Making an Impact

The relative effectiveness of two approaches to early literacy intervention in grades K–2

Barbara Foorman
Sarah Herrera
Jennifer Dombek
Chris Schatschneider
Yaacov Petscher
Florida Center for Reading Research at Florida State University

Key findings

This randomized controlled trial in 55 low-performing schools across Florida compared two pull-out early literacy interventions—one using standalone materials and one using materials embedded in the existing core reading program. The interventions were delivered daily for 45 minutes for 27 weeks in small groups of students at risk of literacy failure in 2013/14 and 2014/15. The standalone intervention significantly improved grade 2 spelling outcomes relative to the embedded intervention, but impacts on other student outcomes were similar for the two interventions. On average, students in schools that used the standalone intervention and students in schools that used the embedded intervention showed similar improvement in reading and language outcomes. The two interventions also had similar impacts on reading and language outcomes among English learner students and non–English learner students, except for some reading outcomes in kindergarten.

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S. Department of Education

REL SOUTHEAST
Regional Educational Laboratory
At Florida State University

U.S. Department of Education
Betsy DeVos, *Secretary*

Institute of Education Sciences
Thomas W. Brock, *Commissioner for Education Research*
Delegated the Duties of Director

National Center for Education Evaluation and Regional Assistance
Audrey Pendleton, *Acting Commissioner*
Elizabeth Eisner, *Acting Associate Commissioner*
Amy Johnson, *Action Editor*
Sandra Garcia, *Project Officer*

REL 2017–251

Foorman, B., Herrera, S., Dombek, J., Schatschneider, C., & Petscher, Y. (2017). *The relative effectiveness of two approaches to early literacy intervention in grades K–2* (REL 2017–251). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs.

This report is available on the Regional Educational Laboratory website at http://ies.ed.gov/ncee/edlabs.

www.manaraa.com

Summary

Understanding written language is crucial to academic success in all content areas. Ensuring a strong foundation in the components of written language—that is, the literacy skills of reading, writing, and oral language (Mehta, Foorman, Branum-Martin, & Taylor, 2005) —is essential if students are to read with understanding and, thus, is a primary goal of early literacy instruction and of the Regional Educational Laboratory Southeast Improving Literacy Research Alliance. When students fall behind in developing literacy skills, early literacy intervention in kindergarten through grade 2 can reduce the number of students failing to attain grade-level expectations (Foorman & Al Otaiba, 2009; Foorman, Breier, & Fletcher, 2003; Foorman & Torgesen, 2001).

There is a strong research base on the skills targeted by effective early literacy intervention (Foorman, Beyler, et al., 2016). Effective early literacy instruction includes explicit instruction in phonological awareness, links from letters to sounds, decoding, and word study, as well as practice reading for accuracy, fluency, and comprehension (Foorman, Beyler, et al., 2016; Foorman & Connor, 2011; National Institute of Child Health and Human Development, 2000; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001; Snow, Burns, & Griffin, 1998). These skills are often delivered in multiple tiers of instruction that include the classroom at tier 1, supplemental, small-group intervention at tier 2, and intensive intervention at tier 3 for students who do not progress after a reasonable amount of time with tier 2 intervention (Gersten et al., 2009).

Furthermore, research has demonstrated the efficacy of directly teaching academic vocabulary and language to students to improve their comprehension (Baker et al., 2014; Foorman, Beyler, et al., 2016). In grades K–2 this includes the oral language skills of listening comprehension, syntax, and vocabulary that predict comprehension outcomes, along with reading skills (Foorman, Herrera, Petscher, Mitchell, & Truckenmiller, 2015).

An important consideration for schools and this study is to determine which instructional materials to use in tier 2 early literacy intervention. One approach is to use the tier 2 materials embedded in the existing core reading program selected for classroom instruction, which is appealing because these materials are aligned with core classroom instruction and do not require the purchase of additional materials. But even though these embedded tier 2 materials may claim to be research-based, they are rarely evaluated empirically. Another approach is to select tier 2 standalone instructional materials and strategies outside the existing core reading program. If the standalone materials are backed by strong evidence that they support learning in reading and language, it is reasonable to expect that the standalone approach will lead to better outcomes for small-group tier 2 intervention than will an embedded approach that has not been empirically evaluated.

