) (by David Freeman, Yvonne Freeman, Aurora Colon Garcia, Margo Gottlieb, Mary Lou McCloskey, Lydia Stack, and Cecilia Silva and published by Rigby in 2003) and Responsive Instruction for Success in English (RISE) (by Clara Amador-Watson and published by Harcourt Achieve in 2004), we proposed that an effect size of at least 0.35 standard deviation was needed. We considered this value a conservative estimate based on the literature on effective interventions with English language learner (ELL) students addressing English language acquisition and academic achievement (U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse 2008).

An intraclass correlation coefficient of 0.10 was selected, based on the work of Raudenbush et al. (2005), who cite coefficients for educational achievement between 0.05 and 0.15.[45] The midpoint value of this interval (0.10) was assumed for the power analysis in this study. The research team assumed a two-tailed test with $p < .05$, and a desired power greater than 0.80. The sample and cluster size were adjusted to reach this goal. Additionally, we assumed that pull-out classes (the smallest possible unit of classes participating in the study) would contain at least 10 students and that covariates such as previous achievement would yield $R^2 = 0.50$. Power analyses were conducted using the smallest class size potentially participating in this study; when whole classes or mainstream classes of ELL students participated, the class sizes were greater than 10 (indicating that the proposed 10 students per class is a conservative estimate).

Optimal Design software indicated that 36 schools would be required for a power of 0.80. To account for the needed power for the primary research question and potential school attrition (estimated at about 25 percent), 53 schools were sampled, two-thirds of which were assigned to the intervention group.

[45] According to Bloom (2005), the limited empirical literature yields ranges of intraclass correlation coefficients of 0.01–0.10, with most falling between 0.01 and 0.05.

APPENDIX C: TEACHER CHARACTERISTICS

This appendix compares the characteristics of teachers in the intervention and control groups in the exploratory sample (table C1). It describes the characteristics of teachers in the intervention (table C2) and control (table C3) groups in the baseline and exploratory samples.

Table C1. Characteristics of teachers in exploratory sample, by treatment condition

	Intervention group (*n* = 81)		Control group (*n* = 59)				
Characteristic	**Percent**	**Number**	**Percent**	**Number**	**Difference**	**Test statistic**	***p*-value**
Gender							
Male	7.41	6	11.86	7	–4.45	0.81	.37
Female	92.59	75	88.14	52	4.45		
Highest degree earned							
Bachelor	34.57	28	47.46	28	–12.89	3.40	.07
Postgraduate	58.02	47	40.68	24	17.34		
Teaching certificate							
Regular/state standard	81.48	66	66.10	39	15.38	3.62	.06
Other	11.11	9	22.03	13	–10.92		
Earned English as a second language certificate							
Yes	53.09	43	45.76	27	7.33	0.36	.55
No	39.51	32	42.37	25	–2.86		
Spanish proficiency [a]							
None	6.17	5	15.25	9	–9.08	6.47	.09
Minimal	35.80	29	42.37	25	–6.57		
Intermediate	27.16	22	16.95	10	10.21		
Advanced/fluent	23.46	19	13.56	8	9.90		

Note: Significance tests are chi-square tests; * $p < .05$. Numbers may not sum to 100 percent because of missing data.
Source: Teacher background and classroom level data survey.

Table C2. Characteristics of teachers in intervention group in baseline and exploratory samples

	Baseline (n = 136)		Exploratory (n = 81)				
Characteristic	**Percent**	**Number**	**Percent**	**Number**	**Difference**	**Test statistic**	***p*-value**
Gender							
Male	13.24	18	7.41	6	5.83	1.75	.19
Female	86.76	118	92.59	75	–5.83		
Highest degree earned							
Bachelor	33.09	45	34.57	28	–1.48	0.00	.95
Postgraduate	54.41	74	58.02	47	–3.61		
Teaching certificate							
Regular/state standard	76.47	104	81.48	66	–5.01	0.02	.90
Other	11.03	15	11.11	9	–0.08		
Earned English as a second language certificate							
Yes	41.91	57	53.09	43	–11.18	1.64	.20
No	45.59	62	39.51	32	6.08		
Spanish proficiency							
None	5.15	7	6.17	5	–1.02	0.27	.97
Minimal	36.03	49	35.80	29	0.23		
Intermediate	26.47	36	27.16	22	–0.69		
Advanced/fluent	19.85	27	23.46	19	–3.61		

Note: Significance tests are chi-square tests; * $p < .05$. Numbers may not sum to 100 percent because of missing data.
Source: Teacher background and classroom level data survey.

Table C3. Characteristics of teachers in control group in baseline and exploratory samples

Characteristic	Baseline (*n* = 67) Percent	Baseline (*n* = 67) Number	Exploratory (*n* = 59) Percent	Exploratory (*n* = 59) Number	Difference	Test statistic	*p*-value
Gender							
Male	10.45	7	11.86	7	–1.41	0.06	.80
Female	89.55	60	88.14	52	1.41		
Highest degree earned							
Bachelor	46.27	31	47.46	28	–1.19	0.05	.82
Postgraduate	43.28	29	40.68	24	2.60		
Teaching certificate							
Regular/state standard	65.67	44	66.10	39	–0.43	0.04	.84
Other	23.88	16	22.03	13	1.85		
Earned English as a second language certificate							
Yes	41.79	28	45.76	27	-3.97	0.31	.58
No	47.76	32	42.37	25	5.39		
Spanish proficiency							
None	14.93	10	15.25	9	–0.32	0.34	.95
Minimal	47.76	32	42.37	25	5.39		
Intermediate	14.93	10	16.95	10	–2.02		
Advanced/fluent	11.94	8	13.56	8	–1.62		

Note: Significance tests are chi-square tests; * $p < .05$. Numbers may not sum to 100 percent because of missing data.
Source: Teacher background and classroom-level data survey.

APPENDIX D: VALIDITY, RELIABILITY, AND SCORING OF LANGUAGE ACQUISITION OUTCOME MEASURE

The validity and reliability of the student outcome measure, the IDEA Proficiency Test (IPT), published in 2005 by Ballard & Tighe, was reported by the test publisher. Specific procedures and evidence on internal consistency, test content, response processes, bias, relations to other criterion measures, and other indicators of the test-takers' ability are detailed in the test technical manuals for grades 1–2 (Ballard & Tighe 2005a) and 3–5 (Ballard & Tighe 2005b). To supplement the psychometric evidence on the IPT, the REL Central researcher team collected additional, externally published research.

Validity

The publisher provides evidence that the content of the listening, speaking, reading, and writing sections is aligned with underlying constructs and skills for both grade spans. To provide further content and construct evidence, the publisher used English as a second language and mainstream classroom teachers as well as experts to reflect on the proposed items and to comment on whether the items under development made sense. We collected additional validity evidence for the IPT and found some evidence of concurrent validity, with overall student IPT score correlations of .86 with WLPB Oral scores (Schrank, Fletcher, and Alvarado 1996) and correlations between .61 and .66 on ACCESS for ELLs subscales of listening, speaking, reading and writing (Abedi 2007).

Reliability

Published internal consistency estimates for the IPT for listening, speaking, reading, writing, comprehension, and overall English sections are 0.83–0.94 for grades 1–2 and 0.84–0.93 for grades 3–5. Constructed item response consistency estimates were provided by the publisher; however, REL Central scorers were trained specifically for the study to rate these items. Additional information is provided in the Scoring section below.

The research team was trained to administer the IPT. The training, conducted by a member of the test publisher's technical team, took place over a two-day period before the administration of the test in fall 2008. We participated in a refresher training on test administration procedures before the spring 2010 administration. To ensure consistency across test administrations, administrators at both intervention and control sites adhered to the procedures learned in this training session and the script in the test's examiner's manual. With the approval of a representative of the IPT, we developed and adhered to an additional script to instruct students on test-taking procedures (the script included instructions on how to fill in response bubbles).

Scoring

The IPT for grades 1–2 included a descriptive writing section comprising four pictures with accompanying fill-in-the-blank sentences. Children were asked to fill in each blank with an appropriate word to describe the picture. For both the training and the implementation years, we used the Ballard & Tighe rubrics to assess student responses. Rubrics were comprised of lists of representative answers for each of three score points: 2 score points were given for real words that were correctly spelled and accurately completed the sentence describing the picture; 1 score

point was given for approximate/phonetic spellings of real words that accurately completed the sentence describing the picture; and 0 points were given for any other responses. When students provided responses that were not already in the list of representative answers, we expanded the rubric to include those responses.

The IPT for grades 1–2 also included one open-ended narrative writing item, and the IPT for grades 3–5 included three open-ended writing prompts: one descriptive, one narrative, and one extended narrative writing. After administration of the pretest, seven researchers were trained as a group to score these items using the Ballard & Tighe four-point writing rubric and writing samples. [46]

Pairs of raters initially scored all writing items from the spring 2008 pretest; double scoring was gradually reduced to random checks for continued reliability. Eleven rater pairs scored 614 of the 1,614 narrative writing items (38 percent) on the grade 1–2 tests. Interrater agreement for these items (that is, exact agreement between raters) was 82 percent. Scores were within one point on 110 of the 112 items (98 percent) on which rater pairs disagreed. As with the IPT scoring for grades 1–2, two researchers initially scored all writing items from the spring 2008 pretest for grades 3–4 (no grade 5 students were tested in Year 1) and double scoring was gradually reduced to random checks for continued reliability. Of the 1,293 grade 3–4 writing tests, 641 descriptive writing items, 642 narrative writing items, and 632 extended narrative writing items were scored by 19 rater pairs. Interrater agreement for these items was 66 percent for the descriptive writing items, 68 percent for the narrative writing items, and 78 percent for the extended narrative writing items. Scores were within one point on 207 of the 218 descriptive writing items (95 percent), 202 of the 208 narrative writing items (97 percent) and 136 of the 138 extended narrative writing items (99 percent) on which rater pairs disagreed. In cases of disagreement for items scored by two raters, three researchers with extensive experience with ELL students and knowledge of writing constructs arbitrated the final score.

After the administration of the posttest in Year 2, five members of the research team participated in refresher training on scoring the writing items. The focus of the Year 2 scoring was on grade 2 and grades 3–5 tests. All raters scored each prompt independently and then compared ratings. In cases of disagreement, the majority score was used as the final score. Of the 862 grade 2 tests, 166 narrative writing items were groups scored (19 percent). Interrater agreement for these items was 52 percent among all five raters; in 78 percent of cases at least four of five reviewers agreed exactly. Of the 2,014 grade 3–5 tests, 378 (19 percent) were group scored. Interrater agreement for the group-scored descriptive writing training item was 53 percent among all five raters; in 75 percent of cases at least four of five reviewers agreed exactly. Similarly, interrater agreement for the narrative writing training items was 49 percent for narrative writing with four of five reviewers in exact agreement in 78 percent of cases, and 46 percent for extended narrative writing with four of five reviewers in exact agreement in 75 percent of cases. Group scoring was gradually reduced to random checks for continued reliability.

[46] Researchers coded all items independently. Interrater agreement was calculated as the agreement among the group of raters who scored an item. In cases of items that were coded by more than one rater, interrater agreement was calculated as the exact agreement among the pair or group of raters that scored an item. In Year 1, in cases where items were double scored interrater agreement was calculated as the exact agreement among both researchers. In Year 2, in cases where items were scored by five researchers, interrater agreement was calculated as the exact agreement among all five researchers.

For the speaking section, examiners scored student responses while administering the test and then entered the responses into a database. The IPT speaking tests (1–2 and 3–5) are each comprised of 19 items. The test administrator uses a picture book to lead students through a series of questions about items and actions depicted in each picture. Questions, follow-up questions, and correct responses are provided in the examiner script. As members of the research team were trained as a group to administer the speaking section of the IPT, and because 18 of the 19 items have correct responses provided in the examiner script (1 item has a 4-point rubric) reliability on these items is assumed to be high. However, we have no information on the reliability of these items.

For the listening, reading, and multiple-choice writing items, student responses were entered into a database and electronically scored. Fifteen percent of pretests were randomly selected and checked for accuracy of data entry. Of the 104 tests rescored during the pretest screen, 8 percent included one or more data entry errors. Reexamination of data-entry records following the pretest screen indicated that all errors were attributable to one individual. All 441 tests scored by this individual (22 percent of all pretests) were rescored to ensure accuracy. Posttest data entry was conducted by an independent subcontractor, which ensured data quality by employing two-pass verification (that is, double data entry). A third-party analyst resolved discrepancies arising from the two-pass verification process.

APPENDIX E: SURVEY INSTRUMENTS

This appendix presents the survey instruments used in this study. These instruments include the survey of site coordinators, the teacher background and classroom-level data survey, the online teacher log, the interview protocol for intervention group teachers, and the interview protocol for control group teachers. An additional data-collection instrument— a modified version of the English Language Learner Classroom Observation Instrument (ELLCOI) (Gersten et al. 2005; Haager et al. 2003) and a corresponding interview protocol—were dropped from the study. There were three reasons that researchers ceased the collection of observation data. First, findings from the teacher log and post-observation interviews conducted in spring 2009 led them to conclude that two observation points would be insufficient to capture variation in school practice, program structure, and implementation over the study period. Second, 29 percent of the teacher observation sample had withdrawn from training by the time implementation began (in Year 2). Third, teacher practices were not seen as central to this study; they were considered useful insofar as these data provided context for interpreting impacts on student achievement. Researchers received approval to drop the observations from the plan and to supplement teacher log data with data derived from 10-minute telephone interviews with all teachers using an abbreviated version of the follow-up interview originally intended to accompany classroom observations.

Implementation measures

Although an examination of implementation fidelity is not of primary interest in this study, collecting minimal implementation information enables researchers to gauge the degree of intervention fidelity, which is relevant for interpreting the impacts of the study (U.S. Dept. of Education, Institute of Education Sciences, What Works Clearinghouse 2008). Two data sources were employed to monitor and examine fidelity of implementation: During Year 1, the focus of the analysis of implementation fidelity was on whether intervention teachers completed the required training. RISE-trained trainers, who also served as site coordinators, reported on teacher attendance at RISE training sessions. During Year 2, the focus of the analysis of implementation fidelity was on whether intervention teachers used OWE materials and RISE strategies in their instruction. Data obtained from items on the online teacher logs were used to determine implementation of critical intervention components, identified through a review of OWE and RISE materials, as well as discussions with trainers and authors of the programs.

OWE/RISE ELL Study - Site Coordinator & School Information - 2008

1. About You:

First Name: _____ Last Name: _____

School: _____

Mailing Address (Home): _____ City, State, ZIP: _____

Email: _____ Summer email (if different): _____

Telephone #: _____ Fax #: _____

2. What is the best method for contacting you?

- □ Phone
- □ Email

3. Demographic Information

Are you...

- □ Female
- □ Male

What is your ethnic/cultural group? (please check all that apply):

- □ White/Caucasian
- □ Black/African-American
- □ Spanish/Hispanic/Latino
- □ American Indian/Native American
- □ Asian/Pacific Islander

If Spanish/Hispanic/Latino, please select one:

- □ Mexican, Mexican American, Chicano
- □ Puerto Rican
- □ Cuban
- □ Other Spanish/Hispanic/Latino

4. Professional Experience (Please tell us about your professional background):

What is your most advanced degree?

- □ BA/BS
- □ MA/MS
- □ PhD/EdD
- □ Other

How many total years have you worked in K–12 education? _____

5. If you have worked as a school administrator (e.g., principal, program supervisor, professional development coordinator) please complete the following table starting with your most recent position.

	Position	School level (elementary, middle, high, other)	Years of experience	Type of school (rural, urban, suburban)
a.				
b.				
c.				
d.				

6. If you have performed the duties of the positions listed in question 5 without having formally held such positions, please provide us with a brief description of your experience.

_____

_____

7. If you worked as a teacher/faculty member, please complete the following table starting with your most recent position.

	Position	Content area	School level (elementary, middle, high, other)	Years of experience
a.				
b.				
c.				
d.				

8. If you have served as a trainer for adult learners, please complete the following table starting with your most recent position. Please include both formal and informal positions.

	Position/type of work	School/agency/organization	Years of experience
a.			
b.			
c.			
d.			

9. Additional comments: Please share with us any additional comments about your experience or training needs related to RISE:

_____

_____

10. What are the program approaches/plans for English language learners' (ELLs) education at this school? (check all that apply):

- □ ESL pullout (special class for ELLs only)
- □ One-way developmental bilingual classes (one language group being schooled through two languages)
- □ Two-way bilingual classes (two language groups receiving integrated schooling through their two languages)
- □ ESL teachers co-teach with mainstream teachers
- □ Instruction in native language in one or more subject areas
- □ ELLs in mainstream classes with ESL certified teachers
- □ Other (please explain):

11. Are classrooms serving ELLs configured by (select one):

- □ Grade levels
- □ Level of English proficiency (mixed grades)
- □ Other (please explain):

12. How much time per week do ELLs receive direct instruction in learning English? (Please indicate hours per week.) _____

13. How much time per week do ELLs receive instruction that is adapted for ELLs? (Please indicate hours per week.) _____

14. What English as a Second Language (ESL) professional development was offered at this school last year and for how many hours?

	Program name	Number of hours in program
a.		
b.		
c.		
d.		
e.		

15. What ESL professional development is being offered at the school this year?

	Program name	Number of hours in program
a.		
b.		
c.		
d.		
e.		

16. Are all teachers required to attend professional development related to teaching ELLs? *(show/hide trigger question)*

- □ Yes
- □ No

17. What percentage of your teachers attend professional development related to the teaching of ELLs? *(hidden question related to question 16)* _____

18. What percentage of your student population is LEP/ELL? _____

19. Of that, what percent speak Spanish as their dominant oral language? _____

20. What is your school's transience rate for ELLs? _____

Teacher Background and Classroom-Level Data Survey

1. What is your most advanced degree?

- □ BA/BS
- □ MA/MS
- □ Education Specialist
- □ PhD/EdD
- □ Other

2. What type of teaching certificate do you have in the state where you currently teach?[1] (select one):

- □ Regular or standard state certificate or advanced professional certificate
- □ Probationary certificate (the initial certificate issued after satisfying all requirements except the completion of a probationary period)
- □ Provisional or other type of certificate given to persons who are still participating in what the state calls an "alternative certification program"
- □ Temporary certificate (requires some additional college coursework and/or student teaching before regular certification can be obtained)
- □ Emergency certificate or waiver (issued to persons with insufficient teacher preparation who must complete a regular certification program in order to continue teaching)
- □ No certificate

[1] Adapted from NCES, NAEP (2004)

3. Did you have coursework specific to teaching English language learners (ELLs) or English as a Second Language (ESL) methodology in your teacher preparation program (pre-service)?

- □ No
- □ Yes (Indicate hours)

4. Do you have certification in ESL?

- □ No
- □ Yes

5. What staff development (not including college/university teacher preparation classes) directly related to teaching ELLs have you had in the past five years? (For each staff development opportunity, please specify the name of the program, the number of hours you spent in the program, and the year in which you were enrolled in this staff development opportunity.)

	Program name	Number of hours in program	Year of enrollment
1			
2			
3			
4			

6. During the current school year, how many times did you use each of the following resources to obtain information about improving the academic performance of English language learners or English as a Second Language students?

	Never	1–2 times	3–4 times	5+ times
Online websites or databases to find lesson plans, curricular materials, or instructional strategies				
Online websites or databases to find research reports or articles				
Professional journals				
In-service classes and workshops				
Local resources including libraries or local cultural centers				
Consultation with local experts				
Consultation with other teachers in your school				

7. How would you characterize your knowledge of Spanish (level/proficiency)?

- □ None
- □ Minimal
- □ Intermediate
- □ Advanced/Fluent

8. What practices do you normally utilize in your classroom teaching? (check all that apply):

- □ Cooperative learning
- □ Thematic units integrating subject areas across curriculum
- □ Learning styles research applied to bilingual students
- □ Whole language approaches combined with balanced literary instruction
- □ Activating background knowledge (multicultural literature, authentic assessment, community connections)
- □ Other (please explain): _____
- □ Other (please explain): _____
- □ Other (please explain): _____

9. What is your name: _____

10. School Name: _____

11. What is your grade level? _____

Participating Teacher Online Log

1. Background Information

This log provides you an opportunity to give researchers data on your classroom practices, activities, and ELL accommodations or ELL-specific strategies that you have used during your classes. When you respond to the items, please respond based on a typical week over this past quarter (this might be the previous week or two weeks ago—we want to be sure that the data we collect reflect what a typical week might involve, rather than capturing an unusual week—such as one filled with student testing). Also, please note that data collected from the log will be reported in the aggregate—that is, individual teacher names will not be associated with responses to items. Completion of this log should take no more than 20 minutes.

1. What is your first name? _____

2. What is your last name? _____

3. What is your school's name? _____

4. What instructional model do you use with your ELLs? (please check all that apply)

- □ Pull-out/small ELL group (ELLs only)
- □ Modified pull-out/small group (can include non–ELLs)
- □ Dual language (two-way, developmental bilingual)
- □ Maintenance (late-exit bilingual)
- □ Transitional (early-exit bilingual)
- □ Push-in
- □ Newcomer
- □ Other (please specify) _____

2. Organization of Curriculum and Pedagogy

When you respond to the items, please respond based on a typical week over this past quarter (this might be the previous week or two weeks ago – we want to be sure that the data we collect reflect what a typical week might involve, rather than capturing an unusual week – such as one filled with student testing).

5. Differentiated Instruction

	Never	Rarely	Sometimes	Always
I scaffold my lessons for students.				
I provide adequate amounts of support to move students from one level of understanding to a higher level.				
I "level" assignments and homework to align with the students' levels of English language acquisition.				

	Never	Rarely	Sometimes	Always
I deliberately tailor instruction based on students' identified level of English language acquisition.				
I know the level of English language proficiency for each ELL across all domains of comprehension (e.g., measures for listening, speaking, reading, and writing).				

6. Sheltering Instruction

	Never	Rarely	Sometimes	Always
I model skills and strategies for students during lessons.				
I provide prompts and cues in how to use strategies, skills, and concepts (e.g., guided practice, steps, and procedures).				
I facilitate student understanding of both English language and content during lessons by providing comprehensible input (e.g., gestures, realia, visuals, photos, pictures, and drawings, etc.)				
I teach difficult vocabulary prior to lessons, or during lessons as needed.				
I adjust my speaking (e.g., clear expression and articulation; short, simple sentences; eye contact with students; high frequency vocabulary; reduction of idiomatic expressions; nouns more than pronouns) to match students' English proficiency level.				
I use strategies to develop students' use of both social and academic language.				

7. Receptive and Expressive Language Instruction

	Never	Rarely	Sometimes	Always
I provide structured opportunities for children to speak (e.g., conversation, recitation, oral storytelling, songs, chants, answer questions aligned with their level of language acquisition).				
I provide opportunities for students to demonstrate comprehension (e.g., pointing, answering yes/no questions, short answers, listing, labeling).				
I encourage students to elaborate on responses (e.g., prompt students to expand short answers; to provide more information; to give more complete responses).				
I provide opportunities for children to listen to various forms of appropriately leveled input (e.g., oral storytelling, songs, chants).				
I provide time for students to demonstrate what they understand on their own (e.g., written or drawn responses).				

8. I instruct ELLs in reading *(show/hide trigger question)*

- □ Yes
- □ No

9. Reading Instruction *(hidden question related to question 8)*

	Never	Rarely	Sometimes	Always
I read aloud to students.				
I have students read independently at levels aligned with students' stage of English language acquisition.				
I read with students and check for their comprehension by asking ELA appropriate questions (e.g., open-ended questions).				
I model how to read a text (e.g., using reading strategies or pictures).				
I prompt students to use reading strategies (e.g., using pictures, decoding strategies, etc.).				
I assess and build on student background knowledge in reading instruction.				

	Never	Rarely	Sometimes	Always
I group students based on their reading ability regardless of their English fluency.				

10. I instruct ELLs in writing *(show/hide trigger question)*

- □ Yes
- □ No

11. Writing Instruction *(hidden question related to question 10)*

	Never	Rarely	Sometimes	Always
I instruct students in the writing process (e.g., prewriting, drafting, revising, editing, and publishing).				
I model writing strategies with my students (e.g., shared writing, modeled writing, etc.).				
I structure independent writing time for students (e.g., journal writing, reader response).				
I provide instruction in writing mechanics/conventions (e.g., grammar, spelling, punctuation, and capitalization).				
I provide instruction in organization and structure (e.g., topic sentence, supporting details).				
I group students based on their writing ability regardless of their English fluency.				

3. Authentic and Meaningful Instruction

When you respond to the items, please respond based on a typical week over this past quarter (this might be the previous week or two weeks ago—we want to be sure that the data we collect reflect what a typical week might involve, rather than capturing an unusual week—such as one filled with student testing).