Regional Educational Laboratory Southeast sought to explore whether providing at-risk students with small-group tier 2 intervention using a standalone intervention leads to better reading and language outcomes than does using an embedded intervention. To address this question, 55 low-performing schools, as identified by the state's school grading system, in south, central, and north Florida were randomly assigned to implement a pull-out standalone or embedded tier 2 intervention for 45 minutes daily throughout the school year. In each school the intervention was used in groups of four students in grades K–1

and five students in grade 2. All students were among those identified as being at risk of literacy failure.

Key findings include:
- Students at risk of literacy failure in grades K–2 improved, on average, 13–25 percentile points on reading outcomes and 6–25 percentile points on language outcomes, in both standalone and embedded intervention schools.
- The standalone intervention did not significantly improve reading or language outcomes relative to the embedded intervention among students in grades K–2, except for spelling in grade 2. The standalone intervention led to significantly better grade 2 spelling outcomes than did the embedded intervention.
- The two interventions had similar impacts on reading and language outcomes in grades K–2 for groups of students who differed on baseline performance and for schools from the 2013/14 and 2014/15 cohorts, except for spelling in grade 2. Again, the standalone intervention led to significantly better grade 2 spelling outcomes among students with low baseline spelling scores than did the embedded intervention.
- The two interventions had similar impacts on reading and language outcomes among English learner students and non–English learner students in grades K–2, except for some reading outcomes in kindergarten.
- In kindergarten, English learner students in embedded intervention schools performed better in phonological awareness than did non–English learner students, but non–English learner students in standalone intervention schools performed better in word reading than did English learner students. In embedded intervention schools, non–English learner students performed better in word reading in kindergarten than did English learner students.

Contents

Tables

Why this study?

Understanding written language is crucial to academic success in all content areas. Ensuring a strong foundation in the components of written language—that is, the literacy skills of reading, writing, and oral language (Mehta et al., 2005)—is essential if students are to read with understanding and, thus, is a primary goal of early literacy instruction and of the Regional Educational Laboratory (REL) Southeast Improving Literacy Research Alliance. When students fall behind in developing literacy skills, early literacy intervention can reduce the number of students failing to attain grade-level expectations (Foorman & Al Otaiba, 2009; Foorman et al., 2003; Foorman & Torgesen, 2001).

Skills targeted in effective early literacy intervention

There is a strong research base on the skills targeted by effective early literacy intervention (Foorman, Beyler, et al., 2016; see box 1 for definitions of key terms). Effective early literacy intervention includes explicit instruction in phonological awareness, links from letters to sounds, decoding, and word study as well as practice reading text for accuracy, fluency, and comprehension (Foorman, Beyler, et al., 2016; Foorman & Connor, 2011; National Institute of Child Health and Human Development, 2000; Rayner et al., 2001; Snow et al., 1998). These skills are often delivered in multiple tiers of instruction that include the classroom at tier 1, supplementary, small-group intervention at tier 2, and intensive intervention at tier 3 for students who do not progress after a reasonable amount of time with tier 2 intervention (Gersten et al., 2009). Although the effectiveness of multiple tiers of intervention was questioned in a national evaluation in which students just above a school's cutscore were compared with students just below (Balu et al., 2015), a recent systematic review of the research on tier 2 interventions in the primary grades from 2002 to 2014 revealed that 23 studies met rigorous design standards and had impacts in all areas of reading but primarily in word and pseudoword reading (Gersten, Newman-Gonchar, Haymond, & Dimino, in press). These tier 2 interventions were administered individually and in small groups by adults who had high levels of ongoing support (Gersten et al., in press).