12. Modification of Instruction/Teacher Responsiveness

	Never	Rarely	Sometimes	Always
I understand how to assess the content the student knows versus what they can communicate (orally, reading, writing) in English.				
I modify instruction for students as needed during the lesson (e.g., breaking down tasks into smaller/simpler components; modifying assignments to promote success; providing specializing instruction).				
I provide extra instruction, practice, or review for students having difficulty with the task at hand.				

13. Student-Centered Instruction

	Never	Rarely	Sometimes	Always
I select and incorporate student responses, ideas, examples and experiences into lessons.				
I strategically use students' native language to help students understand content.				
I use strategies to assess and activate student background knowledge.				
I expose students to different genres across all areas of instruction (e.g., reading texts, music, and written work).				

4. Assessment

When you respond to the items, please respond based on a typical week over this past quarter (this might be the previous week or two weeks ago—we want to be sure that the data we collect reflect what a typical week might involve, rather than capturing an unusual week—such as one filled with student testing).

14. Assessment

	Never	Rarely	Sometimes	Always
I monitor student learning as I teach (on-the-run assessment).				
I use student work as evidence of learning (assessment for learning).				

	Never	Rarely	Sometimes	Always
I use assessment to plan for instruction (assessment for teaching).				
I use assessment to examine the effectiveness of my teaching (self-assessment).				

5. Student Engagement

When you respond to the items, please respond based on a typical week over this past quarter (this might be the previous week or two weeks ago—we want to be sure that the data we collect reflect what a typical week might involve, rather than capturing an unusual week—such as one filled with student testing).

15. Student Engagement

	Never	Rarely	Sometimes	Always
Students stayed on task during instruction and enjoyed participating in the activities.				
Students showed interest in the materials.				
Students made positive comments about the materials, including the illustrations and graphics.				
Students often talked to each other about the materials and regularly asked questions about the reading content.				
Students requested to see/read more and wanted to access the materials independently during the school day and at home.				

6. Materials, Methods and Strategies

When you respond to the items, please respond based on a typical week over this past quarter (this might be the previous week or two weeks ago—we want to be sure that the data we collect reflect what a typical week might involve, rather than capturing an unusual week—such as one filled with student testing).

16. Which of the following instructional methods and/or strategies do you use with ELLs during instructional time specifically focused on improving the receptive and expressive English language skills of ELLs (ELD time)? Please check all that apply.

- □ Small group instruction (mixed ability groups)
- □ Small group instruction (same ability groups)
- □ Whole group lecture
- □ Whole group demonstration
- □ Whole group direct instruction (e.g., introduction, instruction, guided practice, independent practice)
- □ Dyads (paired instruction)
- □ Peer instruction
- □ Computer-based instruction
- □ Other multimedia-based instruction
- □ One-on-one instruction
- □ Sentence strips
- □ Games
- □ Realia/manipulatives
- □ Graphic organizers
- □ Guided/leveled readers
- □ Shared readers/big books
- □ Songs/chants
- □ Cooperative learning
- □ Drama/guided role play/simulations
- □ Student presentations/report back sessions
- □ Total Physical Response
- □ Discovery learning/inquiry-based learning
- □ Journals
- □ K-W-L
- □ Seat work/worksheets
- □ Word walls/vocabulary walls
- □ Flash cards/vocabulary cards
- □ Native language support
- □ Listening centers
- □ Silent reading centers
- □ Other centers (please specify) _____
- □ Other (please specify) _____

17. Did the pace of your instruction allow you adequate time to address the needs of your ELLs?

- □ No
- □ Somewhat
- □ Yes

18. Based on your experience this quarter, how adequate were your classroom materials in meeting the needs of ELLs at each of the following stages of language proficiency?

	Very Inadequate	Inadequate	Somewhat Adequate	Adequate	Very Adequate	N/A (no students at this level)
Preproduction						
Early Production						
Speech Emergence						
Intermediate Fluency						
Advanced Fluency						

19. Based on your experience this quarter, how adequate were your classroom materials in meeting the needs of ELLs at each of the following stages of reading development?

	Very Inadequate	Inadequate	Somewhat Adequate	Adequate	Very Adequate	N/A (no students at this level)
Emergent Readers						
Early Readers						
Early Fluency Readers						
Fluency Readers						

Note: the remaining questions (20–32) were additional questions for intervention teachers only.

7. OWE/RISE Implementation

The purpose of this section is to get a better understanding of your implementation of the On Our Way to English (OWE) and Responsive Instruction for Success in English (RISE) programs. We are interested in knowing which of the components and materials you implement, the extent to which you implement them, and the nature of your use. Some of the questions will ask you to reflect on an average week. When you respond to the items, please respond based on a typical week over this past quarter (this might be the previous week or two weeks ago – we want to be sure that the data we collect reflect what a typical week might involve, rather than capturing an unusual week – such as one filled with student testing).

20. Did you implement OWE this quarter? *(show/hide trigger question)*

- □ Yes
- □ No

21. How many days do you use OWE with your ELLs in an average week? (*hidden question related to question 20)*

- □ 0
- □ 1
- □ 2
- □ 3
- □ 4
- □ 5

22. Did you need to supplement your OWE curriculum with other material(s) this quarter? (*hidden question related to question 20)*

- □ Yes
- □ No

23. If yes, with which students (or for which literacy area [e.g., reading, writing]) are you most likely to supplement OWE? (*hidden question related to question 20 & 22)*

_____

_____

24. If yes, which published curriculum/curricula do you use (e.g., Hampton-Brown Avenues)? (*hidden question related to question 20 & 22)*

_____

_____

25. On average, how many minutes per day do you use OWE materials with ELLs at each stage of language proficiency? (*hidden question related to question 20)*

	0 minutes	1–29 minutes	30+ minutes	N/A (no students at this level)
Preproduction				
Early Production				
Speech Emergence				
Intermediate Fluency				
Advanced Fluency				

26. On average, how many minutes per day do you use OWE materials with ELLs at each stage of reading development? (*hidden question related to question 20)*

	0 minutes	1–29 minutes	30+ minutes	N/A (no students at this level)
Emergent Readers				
Early Readers				
Early Fluency Readers				
Fluency Readers				

27. On average, how many minutes per day do you use OWE materials with ELLs at each grade level? (*hidden question related to question 20)*

	0 minutes	1–29 minutes	30+ minutes	N/A (no students at this level)
Grade 1				
Grade 2				
Grade 3				
Grade 4				
Grade 5				

28. Are reading/ELD instructional periods missed or disrupted in a typical week? (*hidden question related to question 20)*

□ No

□ Yes (please explain) _____

29. How much weekly planning and preparation time was required for one complete OWE lesson? (*hidden question related to question 20)*

Minutes _____

30. Did you use RISE strategies with your ELLs this quarter? *(show/hide trigger question)*

□ Yes

□ No

31. During a typical week, how often do you incorporate the following OWE/RISE strategies during ELL instruction? (*hidden question related to question 30)*

	Never	Rarely	Sometimes	Always
I use the three Instructional Lenses for ELL Responsive Instruction (content-area learning, literacy instruction, and oral language development) when planning instruction.				
I use the Teacher as Decision-Maker Framework to plan my instruction (standards, standards-based assessment, Responsive Instruction process, and assessment for teaching and learning).				
I use the Authentic Oral Reading Assessment to assess the reading proficiency of my ELLs.				
I use the Modified Apprenticeship Model for English Language Learning (modeling, scaffolding, validating) to support my ELLs.				
I use the Balanced Literacy instructional approaches (modeled, shared, guided, and independent) to transition from teacher support to learner responsibility with my ELL students.				
I use the process delineated in the Shared Writing Lesson Planning Form (decide the focus, set the scene, write the text, return to the text) in shared writing lessons.				
I use a three-step process (description,				

	Never	Rarely	Sometimes	Always
demonstration, statement) when coaching ELLs in reading.				
I use a combination of grade level standards, multi-leveled texts, and classroom environment supports to help students make thematic connections.				

32. What else should we know about your use of the OWE materials and/or RISE strategies during this quarter?

_____

Interview Protocol for Teachers (Intervention)

1. What instructional model do you use with your students? *(pullout/small ELL group (ELLs only), modified pullout/small group (can include non-ELLs), dual language (two-way, developmental bilingual), maintenance (late-exit bilingual), transitional (early-exit bilingual), push in, newcomer, other)*
    a. If "other," please specify
2. Please describe how ELD is delivered to your students *(how many days per week/hours per day…)*
3. To what extent have you incorporated RISE strategies this school year? *(a great extent, somewhat, a little, not at all)*
    a. If "a little" or "not at all" please explain why not
    b. Was there anything you particularly liked or disliked about the RISE professional development?
4. To what extent have you incorporated OWE curricular materials in your instruction? *(a great extent, somewhat, a little, not at all)*
    a. If "a little" or "not at all" please explain why not
    b. Was there anything you particularly liked or disliked about the OWE curriculum materials?
5. Have OWE and/or RISE significantly influenced your ELL instruction? *(yes, no)*
    a. If "yes," please explain
6. Do you plan to use OWE and RISE next year? *(yes, no)*
    a. If "no," please explain
7. Are you receiving adequate support at the <u>school level</u> for your ELL instruction? *(yes, no)*
    a. If "yes," please explain; if "no," in what ways/areas do you need additional support?
8. Are you receiving adequate support at the <u>district level</u> for your ELL instruction? *(yes, no)*
    a. If "yes," please explain; if "no," in what ways/areas do you need additional support?
9. To what extent do you feel the instructional practices in use at your school adequately address the needs of your ELL students *(a great extent, somewhat, a little, not at all)*
    a. If "a little" or "not at all" please explain why not
10. Do you have any final thoughts you'd like to share?

Interview Protocol for Teachers (Control)

1. What instructional model do you use with your students? *(pullout/small ELL group (ELLs only), modified pullout/small group (can include non-ELLs), dual language (two-way, developmental bilingual), maintenance (late-exit bilingual), transitional (early-exit bilingual), push in, newcomer, other)*
    a. If "other," please specify
2. Please describe how ELD is delivered to your students *(how many days per week/hours per day…)*
3. Are you receiving adequate support at the <u>school level</u> for your ELL instruction? *(yes, no)*
    a. If "yes," please explain; if "no," in what ways/areas do you need additional support?
4. Are you receiving adequate support at the <u>district level</u> for your ELL instruction? *(yes, no)*
    a. If "yes," please explain; if "no," in what ways/areas do you need additional support?
5. To what extent do you feel the instructional practices in use at your school adequately address the needs of your ELL students *(a great extent, somewhat, a little, not at all)*
    a. If "a little" or "not at all" please explain why not
6. Do you have any final thoughts you'd like to share?

APPENDIX F: IMPLEMENTATION OF THE INTERVENTIONS

This appendix describes the implementations of the interventions the classroom materials used in this study, entitled On Our Way to English (OWE) (by David Freeman, Yvonne Freeman, Aurora Colon Garcia, Margo Gottlieb, Mary Lou McCloskey, Lydia Stack, and Cecilia Silva and published in 2003 by Rigby), and the professional development program, entitled Responsive Instruction for Success in English (RISE) (written by Clara Amador-Watson and published in 2004 by Harcourt Achieve). It describes the timing of RISE training (table F1), the delivery of OWE and RISE materials delivery (table F2), and OWE training during Year 1 (table F3); shows the percentage of study teachers attending RISE training sessions during Year 1 (table F4); and documents the data on fidelity collected from the teacher logs (table F5).

Table F1. Training in Responsive Instruction for Success in English, by state location

	Dates	Number of schools	Number of site coordinators
Colorado (training option 1)	September 3–5, 2008	12	20
Colorado (training option 2)	September 16–18, 2008	8	13
Kansas	September 29–October 1, 2008	7	16
Nebraska	September 10–12, 2008	7	19
Total		34	68

Table F2. Month materials received by intervention schools, by district

District	Responsive Instruction for Success in English	On Our Way to English
1	September 2008	November 2008
2	January–February 2009	December 2008
3	February 2009	November 2008
4	September 2008	November 2008
5	February 2009	January 2009
7	October 2008	January 2009
9	February 2009	February 2009
10	February 2009	January 2009
11	February 2009	January 2009
12	February 2009	November 2008
13	February 2009	November 2008

Note: The middle column lists the date that RISE materials were received by study teachers rather than site coordinators. Site coordinators received RISE materials during training sessions in September 2008. They began delivering RISE instruction before the shipment of materials to individual teachers; because of publisher delays, these materials were not shipped to some sites until February 2009.