To improve comprehension of content area text, students must also learn the vocabulary and discourse elements—the academic language—of the texts. Research is increasingly demonstrating the efficacy of directly teaching academic language to students in order to improve their comprehension (Baker et al., 2014; Foorman, Beyler, et al., 2016). Specifically, in grades K–2, this includes the oral language skills of listening comprehension, syntax, and vocabulary that predict comprehension outcomes, along with reading skills (Foorman, Herrera, et al., 2015). Thus, early literacy interventions that aim to improve comprehension must include instruction in both reading and language skills.

Two approaches to choosing materials and strategies for early literacy intervention

A priority for the REL Southeast Improving Literacy Research Alliance was to find effective tier 2 early literacy interventions for at-risk students in grades K–2. This priority was especially pressing for the alliance members in Florida because of the state's grade 3 retention law and strict teacher evaluation system. In fact, several of these alliance members were instrumental in gaining approval for this study to be conducted in their districts (Foorman, Dombek, & Smith, 2016). In addition to evidence of effectiveness, alliance

Ensuring a strong foundation in the components of written language is essential if students are to read with understanding and, thus, is a primary goal of early literacy instruction and of the Regional Educational Laboratory Southeast Improving Literacy Research Alliance

Box 1. Key terms

Reading and language baseline measures. The study included reading and language baseline measures that were collected in September, prior to the implementation of the interventions, from the Florida Center for Reading Research Reading Assessment (FRA). Reading baseline measures were the Letter Sounds (kindergarten only), Phonological Awareness (kindergarten only), Word Reading (grades 1 and 2), and Spelling (grade 2 only) subtests from the FRA. Language baseline measures were the Vocabulary Pairs, Following Directions, and Sentence Comprehension (kindergarten and grade 1) subtests from the FRA.

Differences in reading and language outcomes. Differences in reading and language outcomes between the standalone and embedded interventions are reported in one of three ways: statistically significant difference, substantively important difference, or no difference.

- *Statistically significant difference* in an outcome between the interventions was defined as a probability of less than 5 percent that the observed difference occurred by chance.
- *Substantively important difference* in an outcome between the interventions was identified using the What Works Clearinghouse criterion: a Hedges's *g* effect size of 0.25 or greater. When a substantively important but not statistically significant effect of one intervention relative to the other was observed, the outcome of one intervention is described as either higher or lower than the outcome of the other intervention.
- *No difference* in an outcome is identified when the difference in an outcome between the interventions is neither statistically significant nor substantively important.

Early literacy intervention. Early literacy intervention is defined as tier 2 pull-out, small-group, targeted intervention that includes explicit instruction in reading and language skills.

Effect size. An effect size describes the magnitude of the difference in an outcome between interventions as the proportion of a standard deviation. The effect size estimate used in this study is Hedges's *g* following What Works Clearinghouse guidance (U.S. Department of Education, 2014).

Fidelity of implementation. The percentage of the lesson in which instruction followed the lesson sequence and script for each of the skills taught. Fidelity of implementation was assessed by the study team twice a year.

Improvement index. The improvement index describes the magnitude of the difference in an outcome between interventions in terms of percentile rank (U.S. Department of Education, 2014). In this study the improvement index reflects the expected change in percentile rank of an average student in an embedded intervention school had the student been in a standalone intervention school.

Low-performing schools. Schools identified by the state's school grading system as having a grade of C or D. Schools receive a grade on a scale of A (best) to F (worst) based on the percentage of students scoring at the proficient level on the state reading test and the percentage making learning gains on the test.

Percentile rank. The percentile rank is the percentage of scores that fall at or below a given score on an outcome.

Reading and language outcomes. The study included reading and language outcomes that were collected in May of each school year from the FRA, the Stanford Early Scholastic Achievement Test (SESAT), and the Stanford Achievement Test, 10th edition (SAT-10). Reading outcomes included the Phonological Awareness (kindergarten only), Word Reading, and Spelling (grade 2 only) subtests from the FRA and the Word Reading subtest from SESAT in kindergarten. Language outcomes included the Vocabulary Pairs, Following Directions, and Sentence Comprehension subtests from the FRA, the Sentence Reading subtest from the SESAT in kindergarten, and the Reading Comprehension subtest from SAT-10 in grades 1 and 2.

Students at risk of literacy failure. Students who scored below the 30th percentile at baseline on the Phonological Awareness (kindergarten only), Word Reading (grades 1 and 2), or Vocabulary Pairs subtest of the FRA.

members were concerned about cost of materials, alignment to curriculum standards, and ease of implementation.

To address the question of effectiveness and ease of implementation, the approach to instructional materials was incorporated into the design of this study (Dombek, Foorman, Garcia, & Smith, 2016). One approach is to use the tier 2 intervention materials embedded in the existing core reading program for classroom instruction. That approach is appealing because the embedded materials are aligned with core classroom instruction and do not require buying additional materials. But even though the embedded tier 2 materials may claim to be research based, they are rarely evaluated empirically.

Another approach is to select tier 2 standalone instructional materials and strategies that are outside the core reading program. Some have been rated by the What Works Clearinghouse as having strong evidence of positive effects on reading and language outcomes. It is reasonable to expect that a standalone intervention with a strong evidence base will lead to better reading and language outcomes for small-group tier 2 intervention than will an embedded intervention that has not been empirically evaluated.

This study used a cluster-level randomized controlled trial conducted across the 2013/14 and 2014/15 school years in 55 low-performing Florida schools to evaluate the effectiveness of the standalone and embedded early literacy interventions

What the study examined

To evaluate the effectiveness of an intervention, it should be compared with logical alternatives, preferably in a random assignment design using appropriate outcome measures. This study used a cluster-level randomized controlled trial conducted across the 2013/14 and 2014/15 school years (referred to as cohort 1 and cohort 2) in 55 low-performing Florida schools, as identified by the state's school grading system.

The study addressed three research questions separately for students in kindergarten, grade 1, and grade 2 who are at risk of literacy failure:
- What are the improvements in percentile rank on reading and language measures in the standalone and embedded early literacy interventions?
- What are the impacts of a standalone early literacy intervention relative to an embedded early literacy intervention on reading and language outcomes? Does the impact differ by baseline performance or cohort (2013/14 and 2014/15)?
- What are the impacts of a standalone early literacy intervention relative to an embedded early literacy intervention on reading and language outcomes for English learner students and non–English learner students? Does the impact differ between English learner students and non–English learner students in each intervention?

Box 2 describes the standalone and embedded interventions and summarizes the data and methods used in the study, and the appendix provides details. Figure 1 describes the two early literacy intervention approaches compared in this impact study. There is no control or "business as usual" group.

Box 2. Descriptions of interventions, data, and methods

Descriptions of interventions

Two approaches to early literacy intervention were compared: a standalone intervention and an intervention embedded in the core curriculum. The standalone intervention combined a reading component and two oral language components: Sound Partners (reading component), a What Works Clearinghouse–reviewed intervention that had strong levels of evidence in alphabetics, fluency, and comprehension (taught daily); Bridge of Vocabulary (oral language component), which focuses on building oral vocabulary and concepts using manipulatives and discussion (taught three times a week); and Language in Motion (oral language component), an inferential language program that uses science-based manipulatives to build oral language components of syntax, inferential language, and listening comprehension (taught twice a week). The intervention embedded in the core curriculum combined a reading component and an oral language component that were both included within Houghton Mifflin Harcourt *Journeys* (the core curriculum followed in all the study schools): the tier 2 Strategic Intervention (reading component) and Curious about Words (a supplementary vocabulary piece that made up the oral language component); both were taught daily.

Each school had three to four interventionists who taught the lessons associated with each intervention, serving four to six small groups daily. Interventionists had experience working with young children in education settings and received two days of training in late September. Some interventionists were school-based paraprofessionals assigned by the schools, and others were hired by Regional Educational Laboratory (REL) Southeast. For cohort 1, REL Southeast provided 66 interventionists, schools provided 17 paraprofessionals, and together they served 370 small groups; 32 percent of the interventionists were certified teachers. For cohort 2, REL Southeast provided 64 interventionists (42 percent of whom were interventionists for cohort 1 schools), schools provided 25 paraprofessionals, and together they served 424 small groups; 37 percent of the interventionists were certified teachers.