Table F3. Month and year of training in On Our Way to English (OWE) in intervention schools, by district

District	Month and year
1	November 2008
2	November 2008
3	November 2008
4	November 2008
5	May 2009
7	April 2009
9	May 2009
10	April 2009
11	April 2009
12	November 2008
13	November 2008

Table F4. Percentage of baseline study teachers attending Responsive Instruction for Success in English (RISE) training sessions during Year 1, by district and module

State/district	Percentage of teachers reporting completing module[a]								Number (percentage) of teachers reporting	Number of enrolled teachers (*n* = 136)
	1	2	3	4	5	6	7	8		
Colorado										
District 1	100	100	94.12	94.12	94.12	94.12	94.12	94.12	17 (94.44)	18
District 2[b]	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	4 (22.22)	18
District 3	100	88.24	100	100	100	100	100	76.47	17 (100)	17
District 4	100	88.89	88.89	88.89	88.89	88.89	88.89	0	9 (100)	9
District 5	100	100	100	100	70.00	70.00	30.00	30.00	10 (90.91)	11
District 7	100	100	100	100	100	100	100	100	12 (100)	12
Kansas										
District 9[c]	86.21	93.10	58.62	62.07	65.52	62.07	62.07	65.52	29 (96.67)	30
District 11	100	100	100	100	100	100	100	60.00	10 (100)	10
Nebraska										
District 6[d]	d.	d.	d.	d.	d.	d.	d.	d.	d.	d.
District 12	d.	d.	d.	d.	d.	d.	d.	d.	d.	d.
District 13	100	100	100	100	100	100	100	100	3 (75.00)	4
Cross-site average (reporting teachers only)	91.30	94.78	85.22	86.09	84.35	83.48	80.00	66.96	115 (84.56)	136

Note: Table A1 describes each RISE module.

a. Percentages are based on number of teachers at each site reporting RISE completion rates.

b. District dropped from study at onset Year 2.

c. Districts 8 and 10 are comprised only of control schools and, as such, are not represented in this table.

d. Data in cells with fewer than three entries are not displayed in order to protect anonymity.

Table F5. Teacher-reported fidelity of implementation

Question	Category	Percentage of Teachers
Did you implement OWE this quarter?	Yes =	86.87
	No =	13.13
How many days do you use OWE with your ELLs in an average week?	1 =	4.23
	2 =	13.67
	3 =	17.83
	4 =	23.40
	5 =	40.90
On average, how many minutes per day do you use OWE with ELLs at each grade level? (averaged across grade levels)	0 min =	6.91
	1–29 min =	40.20
	30+ min =	52.89
Did you use RISE strategies with your ELLs this quarter?	Yes	53.73
	No	46.27
In a typical week, how often do you incorporate the following OWE/RISE strategies during ELL instruction?		
I use the three Instructional Lenses for ELL Responsive Instruction when planning instruction.	Never =	7.67
	Rarely =	9.10
	Sometimes =	59.77
	Always =	23.43
I use the Teacher as Decision-Maker Framework to plan my instruction.	Never =	9.23
	Rarely =	9.90
	Sometimes =	63.33
	Always =	17.53
I use Authentic Oral Reading Assessment to assess the reading proficiency of my ELLs.	Never =	32.60
	Rarely =	21.17
	Sometimes =	34.87
	Always =	11.33
I use the Modified Apprenticeship Model for English Language Learning to support my ELLs.	Never =	18.47
	Rarely =	12.23
	Sometimes =	49.53
	Always =	19.77
I use the Balanced Literacy instructional approaches to transition from teacher support to learner responsibility with my ELLs.	Never =	3.03
	Rarely =	5.37
	Sometimes =	54.87
	Always =	36.70

Question	Category	Percentage of Teachers
I use the process delineated in the Shared Writing Lesson Plan Form in shared writing lessons.	Never =	17.07
	Rarely =	18.60
	Sometimes =	56.57
	Always =	7.77
I use a three-step process when coaching ELLs in reading.	Never =	14.53
	Rarely =	19.70
	Sometimes =	58.23
	Always =	7.60
I use a combination of grade level standards, multi-leveled texts, and classroom environment supports to help students make thematic connections.	Never =	4.63
	Rarely =	10.73
	Sometimes =	51.10
	Always =	33.53

Note: OWE is On Our Way to English.

Note: Percentages represent average teacher responses across all three teacher logs. Missing log data were not figured into the average implementation fidelity score: if data were missing, the implementation fidelity score was calculated as the average of data from the existing log or logs. Where data were missing from two logs, item scores from the one available log were used to calculate the implementation fidelity score.

Source: Online teacher logs.

APPENDIX G: SURVEY SCALES

Each of the measures included in the online teacher logs was created to assess different aspects of teacher instructional practices, instructional responsiveness/modification and assessment practices, and student engagement. The measures were created using Rasch modeling, in which the probability of a response is modeled as a logistic function of person and item parameters. The Rasch measure for each item, also known as the item difficulty, refers to the mean of item threshold estimate in polytomous Rasch models. The internal consistency of the survey items was reported as a scale reliability estimate, which is analogous to Cronbach's alpha.

Student engagement

Student engagement measures the extent to which teachers self–report that their ELL students were interested and engaged in the ELL-focused materials. All five items were retained in the final version of this measure (table G1).

Table G1. Difficulty of online teacher log items on student engagement

Item	Difficulty
Students stayed on task during instruction and enjoyed participating in the activities.	–1.53
Students showed interest in the materials.	–1.47
Students made positive comments about the materials, including illustrations and graphics.	–0.14
Students often talked to each other about the materials and regularly asked questions about the reading content.	0.96
Students requested to see/read more and wanted to access the materials independently during the school day and at home.	2.18

Note: Scale reliability is 0.86.

Source: Online teacher logs.

Teacher instructional practices

Differentiated instruction measures the extent to which teachers tailor their instruction to meet the needs of their English language learner (ELL) students. All five items fit together as expected and were retained in the final measure (table G2).

Table G2. Difficulty of online teacher log items on differentiated instruction

Item	Difficulty
I scaffold my lessons for students.	–0.96
I provide adequate amounts of support to move students from one level of understanding to a higher level.	–0.74
I know the level of English language proficiency for each ELL student across all domains of comprehension (for example, measures for listening, speaking, reading, and writing).	0.38
I deliberately tailor instruction based on students' identified level of English language acquisition.	0.38
I "level" assignments and homework to align with students' levels of English language acquisition.	0.95

Note: Scale reliability is 0.77.

Source: Online teacher logs.

Sheltering instruction measures the extent to which teachers facilitate understanding through the use of various instruction techniques and strategies. One item—"I teach difficult vocabulary before lessons or during lessons, as needed"—was deleted from the final version of this measure because of lack of fit.

Table G3. Difficulty of online teacher log items on sheltering instruction

Item	Difficulty
I model skills and strategies for students during lessons.	–0.63
I provide prompts and cues in how to use strategies, skills, and concepts (for example, guided practice, steps, and procedures).	–0.28
I adjust my speaking to match students' English proficiency levels.	0.04
I facilitate student understanding of both English language and content during lessons by providing comprehensible input (for example, gestures, realia, visuals, photos, pictures, and drawings, and so forth).	0.16
I use strategies to develop students' use of both social and academic language.	0.71

Note: Scale reliability is 0.75.

Source: Online teacher logs.

Receptive and expressive language instruction measures the extent to which teachers self-reported encouraging their ELL students to listen and demonstrate comprehension and understanding in English. All five items fit together as expected and were retained in the final measure (table G4).

Table G4. Difficulty of online teacher log items on receptive and expressive language instruction

Item	Difficulty
I provide opportunities for students to demonstrate comprehension (for example, pointing, answering yes/no questions, short answers, listing, labeling).	–0.94
I provide structured opportunities for children to speak (for example, conversation, recitation, oral storytelling, songs, chants, answer questions aligned with their level of language acquisition).	–0.53
I encourage students to elaborate on responses (for example, prompt students to expand short answers; to provide more information; to give more complete responses).	0.32
I provide time for students to demonstrate what they understand on their own (for example, written or drawn responses).	0.35
I provide opportunities for children to listen to various forms of appropriately leveled input (for example, oral storytelling, songs, chants).	0.79

Note: Scale reliability is 0.77.

Source: Online teacher logs.

Reading instruction measures the extent to which teachers provide their ELL students with skills, strategies, and opportunities to improve their reading skills in English. One item—"I group students based on their reading ability regardless of their English fluency"—was dropped from this measure because of lack of fit (table G5).

Table G5. Difficulty of online teacher log items on reading instruction

Item	Difficulty
I prompt students to use reading strategies (for example, using pictures, decoding strategies, and so forth).	–0.99
I assess and build on student background knowledge in reading instruction.	–0.64
I model how to read a text (for example, using reading strategies or pictures).	–0.31
I read with students and check for their comprehension by asking English language acquisition-appropriate questions (for example, open-ended questions).	0.86
I have students read independently at levels aligned with students' stage of English language acquisition.	1.08

Note: Scale reliability is 0.76.

Source: Online teacher logs.

Writing instruction measures the extent to which teachers provide their ELL students with skills, strategies, and opportunities to improve their writing skills in English. One item—"I group students based on their writing ability regardless of their English fluency"— was deleted from this measure because of lack of fit.

Table G6. Difficulty of online teacher log items on writing instruction

Item	Difficulty
I provide instruction in writing mechanics/conventions (for example, grammar, spelling, punctuation, and capitalization).	–0.58
I model writing strategies with my students (for example, shared writing, modeled writing, and so forth).	–0.53
I instruct students in the writing process (for example, prewriting, drafting, revising, editing, and publishing).	–0.14
I structure independent writing time for students (for example, journal writing, read response).	0.43
I provide instruction in organization and structure (for example, topic sentence, supporting details).	0.82

Note: Scale reliability is 0.85.

Source: Online teacher logs.

Teacher instructional responsiveness/modification and assessment practices

Modification of instruction/teacher responsiveness measures the extent to which teachers respond and modify their instructional practices to meet the needs of their ELL students. All three items were retained in the final version of this measure (table G7).

Table G7. Difficulty of online teacher log items on modification of instruction/teacher responsiveness

Item	Difficulty
I provide extra instruction, practice, or review for students having difficulty with the task at hand.	–0.99
I modify instruction for students as needed during the lesson (for example, breaking down tasks into smaller/simpler components; modifying assignments to promote success; providing specializing instruction).	–0.69
I understand how to assess the content the student knows versus what they can communicate (orally, reading, writing) in English.	1.68

Note: Scale reliability is 0.66.

Source: Online teacher logs.

Student-centered instruction measures the extent to which teachers incorporate student background knowledge and interests into their curriculum and instructional practices. All four items were retained in the final version of this measure (table G8).

Table G8. Difficulty of online teacher log items on student-centered instruction

Item	Difficulty
I use strategies to assess and activate student background knowledge.	–1.22
I expose students to different genres across all areas of instruction (for example, reading texts, music, and written work).	–0.55
I select and incorporate student responses, ideas, examples, and experiences into lessons.	–0.13
I strategically use students' native language to help students understand content.	1.90

Note: Scale reliability is 0.70.
Source: Online teacher logs.

Assessment practices measures teachers' use of a variety of assessment practices. All four items were retained in the final version of this measure (table G9).

Table G9. Difficulty of online teacher log items on assessment practices

Item	Difficulty
I use student work as evidence of learning (assessment for learning).	–0.43
I monitor student learning as I teach (on-the-run assessment).	–0.38
I use assessment to examine the effectiveness of my teaching (self-assessment).	0.39
I use assessment to plan for instruction (assessment for teaching).	0.42

Note: Scale reliability is 0.82.
Source: Online teacher logs.

APPENDIX H: BASELINE EQUIVALENCE, CONFIRMATORY IMPACT, SENSITIVITY, AND EXPLORATORY ANALYSIS MODEL SPECIFICATIONS

This appendix presents the estimation models used in the baseline equivalence, confirmatory impact, sensitivity, and exploratory analysis models.

Model 1: specification of baseline equivalence model

The baseline equivalence models for each level are as follows:

Level 1:

$$PRE_{ij} = \beta_{0j} + \beta_{1j}(Grade\ 3)_{ij} + \beta_{2j}(Grade\ 4)_{ij} + \beta_{3j}(Grade\ 5)_{ij} + r_{ij}$$

where

- PRE_{ij}, is the IPT pretest score for student *i* in school *j*.
- β_{0j} is the adjusted mean IPT pretest score for students in school *j*.
- $\beta_{1j...}\beta_{3j,}$ are the fixed level 1 covariate effects for grades 3–5.
- r_{ij} is the random error for student *i* in school *j*.