The study team observed interventionists once in the fall and once in the spring to rate fidelity of implementation. Separate fidelity ratings were calculated for each small group for the reading and oral language components, and the fall and spring fidelity ratings for each small group were averaged to create overall fidelity ratings for each component. For both interventions, 72–91 percent of small groups demonstrated at least 80 percent fidelity on the reading and oral language components (see table A3 in the appendix). The median overall fidelity across interventions was 96 percent in kindergarten, 94 percent in grade 1, and 96 percent in grade 2.

Across grades K–2, interventionists covered an average of 55–80 percent of the reading component and 77–79 percent of the oral language component in the standalone intervention and 86–88 percent of the reading and oral language components in the embedded intervention (see table A4 in the appendix). Out of 134 days of instruction, students in standalone intervention schools attended 92–95 days of intervention on average, and students in embedded intervention schools attended 96–98 days (see table A5 in the appendix).

Data

The study used data provided by schools in a large urban district in south Florida, a medium-size urban district in central Florida, and three small rural districts in north Florida. There were two nonoverlapping cohorts of schools: cohort 1 included 27 schools and 1,598 students that participated in the 2013/14 school year, and cohort 2 included 28 schools and 1,870 students that participated in the 2014/15 school year (see figures A1–A3 in the appendix).[1] All participating schools were low performing, as identified by the state's school grading system. Participating students were in grades K–2, were at risk of literacy failure, and had parent consent to participate. The average percentage of students who qualified for the federal school lunch program (a proxy for low-income status) ranged from 72 percent to 78 percent for cohorts 1 and 2 combined across interventions and grades (see table A1 in the appendix for school demographics). Approximately 30–42 percent of participating students in cohorts 1 and 2 combined across interventions and grades were English learner students (see table A2 for student demographics).

Several reading and language measures were included at baseline and outcome. Reading baseline measures were the Letter Sounds (kindergarten only), Phonological Awareness (kindergarten only), Word Reading (grades 1

Box 2. Descriptions of interventions, data, and methods *(continued)*

and 2), and Spelling (grade 2 only) subtests from the Florida Center for Reading Research Assessment (FRA; see table A6 in the appendix). Language baseline measures were the Vocabulary Pairs, Following Directions, and Sentence Comprehension (kindergarten and grade 1) subtests from the FRA. Reading outcomes were the Phonological Awareness (kindergarten only), Word Reading, and Spelling (grade 2 only) subtests from the FRA (see table A6 in the appendix) and the Word Reading subtest from the Stanford Early Scholastic Achievement Test (SESAT) in kindergarten. Language outcomes were the Vocabulary Pairs, Following Directions, and Sentence Comprehension subtests from the FRA; the Sentence Reading subtest from the SESAT in kindergarten; and the Reading Comprehension subtest from the Stanford Achievement Test, 10th edition in grades 1 and 2.

Methods

Participating schools were randomly assigned to use a standalone or embedded approach to early literacy intervention. Students received daily pull-out intervention for 45 minutes from mid-October through May, about 27 weeks, in small groups of four (kindergarten and grade 1) or five (grade 2). About 30 minutes were devoted to the reading component, and about 15 minutes to the oral language component.

Prior to analyses, baseline equivalence was assessed by comparing differences between the interventions on all reading and language baseline measures by grade at the school and student levels. Most of the differences in baseline scores by grade at the school and student levels between students in standalone intervention schools and students in embedded intervention schools were not statistically significant (see tables A7–A10 in the appendix). One exception was the FRA Word Reading subtest for grade 1, where baseline scores were significantly higher for students in embedded intervention schools than for students in standalone intervention schools (see the appendix).

Multilevel analyses of student outcomes were conducted by grade, with students nested in small groups, nested within schools. All analyses included student, small-group, and school-level baseline measures as covariates (see the appendix). Baseline scores were aggregated by small group and then by school and were used as covariates at their respective levels. Cohort and region were also included as school-level covariates. Cohort was included as an analytic variable because different schools participated each year and the calculation of school grades changed with a change in the state reading test in 2013/14. As a result, participating districts recommended even lower performing schools in cohort 2 (2014/15) than in cohort 1 (2013/14).

Differences in outcomes between the interventions are reported in three ways: statistical significance, effect size, and improvement index (see box 1 and the appendix).

Note

1. One of the standalone intervention schools in cohort 2 was excluded from the grade 2 analyses because scheduling conflicts resulted in the withdrawal of the 21 participating grade 2 students at that school. The cohort total includes these 21 students.

What the study found

This section discusses the findings of the study, starting with baseline and outcome percentile ranks on the reading and language measures by grade and intervention. It then reports differences in reading and language outcomes between the standalone and embedded interventions for all students and by cohort and baseline performance. Finally, it reports differences in reading and language outcomes by English learner status.

Comparable improvements in percentile ranks between standalone and embedded interventions on reading and language measures

In grades K–2, students in schools in both intervention groups started, on average, at or below the 10th percentile on FRA reading measures (Phonological Awareness in

Figure 1. Two approaches to early literacy intervention, standalone and embedded, were delivered to at-risk students in grades K–2 in small groups

Early literacy interventions
Daily (October–May) for 45 minutes per session
Kindergarten and grade 1: four students per group
Grade 2: five students per group

Standalone intervention

Sound Partners
Daily (25-30 minutes)
Explicit and systematic instruction in phonemic awareness, phonics, spelling, and fluent blending

Bridge of Vocabulary
Three times per week (15 minutes)
Uses manipulatives and discussion to build vocabulary skills in listening, speaking, and reading

Language in Motion
Two times per week (15 minutes)
Builds syntax, language, and listening comprehension skills through the use of science-based manipulatives, interactive stories, and games to apply skills learned

Embedded intervention

Journeys **Tier 2 Strategic Intervention**
Daily (25–30 minutes)
Aligned with tier 1 content and builds skills in phonemic awareness, phonics, fluency, comprehension, and vocabulary

Journeys **Tier 2 Curious About Words**
Daily (15 minutes)
Supplements tier 1 content and introduces 8–12 new vocabulary words per week (depending on grade)
Lessons include: readalouds, using graphic organizers, teacher-led discussion, and partner activities

Source: Authors' compilation based on information provided in curricula materials.

kindergarten, Word Reading in grades 1 and 2, and Spelling in grade 2) and ended the year above the 20th percentile, except FRA Spelling in grade 2 (table 1). The average difference between baseline and outcome percentile ranks on FRA reading measures was 13–25 percentile points across grades.

In kindergarten, students in both intervention groups started, on average, at or below the 10th percentile on two of the FRA language measures (Following Directions and Sentence Comprehension) and ended the year above the 25th percentile. The average difference between baseline and outcome percentile ranks for these FRA language measures was 20–25 percentile points.

In grades 1 and 2, students in schools in both intervention groups started, on average, between the 10th and 15th percentiles on two of the FRA language measures (Following Directions and Vocabulary Pairs) and ended the year between the 18th and 30th percentiles. The average difference between baseline and outcome percentile ranks for these FRA language measures was 6–15 percentile points.

The largest average difference between baseline and outcome percentile ranks for any FRA measure was Sentence Comprehension in grade 1. Students in schools in both intervention groups began just below the 30th percentile and ended the year above the 60th percentile. This reflects an average difference of 35–39 percentile points between baseline and outcome percentile ranks across interventions. However, the norms for FRA Sentence Comprehension are based on kindergarten students, which means that the percentile ranks for all grades reflect ability on a kindergarten scale.