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(Intervention)_j + \gamma_{02}(BLOCK2)_j + \gamma_{03}(BLOCK3)_j + \gamma_{04}(BLOCK4)_j + \gamma_{05}(BLOCK5)_j + \gamma_{06}(BLOCK6)_j + \gamma_{07}(BLOCK7)_j + \gamma_{08}(BLOCK8)_j + \gamma_{09}(BLOCK9)_j + \gamma_{010}(BLOCK10)_j + \gamma_{011}(BLOCK11)_j + \gamma_{012}(BLOCK12)_j + u_{0j},$$

$$\beta_{1j} = \gamma_{10} \ldots \beta_{3j} = \gamma_{30}$$

where

- γ_{00} is the adjusted mean IPT pretest score across schools.
- γ_{01} is the difference between intervention and control schools in the adjusted mean school IPT pretest score.
- γ_{02}–γ_{012} are the additive effects of each district block used in the random assignment of schools (with the first block omitted).
- γ_{10}–γ_{30} are the average regression slopes for highest degree earned, English as a second language certification, and Spanish fluency fixed across schools.
- u_{0j} is the random error for school *j*.

Model 2: specification of confirmatory impact analysis model

Level 1 of the model was specified as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(IPT\ Pretest) + \beta_{2j}(Grade\ 3) + \beta_{3j}(Grade\ 4) + \beta_{4j}(Grade\ 5) + r_{ij}$$

where

- Y_{ij} is the IPT posttest score for student *i* in school *j*.
- β_{0j} is the adjusted mean IPT posttest score for students in school *j*.
- β_{1j} is the effect of previous English language ability on current English language ability.
- $\beta_{2j...}\beta_{4j,}$ are the fixed level 1 covariate effects for grades 3–5.
- r_{ij} is the random error for student *i* in school *j*.

Level 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(Intervention)_j + \gamma_{02}(\%\ White)_j + \gamma_{03}(\%\ Hispanic)_j + \gamma_{04}(Block2)_j + \gamma_{05}(Block3)_j + \gamma_{06}(Block4)_j + \gamma_{07}(Block5)_j + \gamma_{08}(Block6)_j + \gamma_{09}(Block7)_j + \gamma_{10}(Block8)_j + \gamma_{011}(Block9)_j + \gamma_{012}(Block10)_j + \gamma_{013}(Block11)_j + \gamma_{014}(Block12)_j + \gamma_{015}(Block13)_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} \ldots \beta_{4j} = \gamma_{40}$$

where

- γ_{00} is the regression adjusted mean IPT basic composite score for control schools.
- γ_{01} is the difference between intervention and control schools in the regression-adjusted mean IPT basic composite score (intervention effect).
- γ_{02} is the effect of the percentage of White students in the school on the adjusted mean IPT basic composite score.
- γ_{03} is the effect of the percentage of Hispanic students in the school on the adjusted mean IPT basic composite score.
- γ_{04}–γ_{01} are the additive effects of each district block used in the random assignment of schools, with Block 1 as the omitted block.
- γ_{10}–γ_{40} are the average regression slopes for IPT pretest and grade fixed across schools.
- u_{0j} is the random error for school *j*.

Models 3-10: specification of sensitivity analysis models

Researchers conducted eight sensitivity analyses on the confirmatory model. Each is described below.

Model 3: case deletion of missing data

To examine the robustness of the confirmatory impact model to the method of treating missing data, the confirmatory impact model (Model 2) was estimated using case deletion of missing data (rather than imputing data).

Model 4: analysis with no race/ethnicity and pretest covariates

To test the robustness of the confirmatory impact model (Model 2) to the exclusion of covariates, researchers modeled differences in IPT posttest scores between intervention and control group schools excluding the IPT pretest covariate and the control for the percent of White students and the percent of Hispanic students in the school.

Model 5: exclusion of race/ethnicity covariates

To test the robustness of the confirmatory impact model (Model 2) to the inclusion of race/ethnicity covariates, researchers ran analyses modeling differences in IPT posttest scores between intervention and control group schools excluding the percentage of White students and the percentage of Hispanic students in the school.

Models 6, 7 & 8: subtest analyses

To test the robustness of the confirmatory impact model (Model 2) to the use of a composite score, researchers modeled the differences between schools in the intervention and control groups on the IPT posttest listening (Model 6), reading (Model 7), and writing (Model 8) scores.

Model 9: inclusion of eligibility for free or reduced-price lunch as a covariate

To test the robustness of the confirmatory impact model (Model 2) to the inclusion of the percentage of students eligible for free or reduced-price lunch as a covariate, researchers modeled differences in IPT posttest scores between intervention and control group schools including the IPT pretest covariate and the control for the percentage of White students, the percentage of Hispanic students, and the percentage of students eligible for free or reduced-price lunch.

Model 10: exclusion of schools with low enrollment of ELL students

To test the robustness of the confirmatory impact model (Model 2) to the ELL enrollment criteria for schools that were eligible for participation, researchers ran analyses modeling differences in IPT posttest scores between intervention and control group schools excluding schools that did not meet the 20 percent ELL enrollment criteria for the study.

Models 11–19: specification of exploratory analysis models

For the analyses of the impact of OWE and RISE on the intermediate outcomes, an intervention indicator at level 2 was used to estimate the impact of the intervention on the outcome as the difference between the adjusted mean outcome for intervention and control schools. Nine exploratory models (Models 11–19) were estimated, one for each exploratory outcome:

Model 11	*Student engagement:* Teacher reports of student engagement
Model 12	*Instructional practice:* Differentiated instruction
Model 13	*Instructional practice:* Sheltering instruction
Model 14	*Instructional practice:* Receptive and expressive language instruction
Model 15	*Instructional practice:* Reading instruction
Model 16	*Instructional practice:* Writing instruction
Model 17	*Teacher instructional responsiveness and assessment practice:* Modification of instruction/teacher responsiveness
Model 18	*Teacher instructional responsiveness and assessment practice:* Student-centered instruction
Model 19	*Teacher instructional responsiveness and assessment practice:* Assessment practices

Researchers used the model specification for each, varying only the outcome as shown above. At the teacher level (level 1), the models include measures of teachers' education (highest degree earned), expertise in English as a second language (certification), and fluency in Spanish. All variables in the level 1 models are grand-mean centered, which is appropriate when a level 2 predictor is the variable of substantive interest (Enders and Tofighi 2007). Having a bachelor's degree as the highest degree attained and having no Spanish-speaking ability are the omitted variable categories.

Level 1:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(Masters)_{ij} + \beta_{2j}(Ph.D)_{ij} + \beta_{3j}(ESL\ Cert)_{ij} + \beta_{4j}(Min\ Spanish)_{ij} + \beta_{5j}(Int\ Spanish)_{ij} + \beta_{6j}(Adv\ Span)_{ij} + r_{ij}$$

where

- Y_{ij} is the teacher pedagogical practice outcome for teacher *i* in school *j*.
- β_{0j} is the adjusted mean teacher pedagogical practice outcome in school *j*.
- β_{1j} is the relationship between having a master's degree and the adjusted mean teacher pedagogical practice outcome in school *j*.
- β_{2j} is the relationship between having a Ph.D. and the adjusted mean teacher pedagogical practice outcome in school *j*.

- β_{3j} is the relationship between having certification in English as a second language and the adjusted mean teacher pedagogical practice outcome in school *j*.

- β_{4j} is the relationship between having minimal Spanish fluency and the adjusted mean teacher pedagogical practice outcome in school *j*.

- β_{5j} is the relationship between having intermediate Spanish fluency and the adjusted mean teacher pedagogical practice outcome in school *j*.

- β_{6j} is the relationship between having advanced Spanish fluency and the adjusted mean teacher pedagogical practice outcome in school *j*.

- r_{ij} is the random error for teacher *i* within school j.

At the school level (level 2), the models include dummy variables to indicate group assignment (intervention or control), as well as the district blocking variables used in randomization. The intervention indicator variable remains in the original dummy metric; the district blocking variables are grand mean centered. Block 1 is the omitted district blocking variable.

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(Intervention)_j + \gamma_{02}(Block3)_j + \gamma_{03}(Block4)_j + \gamma_{04}(Block5)_j + \gamma_{05}(Block6)_j + \gamma_{06}(Block7)_j + \gamma_{07}(Block8)_j + \gamma_{08}(Block9)_j + \gamma_{09}(Block10)_j + \gamma_{010}(Block11)_j + \gamma_{011}(Block12)_j + \gamma_{012}(Block13)_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} \ldots \beta_{6j} = \gamma_{60}$$

where

- γ_{00} is the adjusted mean teacher pedagogical practice outcome for control schools.

- γ_{01} is the difference between intervention and control schools in adjusted mean teacher outcome (intervention effect).

- $\gamma_{02} - \gamma_{012}$ are the additive effects of each district block used in the random assignment of schools (with the first block omitted).

- γ_{10}–γ_{60} are the average regression slopes for highest degree earned, certification in English as a second language, and Spanish fluency fixed across schools.

- u_{0j} is the random error for school *j*.

APPENDIX I: VARIANCE COMPONENTS

Partitioning variance and covariance allows researchers to examine between and within group variance. Variance components were estimated from an unconditional model (that is, a model with no covariates) to estimate the proportion of variance in the between schools outcome. The variance components were estimated using the IDEA Proficiency Test (IPT) (published in 2005 by Ballard & Tighe) pretest scores to confirm that the use of multilevel modeling was warranted. Researchers partitioned the variance into σ^2—the variance of the residual or individual-level component from level 1 (e_{ij}), and τ_{00}—the variance of the intercept or school-level residual component from level 2 (u_{0j}). These estimates are necessary for calculating the intraclass correlation (ρ, the measure of the proportion of total variance that is between schools). Intraclass correlation coefficients were calculated by dividing the total explainable variance at level 2 (schools) by the total explainable variance in the model [$\tau_{00} + \sigma^2$ *or* Var ($u_{0j} + r_{ij}$)]—the total variance is equal to the student-level variance (σ^2) plus the school-level variance (τ_{00}).

Results from the unconditional model fit to the English language acquisition data for the IPT pretest yielded a student-level variance of 0.76 and a school-level variance of 0.10, (table I1). Therefore, the estimated intraclass correlation between any two students in the same school was 0.12, slightly higher than the expected intraclass correlation coefficient (.10).

Table I1. Variance components and intraclass correlation for student language acquisition

Outcome	Estimate
Level 1 (student) variance	0.76
Level 2 (school) variance	0.10
Total variance	0.86
Intraclass correlation	0.12

Source: IPT data.

Unconditional models were also fit to the IPT section pretests (table I2) and intermediate outcomes (table I3) data.

Table I2. Variance components and intraclass correlation for listening, reading, and writing pretests

Outcome	Estimate
Listening	
Level 1 (student) variance	0.94
Level 2 (school) variance	0.11
Total variance	1.05
Intraclass correlation	0.10
Reading	
Level 1 (student) variance	0.94
Level 2 (school) variance	0.08
Total variance	1.02
Intraclass correlation	0.08
Writing	
Level 1 (student) variance	0.93
Level 2 (school) variance	0.14
Total variance	1.07
Intraclass correlation	0.13

Source: IPT data.

Table I3. Variance components and intraclass correlation for intermediate outcomes

Outcome	Estimate
Student engagement	
Level 1 (student) variance	6.69
Level 2 (school) variance	1.16
Total variance	7.85
Intraclass correlation	0.15
Differentiated instruction	
Level 1 (student) variance	2.21
Level 2 (school) variance	0.44
Total variance	2.65
Intraclass correlation	0.17
Sheltering instruction	
Level 1 (student) variance	3.40
Level 2 (school) variance	0.60
Total variance	4.00
Intraclass correlation	0.15
Level 1 (student) variance	2.16
Level 2 (school) variance	0.14
Total variance	2.30
Intraclass correlation	0.06
Reading instruction	
Level 1 (student) variance	2.40
Level 2 (school) variance	0.10
Total variance	2.50
Intraclass correlation	0.04
Writing instruction	
Level 1 (student) variance	2.71
Level 2 (school) variance	0.34
Total variance	3.05
Intraclass correlation	0.11
Level 1 (student) variance	2.72
Level 2 (school) variance	0.70
Total variance	3.42
Intraclass correlation	0.20
Student-centered instruction	
Level 1 (student) variance	1.84
Level 2 (school) variance	0.32
Total variance	2.16
Intraclass correlation	0.15
Assessment practices	
Level 1 (student) variance	2.22
Level 2 (school) variance	0.26
Total variance	2.48
Intraclass correlation	0.10

Source: IPT data; online teacher logs.