The average difference between baseline and outcome percentile ranks on FRA reading measures was 13–25 percentile points across grades

Table 1. Early literacy intervention average student baseline and outcome percentile rank and difference, by grade, outcome type, measure, and intervention group, 2013/14 and 2014/15

Grade, outcome type, and measure	Standalone intervention			Embedded intervention		
	Baseline	Outcome	Difference	Baseline	Outcome	Difference
Kindergarten						
Reading outcomes						
FRA Phonological Awareness	1	21	20	1	26	25
FRA Word Reading	na	31	na	na	29	na
SESAT Word Reading	na	26	na	na	20	na
Language outcomes						
FRA Vocabulary Pairs	25	34	9	24	33	9
FRA Following Directions	7	27	20	5	26	21
FRA Sentence Comprehension[a]	10	35	25	9	32	23
SESAT Sentence Reading	na	23	na	na	22	na
Grade 1						
Reading outcomes						
FRA Word Reading	1	23	22	1	26	25
Language outcomes						
FRA Vocabulary Pairs	12	18	6	12	18	6
FRA Following Directions	10	19	9	11	21	10
FRA Sentence Comprehension[a]	29	64	35	27	66	39
SAT-10 Reading Comprehension	na	13	na	na	13	na
Grade 2						
Reading outcomes						
FRA Word Reading	5	24	19	9	26	17
FRA Spelling	3	22	19	4	17	13
Language outcomes						
FRA Vocabulary Pairs	12	22	10	10	18	8
FRA Following Directions	15	30	15	13	26	13
FRA Sentence Comprehension[a]	58[b]	87	29	57[b]	82	25
SAT-10 Reading Comprehension	na	15	na	na	14	na

FRA is the Florida Center for Reading Research Reading Assessment. SESAT is the Stanford Early Scholastic Achievement Test. SAT-10 is the Stanford Achievement Test, 10th edition. na is not applicable.

Note: Percentile ranks are based on winter norms.

a. The FRA Sentence Comprehension subtest is a kindergarten-normed assessment, so the percentile ranks for all grades reflect ability on a kindergarten scale.

b. Available only for cohort 2.

Source: Authors' analysis based on data from participating districts in Florida (see the appendix).

Relative impacts of the two interventions

In grade 2 the standalone intervention resulted in significantly improved spelling outcomes relative to the embedded intervention, including among students with a low FRA Spelling baseline score. The average FRA Spelling outcome among grade 2 students was a score of 434 in standalone intervention schools and 417 in embedded intervention schools (see table A15 in the appendix). This statistically significant 17-point difference is equivalent to 0.18 standard deviation and 7 percentile rank points. In other words, grade 2 students in embedded intervention schools would have improved, on average, by 7 percentile points had they been in a standalone intervention school. The FRA Spelling outcome

was also significantly higher in standalone intervention schools than in embedded intervention schools for grade 2 students with a low FRA Spelling baseline score of 233 (one standard deviation below the mean; table 2). The average FRA Spelling outcome among a subgroup of grade 2 students with a low FRA Spelling baseline score was 404 in standalone intervention schools and 379 in embedded intervention schools. This statistically significant 25-point difference is equivalent to 0.27 standard deviation and 11 percentile rank points.

Table 2. Early literacy intervention reading outcomes among grade K–2 students, by grade, outcome measure, cohort, baseline score, and intervention group, 2013/14 and 2014/15

Grade and outcome measure, cohort, and baseline score	Sample size		Adjusted mean					
	Standalone intervention	Embedded intervention	Standalone intervention (standard deviation)	Embedded intervention (standard deviation)	Difference (standard error)	*p* value	Effect size	Improvement index[a]
Kindergarten, SESAT Word Reading outcome								
Cohort 1								
Students with a high FRA Sentence Comprehension baseline score[b]	255	213	435 (39)	421 (33)	14 (8)	.09	0.37	14
Students with a low FRA Sentence Comprehension baseline score[c]	255	213	426 (39)	418 (33)	8 (8)	.34	0.22	9
Cohort 2								
Students with a high FRA Sentence Comprehension baseline score[b]	276	317	435 (36)	437 (35)	–2 (7)	.76	–0.06	–2
Students with a low FRA Sentence Comprehension baseline score[c]	276	317	435 (36)	425 (35)	10 (7)	.22	0.28	11
Grade 1, FRA Word Reading outcome								
Cohort 1	267	239	461 (63)	483 (89)	–22 (17)	.19	–0.29	–11
Cohort 2	267	325	433 (131)	411 (129)	22 (15)	.13	0.17	7
Grade 2, FRA Spelling outcome								
Students with a high FRA Spelling baseline score[b]	618	670	462 (89)	453 (98)	9 (8)	.29	0.10	4
Students with a low FRA Spelling baseline score[c]	618	670	404 (89)	379 (98)	25 (8)	.001*	0.27	11