APPENDIX J: PRELIMINARY, IMPACT, AND SENSITIVITY ANALYSES

This appendix presents the results of various analyses of the pretest and posttest data (tables J1–J10) of the IDEA Proficiency Test (IPT), published in 2005 by Ballard & Tighe.

Table J1. Model 1, Baseline equivalence model, preliminary analysis of unadjusted IDEA Proficiency Test (IPT) pretest data

Parameter	Regression coefficient	Standard error	t ratio	*p*-value
Intercept	0.04	0.06	0.63	.53
Intervention	–0.11	0.08	–1.40	.17
Block 2	0.10	0.14	0.74	.46
Block 3	0.10	0.13	0.75	.46
Block 4	–0.34	0.25	–1.37	.18
Block 5	–0.28	0.14	–1.98	.05
Block 6	–0.11	0.16	–0.72	.48
Block 7	0.14	0.12	1.19	.24
Block 8	0.15	0.25	0.62	.54
Block 9	0.03	0.16	0.17	.87
Block 10	0.14	0.26	0.53	.60
Block 11	–0.42	0.18	–2.30	.03
Block 12	0.02	0.27	0.09	.93
Block 13	–0.01	0.14	–0.09	.93
Grade 3	1.03	0.04	26.28	<.0001
Grade 4	1.15	0.04	29.04	<.0001
Grade 5	1.63	0.04	38.43	<.0001

Note: IPT is a series of language proficiency assessments for English language learners in K–12.
Source: IPT data.

Table J2. Model 2, Confirmatory impact analysis model, adjusted IDEA Proficiency Test (IPT) posttest data

Parameter	Regression coefficient	Standard error	t ratio	*p*-value
Intercept	0.02	0.05	0.43	.67
Intervention*	–0.03	0.06	–0.44	.66
Percent White	–0.00	0.00	–0.15	.88
Percent Hispanic	–0.00	0.00	–0.14	.89
Block 2	0.15	0.10	1.40	.17
Block 3	0.16	0.10	1.72	.10
Block 4	0.06	0.19	0.31	.76
Block 5	0.16	0.12	1.41	.17
Block 6	0.07	0.13	0.52	.61
Block 7	0.15	0.09	1.78	.09
Block 8	–0.23	0.18	–1.31	.20
Block 9	0.03	0.12	0.26	.80
Block 10	0.13	0.20	0.68	.50
Block 11	0.31	0.16	1.91	.07
Block 12	0.33	0.21	1.52	.14
Block 13	–0.04	0.12	–0.34	.73
IPT pretest score	0.79	0.02	40.23	<.0001
Grade 3	–0.72	0.04	–19.10	<.0001
Grade 4	–0.43	0.04	–10.70	<.0001
Grade 5	–0.48	0.05	–9.75	<.0001

Note: IPT is a series of language proficiency assessments for English language learners in K–12.

Note: $R^2 = 0.52$.

* The confirmatory impact analysis model included three covariates—the IPT pretest, the percent of white students enrolled in school and the percent of Hispanic students enrolled in school.

Source: IPT data; online teacher logs.

Table J3. Model 3, Sensitivity analysis of adjusted IDEA Proficiency Test (IPT) posttest data with no imputation

Parameter	Regression coefficient	Standard error	t ratio	*p*-value
Intercept	0.03	0.04	0.82	.42
Intervention*	–0.06	0.05	–1.27	.21
Percent White	0.00	0.00	0.14	.89
Percent Hispanic	0.00	0.00	0.00	1.00
Block 2	0.24	0.08	2.79	.01
Block 3	0.27	0.08	3.52	.00
Block 4	0.09	0.15	0.60	.55
Block 5	0.20	0.10	2.08	.04
Block 6	0.17	0.11	1.59	.12
Block 7	0.24	0.07	3.25	.00
Block 8	–0.11	0.14	–0.78	.44
Block 9	0.14	0.11	1.22	.23
Block 10	0.24	0.16	1.52	.14
Block 11	0.32	0.15	2.12	.04
Block 12	0.44	0.19	2.33	.03
Block 13	0.04	0.10	0.35	.73
IPT pretest score	0.75	0.02	34.93	<.0001
Grade 3	–0.65	0.04	–15.46	<.0001
Grade 4	–0.38	0.04	–8.53	<.0001
Grade 5	–0.42	0.05	–7.96	<.0001

Note: IPT is a series of language proficiency assessments for English language learners in K–12.

*Missing data were case-deleted for this sensitivity analysis.

Source: IPT data; online teacher logs.

Table J4. Model 4, Sensitivity analysis of unadjusted IDEA Proficiency Test (IPT) posttest data with no race/ethnicity and pretest covariates

Parameter	Regression coefficient	Standard error	t ratio	*p*-value
Intercept	0.05	0.06	0.84	.41
Intervention	–0.10	0.07	–1.47	.15
Block 2	0.18	0.12	1.51	.14
Block 3	0.18	0.12	1.52	.14
Block 4	–0.22	0.23	–0.97	.34
Block 5	–0.10	0.13	–0.78	.44
Block 6	–0.07	0.15	–0.49	.62
Block 7	0.22	0.10	2.14	.04
Block 8	–0.20	0.22	–0.90	.37
Block 9	–0.00	0.15	–0.03	.98
Block 10	0.20	0.25	0.82	.42
Block 11	–0.13	0.15	–0.87	.39
Block 12	0.34	0.26	1.34	.19
Block 13	–0.10	0.13	–0.81	.43
Grade 3	0.07	0.04	1.70	.09
Grade 4	0.47	0.04	11.25	<.0001
Grade 5	0.81	0.04	18.45	<.0001

Note: IPT is a series of language proficiency assessments for English language learners in K–12.
Source: IPT data.

Table J5. Model 5, Sensitivity analysis of adjusted IDEA Proficiency Test (IPT) posttest data with no race/ethnicity covariate

Parameter	Regression coefficient	Standard error	t ratio	*p*-value
Intercept	0.02	0.04	0.49	.63
Intervention*	–0.02	0.05	–0.46	.65
Block 2	0.15	0.09	1.57	.13
Block 3	0.17	0.09	1.83	.08
Block 4	0.06	0.18	0.34	.74
Block 5	0.15	0.10	1.54	.14
Block 6	0.07	0.11	0.66	.52
Block 7	0.15	0.08	1.84	.08
Block 8	–0.23	0.17	–1.34	.19
Block 9	0.03	0.11	0.29	.77
Block 10	0.14	0.19	0.75	.46
Block 11	0.31	0.13	2.27	.04
Block 12	0.33	0.20	1.65	.11
Block 13	–0.05	0.10	–0.52	.61
IPT pretest score	0.79	0.02	40.28	<.0001
Grade 3	–0.72	0.04	–19.10	<.0001
Grade 4	–0.43	0.04	–10.70	<.0001
Grade 5	–0.48	0.05	–9.76	<.0001

Note: IPT is a series of language proficiency assessments for English language learners in K–12.

* The two significant race covariates "percent white" and "percent Hispanic" were not included in this model.

Source: IPT data.

Table J6. Model 6, Sensitivity analysis of adjusted IDEA Proficiency Test (IPT) posttest listening data

Parameter	Regression coefficient	Standard error	t ratio	*p*-value
Intercept	0.01	0.05	0.11	.91
Intervention	–0.02	0.06	–0.29	.77
Percent White	–0.00	0.00	–0.20	.85
Percent Hispanic	0.00	0.00	0.04	.97
Block 2	0.14	0.10	1.34	.19
Block 3	0.11	0.10	1.06	.30
Block 4	–0.19	0.20	–0.94	.36
Block 5	0.10	0.12	0.85	.40
Block 6	–0.02	0.14	–0.14	.89
Block 7	0.10	0.09	1.11	.28
Block 8	–0.28	0.18	–1.53	.14
Block 9	–0.01	0.14	–0.04	.97
Block 10	0.17	0.21	0.79	.44
Block 11	0.07	0.18	0.38	.70
Block 12	0.22	0.24	0.92	.37
Block 13	–0.02	0.13	–0.16	.87
IPT pretest score	0.52	0.03	20.31	<.0001
Grade 3	–0.14	0.05	–2.78	.01
Grade 4	0.16	0.05	3.02	.01
Grade 5	0.22	0.06	3.38	.00

Note: IPT is a series of language proficiency assessments for English language learners in K–12.
Source: IPT data; online teacher logs.

Table J7. Model 7, Sensitivity analysis of adjusted IDEA Proficiency Test (IPT) posttest reading data

Parameter	Regression coefficient	Standard error	t ratio	*p*-value
Intercept	0.07	0.06	1.11	.28
Intervention	–0.06	0.07	–0.95	.35
Percent White	–0.00	0.00	–0.21	.83
Percent Hispanic	0.00	0.00	0.08	.94
Block 2	0.02	0.12	0.19	.85
Block 3	0.12	0.11	1.08	.29
Block 4	–0.00	0.22	–0.02	.98
Block 5	–0.02	0.13	–0.18	.86
Block 6	–0.11	0.15	–0.74	.47
Block 7	0.11	0.10	1.08	.29
Block 8	–0.20	0.21	–0.96	.35
Block 9	0.02	0.15	0.16	.87
Block 10	0.14	0.23	0.62	.54
Block 11	0.24	0.19	1.23	.23
Block 12	0.45	0.26	1.76	.09
Block 13	–0.03	0.15	–0.21	.83
IPT pretest score	0.68	0.03	27.08	<.0001
Grade 3	–0.56	0.05	–10.89	<.0001
Grade 4	–0.34	0.06	–6.11	<.0001
Grade 5	–0.33	0.07	–4.94	<.0001

Note: IPT is a series of language proficiency assessments for English language learners in K–12.
Source: IPT data; online teacher logs.

Table J8. Model 8, Sensitivity analysis of adjusted IDEA Proficiency Test (IPT) posttest writing data

Parameter	Regression coefficient	Standard error	t ratio	*p*-value
Intercept	0.00	0.06	0.05	.96
Intervention	–0.03	0.07	–0.51	.61
Percent White	0.00	0.00	0.12	.90
Percent Hispanic	–0.00	0.00	0.00	1.00
Block 2	0.28	0.12	2.36	.03
Block 3	0.25	0.11	2.27	.03
Block 4	0.23	0.22	1.04	.31
Block 5	0.17	0.13	1.28	.21
Block 6	0.18	0.15	1.22	.24
Block 7	0.26	0.10	2.58	.02
Block 8	–0.28	0.20	–1.35	.19
Block 9	0.07	0.15	0.46	.65
Block 10	0.13	0.23	0.56	.58
Block 11	0.30	0.19	1.53	.14
Block 12	0.22	0.26	0.85	.40
Block 13	–0.16	0.14	–1.13	.27
IPT pretest score	0.66	0.02	31.56	<.0001
Grade 3	–1.01	0.05	–20.63	<.0001
Grade 4	–0.55	0.05	–11.13	<.0001
Grade 5	–0.50	0.06	–8.71	<.0001

Note: IPT is a series of language proficiency assessments for English language learners in K–12.
Source: IPT data; online teacher logs.

Table J9. Model 9, Sensitivity analysis of adjusted IDEA Proficiency Test (IPT) posttest data using eligibility for free or reduced-price lunch as an additional covariate

Parameter	Regression coefficient	Standard error	t ratio	*p*-value
Intercept	–0.02	0.05	–0.45	.66
Intervention	–0.02	0.05	–0.33	.75
Percent White	–0.00	0.00	–0.63	.53
Percent Hispanic	0.00	0.00	0.81	.42
Percent eligible for free or reduced–price lunch	0.01	0.00	2.31	.03
Block 2	0.16	0.10	1.65	.11
Block 3	0.34	0.12	2.90	.01
Block 4	0.07	0.17	0.39	.70
Block 5	0.14	0.11	1.27	.22
Block 6	0.00	0.12	0.01	.99
Block 7	0.09	0.08	1.07	.30
Block 8	–0.23	0.16	–1.41	.17
Block 9	0.01	0.12	0.06	.96
Block 10	0.06	0.18	0.34	.74
Block 11	0.34	0.15	2.17	.04
Block 12	0.28	0.20	1.40	.17
Block 13	–0.27	0.15	–1.80	.08
IPT pretest score	0.79	0.02	40.17	<.0001
Grade 3	–0.72	0.04	–19.08	<.0001
Grade 4	–0.43	0.04	–10.64	<.0001
Grade 5	–0.48	0.05	–9.70	<.0001

Note: IPT is a series of language proficiency assessments for English language learners in K–12.
Source: IPT data; online teacher logs.

Table J10. Model 10, Sensitivity analysis of adjusted IDEA Proficiency Test (IPT) posttest data excluding schools with low enrollments of English language learner students

Parameter	Regression coefficient	Standard error	t ratio	*p*-value
Intercept	0.04	0.05	0.83	0.42
Intervention*	–0.01	0.05	–0.25	0.81
Percent White	–0.00	0.00	–0.03	0.98
Percent Hispanic	0.00	0.00	0.98	0.34
Block 2	0.22	0.10	2.27	0.05
Block 3	0.30	0.08	3.85	0.00
Block 4	0.15	0.14	1.08	0.29
Block 5	0.12	0.09	1.38	0.19
Block 6	0.08	0.13	0.61	0.55
Block 7	0.13	0.06	2.02	0.06
Block 8	–0.26	0.13	–1.95	0.07
Block 9	0.32	0.14	2.31	0.03
Block 10	0.08	0.15	0.50	0.62
Block 11	0.20	0.16	1.25	0.24
Block 12	0.23	0.17	1.36	0.19
Block 13	–0.14	0.10	–1.38	0.18
IPT pretest score	0.80	0.02	36.79	<.0001
Grade 3	–0.73	0.04	–18.00	<.0001
Grade 4	–0.44	0.04	–9.79	<.0001
Grade 5	–0.49	0.05	–9.09	<.0001

Note: IPT is a series of language proficiency assessments for English language learners in K–12.

*In this model, students from those schools that have fewer than 20% ELL students were excluded from the analysis sample.

Source: IPT data; online teacher logs.

APPENDIX K: EXPLORATORY ANALYSES

This appendix presents the results of the exploratory analyses of the effect of On Our Way to English (OWE) (by David Freeman, Yvonne Freeman, Aurora Colon Garcia, Margo Gottlieb, Mary Lou McCloskey, Lydia Stack, and Cecilia Silva and published in 2003 by Rigby), and Responsive Instruction for Success in English (RISE) (written by Clara Amador-Watson and published in 2004 by Harcourt Achieve) on student engagement with ELL-specific educational materials (table K1). This appendix also contains results of teachers' instructional practices (differentiated instruction [table K2], sheltering instruction [table K3], receptive and expressive language instruction [table K4], reading instruction [table K5], and writing instruction [table K6]); and teachers' instructional responsiveness and assessment practices (modification of instruction or teacher responsiveness [table K7], student-centered instruction [table K8], and assessment use [table K9]).

Table K1. Model 11, Exploratory analysis, adjusted results of school-level outcomes on student engagement

Parameter	Regression coefficient	Standard error	t ratio	*p*-value
Intercept	3.28	0.98	3.35	.00
Intervention	–0.79	0.71	–1.11	.27
Block 3	1.27	1.22	1.04	.30
Block 4	–0.56	1.79	–0.31	.76
Block 5	0.91	1.44	0.63	.53
Block 6	–4.91	1.66	–2.96	.01
Block 7	–0.45	1.13	–0.40	.69
Block 8	0.49	2.37	0.21	.84
Block 9	1.58	1.46	1.08	.29
Block 10	–0.40	1.87	–0.21	.83
Block 11	–0.59	1.26	–0.47	.64
Block 12	–0.09	2.96	–0.03	.98
Block 13	0.45	1.49	0.30	.76
Masters/Ph.D.	0.66	0.53	1.24	.22
Minimal Spanish	–0.21	0.80	–0.27	.79
Intermediate Spanish	0.90	0.92	0.97	.33
Advanced/fluent Spanish	0.02	0.99	0.02	.98
English as a second language certificate	1.08	0.61	1.77	.08

Note: IPT is a series of language proficiency assessments for English language learners in K–12.
Source: IPT data; online teacher logs.

Table K2. Model 12, Exploratory analysis, adjusted results of school-level outcomes on differentiated instruction

Parameter	Regression coefficient	Standard error	t ratio	*p*-value
Intercept	2.84	0.52	5.42	<.0001
Intervention	0.29	0.32	0.89	.38
Block 3	0.41	0.58	0.70	.49
Block 4	–0.76	0.72	–1.07	.29
Block 5	0.16	0.72	0.23	.82
Block 6	0.30	0.85	0.35	.73
Block 7	0.71	0.56	1.27	.21
Block 8	0.07	1.16	0.06	.95
Block 9	1.54	0.68	2.25	.03
Block 10	–0.63	0.75	–0.83	.41
Block 11	0.37	0.61	0.61	.55
Block 12	–1.66	1.58	–1.05	.30
Block 13	0.48	0.79	0.61	.55
Masters/Ph.D.	–0.00	0.30	–0.01	1.00
Minimal Spanish	0.18	0.46	0.40	.69
Intermediate Spanish	–0.00	0.53	0.00	1.00
Advanced/fluent Spanish	0.10	0.57	0.18	.86
English as a second language certificate	0.97	0.34	2.90	.00

Note: IPT is a series of language proficiency assessments for English language learners in K–12.
Source: IPT data; online teacher logs.

Table K3. Model 13, Exploratory analysis, adjusted results of school-level outcomes on sheltering instruction

Parameter	Regression coefficient	Standard error	t ratio	*p*-value
Intercept	3.10	0.39	7.98	<.0001
Intervention	0.43	0.24	1.80	.08
Block 3	0.25	0.43	0.58	.57
Block 4	–0.97	0.53	–1.82	.08
Block 5	–0.25	0.53	–0.46	.65
Block 6	0.40	0.63	0.63	.53
Block 7	0.40	0.41	0.96	.35
Block 8	–0.48	0.86	–0.56	.58
Block 9	0.13	0.51	0.25	.80
Block 10	0.09	0.56	0.15	.88
Block 11	0.16	0.45	0.35	.73
Block 12	–2.22	1.17	–1.89	.07
Block 13	0.13	0.58	0.22	.83
Masters/Ph.D.	–0.24	0.22	–1.05	.30
Minimal Spanish	0.23	0.34	0.66	.51
Intermediate Spanish	0.06	0.40	0.16	.87
Advanced/fluent Spanish	–0.32	0.42	–0.75	.46
English as a second language certificate	0.46	0.25	1.83	.07

Note: IPT is a series of language proficiency assessments for English language learners in K–12.
Source: IPT data; online teacher logs.

Table K4. Model 14, Exploratory analysis, adjusted results of school-level outcomes on receptive and expressive language instruction

Parameter	Regression coefficient	Standard error	t ratio	*p*-value
Intercept	3.27	0.53	6.16	<.0001
Intervention	–0.11	0.33	–0.33	.74
Block 3	0.14	0.60	0.23	.82
Block 4	–0.58	0.75	–0.76	.45
Block 5	–0.15	0.73	–0.21	.84
Block 6	–0.36	0.86	–0.42	.68
Block 7	0.27	0.57	0.47	.64
Block 8	–0.33	1.19	–0.28	.78
Block 9	–0.05	0.70	–0.07	.95
Block 10	–0.76	0.79	–0.95	.35
Block 11	–0.14	0.62	–0.23	.82
Block 12	–1.98	1.60	–1.24	.22
Block 13	–1.07	0.80	–1.35	.19
Masters/Ph.D.	–0.05	0.30	–0.17	.86
Minimal Spanish	0.13	0.46	0.29	.77
Intermediate Spanish	0.06	0.54	0.12	.90
Advanced/fluent Spanish	–0.21	0.57	–0.37	.72
English as a second language certificate	0.45	0.34	1.31	.19

Note: IPT is a series of language proficiency assessments for English language learners in K–12.
Source: IPT data; online teacher logs.

Table K5. Model 15, Exploratory analysis, adjusted results of school-level outcomes on reading instruction

Parameter	Regression coefficient	Standard error	t ratio	*p*-value
Intercept	3.66	0.60	6.07	<.0001
Intervention	0.10	0.42	0.24	.81
Block 3	1.02	0.77	1.33	.20
Block 4	–0.28	0.99	–0.28	.78
Block 5	1.22	0.91	1.33	.19
Block 6	0.17	1.03	0.17	.87
Block 7	1.11	0.74	1.50	.14
Block 8	2.03	1.41	1.44	.16
Block 9	0.52	0.89	0.59	.56
Block 10	0.11	1.05	0.10	.92
Block 11	0.30	0.79	0.38	.71
Block 12	–1.46	1.81	–0.81	.43
Block 13	–0.51	0.95	–0.54	.60
Masters/Ph.D.	0.18	0.35	0.51	.61
Minimal Spanish	–0.45	0.51	–0.89	.37
Intermediate Spanish	–0.59	0.59	–1.00	.32
Advanced/fluent Spanish	–1.29	0.63	–2.03	.05
English as a second language certificate	0.24	0.39	0.61	.54

Note: IPT is a series of language proficiency assessments for English language learners in K–12.
Source: IPT data; online teacher logs.

Table K6. Model 16, Exploratory analysis, adjusted results of school-level outcomes on writing instruction

Parameter	Regression coefficient	Standard error	t ratio	*p*-value
Intercept	3.36	0.61	5.47	<.0001
Intervention	0.17	0.37	0.47	.64
Block 3	2.05	0.66	3.09	.00
Block 4	1.65	0.78	2.10	.04
Block 5	1.92	0.90	2.13	.04
Block 6	0.21	1.01	0.21	.84
Block 7	2.06	0.61	3.36	.00
Block 8	3.24	1.63	1.98	.06
Block 9	2.32	0.74	3.14	.00
Block 10	–0.15	0.94	–0.16	.87
Block 11	–0.20	0.66	–0.30	.77
Block 12	–3.04	1.70	–1.79	.08
Block 13	–0.07	0.85	–0.09	.93
Masters/Ph.D.	0.45	0.35	1.30	.20
Minimal Spanish	0.26	0.57	0.45	.65
Intermediate Spanish	0.02	0.65	0.03	.97
Advanced/fluent Spanish	0.18	0.73	0.24	.81
English as a second language certificate	–0.17	0.41	–0.42	.68

Note: IPT is a series of language proficiency assessments for English language learners in K–12.
Source: IPT data; online teacher logs.

Table K7. Model 17, Exploratory analysis, adjusted results of school-level outcomes on modification of instruction/teacher responsiveness

Parameter	Regression coefficient	Standard error	t ratio	*p*-value
Intercept	4.47	0.63	7.13	<.0001
Intervention	0.31	0.41	0.75	.46
Block 3	1.10	0.73	1.52	.14
Block 4	0.36	0.96	0.38	.71
Block 5	0.13	0.88	0.15	.88
Block 6	1.97	1.03	1.92	.06
Block 7	1.13	0.68	1.65	.11
Block 8	0.26	1.42	0.18	.86
Block 9	1.99	0.86	2.31	.03
Block 10	0.79	1.01	0.78	.44
Block 11	0.94	0.75	1.25	.22
Block 12	–0.85	1.88	–0.45	.66
Block 13	0.41	0.94	0.44	.67
Masters/Ph.D.	–0.04	0.35	–0.11	.91
Minimal Spanish	0.11	0.54	0.20	.84
Intermediate Spanish	0.08	0.62	0.13	.90
Advanced/fluent Spanish	–0.21	0.66	–0.32	.75
English as a second language certificate	0.94	0.40	2.34	.02

Note: IPT is a series of language proficiency assessments for English language learners in K–12.
Source: IPT data; online teacher logs.

Table K8. Model 18, Exploratory analysis, adjusted results of school-level outcomes on student-centered instruction

Parameter	Regression coefficient	Standard error	t ratio	*p*-value
Intercept	1.73	0.48	3.61	.001
Intervention	0.36	0.29	1.22	.23
Block 3	0.37	0.53	0.70	.49
Block 4	–0.94	0.65	–1.44	.16
Block 5	0.15	0.66	0.23	.82
Block 6	0.65	0.78	0.84	.41
Block 7	0.14	0.51	0.28	.78
Block 8	0.22	1.06	0.21	.84
Block 9	0.11	0.63	0.18	.86
Block 10	0.15	0.69	0.21	.83
Block 11	–0.36	0.55	–0.64	.53
Block 12	–2.58	1.45	–1.78	.08
Block 13	–0.44	0.72	–0.61	.55
Masters/Ph.D.	–0.07	0.28	–0.26	.80
Minimal Spanish	0.16	0.42	0.38	.71
Intermediate Spanish	0.80	0.49	1.63	.11
Advanced/fluent Spanish	0.92	0.52	1.77	.08
English as a second language certificate	0.54	0.31	1.76	.08

Note: IPT is a series of language proficiency assessments for English language learners in K–12.
Source: IPT data; online teacher logs.

Table K9. Model 19, Exploratory analysis, adjusted results of school-level outcomes on assessment practices

Parameter	Regression coefficient	Standard error	t ratio	*p*-value
Intercept	2.93	0.54	5.43	<.0001
Intervention	–0.07	0.34	–0.22	.83
Block 3	1.11	0.61	1.82	.08
Block 4	1.51	0.77	1.94	.06
Block 5	0.09	0.75	0.12	.91
Block 6	0.29	0.88	0.33	.74
Block 7	1.00	0.58	1.72	.09
Block 8	1.34	1.21	1.11	.28
Block 9	0.84	0.72	1.16	.25
Block 10	–0.67	0.82	–0.82	.42
Block 11	0.68	0.63	1.08	.29
Block 12	–2.38	1.62	–1.47	.15
Block 13	–0.70	0.81	–0.87	.39
Masters/Ph.D.	0.40	0.31	1.30	.20
Minimal Spanish	–0.02	0.47	–0.04	.97
Intermediate Spanish	–0.12	0.54	–0.22	.82
Advanced/fluent Spanish	–0.12	0.58	–0.21	.83
English as a second language certificate	0.14	0.35	0.39	.70

Note: IPT is a series of language proficiency assessments for English language learners in K–12.
Source: IPT data; online teacher logs.

REFERENCES

Abedi, J. (Ed.). (2007). *English language proficiency assessment in the nation: Current status and future practice*. Davis, CA: University of California, School of Education.

Amador-Watson, C. (2007). *Responsive Instruction for Success in English kit*. Austin, TX: Harcourt Achieve.

Bachman, F. L. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing, 17*(1), 1–42.

Ballard & Tighe Publishers. (2001). *Overview and advantages of IPT Oral Tests*. Retrieved September 16, 2005, from http://www.ballard- tighe.com /IPTOnlineInserviceTraining/IPTOral/IPTOral_Overview and Advantages.htm

Ballard & Tighe Publishers. (2005a). *IPT 1–2 technical manual: Listening, reading, writing, speaking*. Brea, CA: Ballard & Tighe.

Ballard & Tighe Publishers. (2005b). *IPT 3–5 technical manual: Listening, reading, writing, speaking*. Brea, CA: Ballard & Tighe.

Ballard & Tighe Publishers. (2009). *IPT test description.* Retrieved October 16, 2009, from http://www.nclb.ballard-tighe.com/system/description.html

Bloom, H. S. (2005). Randomizing groups to evaluate place-based programs. In H. S. Bloom (Ed.), *Learning more from social experiments: Evolving analytic approaches* (115–172). New York: Russell Sage Foundation.

Center on Education Policy. (2006). *From the capital to the classroom: Year 4 of the No Child Left Behind Act.* Washington, DC: Center on Education Policy.

Chapman, E. (2003). Alternative approaches to assessing student engagement rates. *Practical Assessment, Research & Evaluation, 8*(13), 1–10.

Darder, A. (1993). How does the culture of the teacher shape the classroom experience of Latino students? The unexamined question in critical pedagogy. In S. Rothstein (Ed.), *Handbook of schooling in urban America* (pp. 195–222). Westport, CT: Greenwood Press.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures [Electronic Version]. *Educational Researcher, 38,* 181–199.

Dobson, J., and Henthorne, K. (1999). *Pupil mobility in schools: Interim report*. London: DfEE.

Dweck, C. S., and Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95,* 256–273.

Editorial Projects in Education Research Center. (2009). *Perspectives on a population: English-language learners in American schools*. Bethesda, MD: Editorial Projects in Education Research Center.

Educational Research Institute of America. (2004, June). *A study of the instructional effectiveness of On Our Way to English*. Brewster, NY: Educational Research Institute of America.

Elementary and Secondary Education Act (ESEA). (1965). Pub.L. 89-10, 79 Stat. 27, 20 U.S.C. ch.70.

Enders, C. K., and Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121–138.

Freeman, D., and Freeman, Y. (2000). Meeting the needs of English language learners. *Talking Points, 12*(1), 2–7.

Freeman, D., Freeman, Y., Garcia, A. C., Gottlieb, M., McCloskey, M. L., Stack, L., and Silva, C. (2003). *On Our Way to English, K–5*. Barrington, IL: Rigby.

Freeman, D., Freeman, Y., Garcia, A. C., Gottlieb, M., McCloskey, M. L., Stack, L., and Silva, C. (2004). *On Our Way to English: Guided reading teacher's guide (grade 1).* Barrington, IL: Rigby.

Fry, R. T. (2008). The role of schools in the English language learner achievement gap. Washington, DC: Pew Hispanic Research Center.

Gersten, R., Baker, S. K., Haager, D., and Graves, A. W. (2005). Exploring the role of teacher quality in predicting reading outcomes for first-grade English learners. *Remedial and Special Education, 26*(4), 197–206.

Glass, G. V., McGaw, B., and Smith, M. L. (1981). *Meta-analysis in social research*. Thousand Oaks, CA: Sage.

Godina, H. (1998). Mexican-American high school students and the role of literacy across home-school-community settings. *Dissertation Abstracts International, 59*(8), 2825.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60,* 549–576

Grissom, R. J., and Kim, J. J. (2005). *Effect sizes for research: A broad approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Haager, D., Gersten, R., Baker, S., and Graves, A. W. (2003). The English-language learner classroom observation instrument for beginning readers. In S. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: Systems for the observation of teaching and learning* (pp. 111–144). Baltimore, MD: Brookes.

Harcourt Achieve. (2004a). *On Our Way to English.* Austin, TX: Rigby-Harcourt Achieve.

Harcourt Achieve. (2004b). *On Our Way to English: Research base*. Retrieved September 8, 2005, from http://rigby.harcourtachieve.com/HA/correlations/pdf/o/*OWE*_RB.pdf

Harcourt Achieve. (2005). *The instructional effectiveness of On Our Way to English on English immersion and bilingual students.* Retrieved September 8, 2005, from http://rigby.hmhco.com/HA/correlations/pdf/o/OWE_Effectiveness_research.pdf

Herman, J. L., and Abedi, J. (2004, April). *Issues in assessing English language learners' opportunity to learn mathematics*. Paper presented at the meeting of the American Educational Research Association, San Diego, CA.

Hill, J. D., and Flynn, K. (2004). *English language learner resource guide: A guide for rural districts with a low incidence of ELLs.* Aurora, CO: Mid-continent Research for Education and Learning.

IPT: The New Title II Testing System. (2004). Brea, CA: Ballard & Tighe.

IPT: Testing System. (2005). Retrieved September 13, 2005 from http://nclb.ballard-tighe.com/main.html

Knox, C., and Amador-Watson, C. (2002). *RISE Facilitators Resource Notebook*. Barrington, IL: Rigby.

Krashen, S. (1996). *Under attack: The case against bilingual education*. Culver City, CA: Language Education Associates.

LAS Links K–12 Assessments. (n.d.). Retrieved from http://www.ctb.com/laslinks

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12*(3), 251–285.

Marketing Works. (2005). The instructional effectiveness of On Our Way to English on English immersion and bilingual students. Yardley, PA: Marketing Works.

National Council of Teachers of English. (2006). *NCTE position paper on the role of English teachers in educating English language learners (ELLs).* Retrieved January 20, 2007, from http://www.ncte.org/about/over/positions/category/div/124545.htm

No Child Left Behind Act of 2001. (2002). Pub. L. No. 107–110, 115 Stat. 1425.

Patkowski, M. (1982). The sensitive period for the acquisition of syntax in a second language. In S. Krashen, R. Scarcell, & M. Long (Eds.), *Issues in second language research* (pp. 52–63). London: Newbury House.

Penuel, W. R., Fishman, B. J., Yamaguchi, R., and Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal, 44*(4), 921–958.

Planty, M., Hussar, W., Snyder, T., Kena, G., Kewal Ramani, A., Kemp, J., Bianco, K., and Dinkes, R. (2009). *The condition of education 2009* (NCES Publication No. 2009-081). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Educational Statistics.

Puma, M. J., Olsen, R. B., Bell, S. H., and Price, C. (2009). *What to do when data are missing in group randomized controlled trials* (NCEE 2009-0049). Washington, DC: National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Raudenbush, S., and Bryk, A. (2002). Hierarchical linear models: Applications and data analysis methods. Thousand Oaks, CA: Sage.

Raudenbush, S. W., Spybrook, J., Liu, X., and Congdon, R. (2005). *Optimal design for longitudinal and multilevel research: Documentation for the "Optimal Design" software*. Retrieved January 12, 2006, from http://sitemaker.umich.edu/group-based

Renaissance Learning. (2003). *STAR Reading - Overview*. Retrieved July 2, 2005, from http://www.renlearn.com/starreading

Rumberger, R., and Gandara, P. (2004). Seeking equity in the education of California's English learners. *Teachers College Record, 106*(10), 2031–2055.

Schafer, J. L. (1999) Multiple imputation: A primer. *Statistical Methods in Medical Research, 8,* 3–15.

Schochet, P. (2008). *Technical methods report: Guidelines for multiple testing in impact evaluations* (NCEE 2008-4018). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Schrank, F. A., Fletcher, T. V., & Alvarado, C. G. (1996). Comparative validity of three English oral language proficiency tests. *The Bilingual Research Journal,* 20(1), 55–68

Singleton, D., and Ryan, L. (2004). *Language acquisition: The age factor* (2nd ed.). Clevedon, UK: Multilingual Matters.

Slavin, R. E., and Cheung, A. (2003). *Effective reading programs for English language learners* (Report No. 66). Baltimore, MD: Center for Research on the Education of Students Placed at Risk.

Smith, F. (1983). Essays into literacy: Selected papers and some afterthoughts. Portsmouth, NH: Heinemann.

Téllez, K., and Waxman, H. (2005). *Quality teachers for English language learners.* Philadelphia, PA: Mid-Atlantic Regional Educational Laboratory.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2002). 1999–2000 Schools and staffing survey: Overview of the data for public, private, public charter and Bureau of Indian Affairs elementary and secondary schools. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2007). *The condition of education 2007* (NCES Publication No. 2007-064). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2010a). *Mathematics 2009: National assessment of educational progress at grades 4 and 8* (NCES Publication No. 2010-451). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2010b). *Reading 2009: National assessment of educational progress at grades 4 and 8* (NCES Publication No. 2010-458). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Common Core of Data. (2005–2006). *Build a Table* online database. Retrieved July 10, 2009, from http://nces.ed.gov/ccd/bat/

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Common Core of Data. (2007–2008). *Build a Table* online database. Retrieved February 22, 2010, from http://nces.ed.gov/ccd/bat/

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Common Core of Data. (2008–2009). *Build a Table* online database. Retrieved June 6, 2010, from http://nces.ed.gov/ccd/bat/.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2008). *WWC study review standards - revised*. Retrieved January 10, 2009, from http://whatworks.ed.gov/reviewprocess/study_standards_final.pdf

U.S. Department of Education, Office of English Language Acquisition, Language Enhancement, and Academic Achievement for Limited English Proficient Students. (2010a). *Colorado: Rate of ELL Growth 1997/1998-2007/2008.* Retrieved from: http://www.ncela.gwu.edu/files/uploads/20/Colorado_G_0708.pdf.

U.S. Department of Education, Office of English Language Acquisition, Language Enhancement, and Academic Achievement for Limited English Proficient Students. (2010b). *Kansas: Rate of ELL Growth 1997/1998-2007/2008.* Retrieved from: http://www.ncela.gwu.edu/files/uploads/20/Kansas_G_0708.pdf.

U.S. Department of Education, Office of English Language Acquisition, Language Enhancement, and Academic Achievement for Limited English Proficient Students. (2010c). *Missouri: Rate of ELL Growth 1997/1998-2007/2008.* Retrieved from: http://www.ncela.gwu.edu/files/uploads/20/Missouri_G_0708.pdf.

U.S. Department of Education, Office of English Language Acquisition, Language Enhancement, and Academic Achievement for Limited English Proficient Students. (2010d). *Nebraska: Rate of ELL Growth 1997/1998-2007/2008.* Retrieved from: http://www.ncela.gwu.edu/files/uploads/20/Nebraska_G_0708.pdf.


DEPARTMENT OF EDUCATION
UNITED STATES OF AMERICA