SESAT is the Stanford Early Scholastic Achievement Test. FRA is the Florida Center for Reading Research Reading Assessment.

* *p*-value is significant after applying the Benjamini–Hochberg Correction procedure (1995) where the identified *p*-value cutoff is $p \leq .0025$.

Note: A hierarchical linear model with students nested in small groups and small groups nested in schools was estimated for each of grades K–2. For each outcome measure the full subgroup model included all grade-specific baseline scores; several dichotomous indicators for region, cohort, and treatment; and several interactions, including baseline score by treatment, cohort by treatment, baseline score by cohort, and baseline score by cohort by treatment (see the appendix for the model equation). Only outcomes with a significant interaction involving the treatment indicator (baseline score by treatment, cohort by treatment, or baseline score by cohort by treatment) were probed further and included in the table.

a. The expected change in percentile rank of an average student in an embedded intervention school had the student been in a standalone intervention school.

b. Refers to baseline scores that are one standard deviation above the mean.

c. Refers to baseline scores that are one standard deviation below the mean.

Source: Authors' analysis based on data from participating districts in Florida (see the appendix).

There were no other differences in reading outcomes between students in standalone and embedded intervention schools for grades K–2.

The standalone intervention resulted in substantively important differences (effect size greater than 0.25) relative to the embedded intervention on the Stanford Early Scholastic Achievement Test (SESAT) Word Reading outcome in kindergarten and the FRA Word Reading outcome in grade 1. In kindergarten the SESAT Word Reading outcome was higher in standalone intervention schools than in embedded intervention schools among students in both cohorts. For cohort 1 (students in 2013/14) there was a 14-point difference—equivalent to 0.37 standard deviation and 14 percentile rank points—among kindergarten students with an FRA Sentence Comprehension baseline score of 487 (one standard deviation above the mean; see table 2). For cohort 2 (students in 2014/15) there was a 10-point difference—equivalent to 0.28 standard deviation and 11 percentile rank points—among kindergarten students with an FRA Sentence Comprehension baseline score of 313 (one standard deviation below the mean). In grade 1 the FRA Word Reading outcome was higher in embedded intervention schools than in standalone intervention schools among students in cohort 1. The difference was 22 points—equivalent to 0.29 standard deviation and 11 percentile rank points.

There were no differences in the FRA Phonological Awareness outcome in kindergarten or FRA Word Reading outcomes in kindergarten and grade 2 by cohort or baseline score between students in standalone and embedded intervention schools.

In grade 2 the standalone intervention resulted in a significantly improved FRA Sentence Comprehension outcome relative to the embedded intervention among students in cohort 1 with a low FRA Vocabulary Pairs baseline score. The average estimated FRA Sentence Comprehension outcome among grade 2 students in cohort 1 with an FRA Vocabulary Pairs baseline score of 414 (one standard deviation below the mean) was a score of 597 in standalone intervention schools and 559 in embedded intervention schools (table 3). This statistically significant 38-point difference is equivalent to 0.38 standard deviation and 15 percentile rank points.

There were no differences in any language outcomes in kindergarten by cohort or baseline performance between students in standalone and students in embedded intervention schools. Nor were there any differences in the FRA Vocabulary Pairs outcomes in grades 1 and 2, the FRA Following Directions outcome in grade 2, the FRA Sentence Comprehension outcome in grade 1, or the SAT-10 Reading Comprehension outcomes in grades 1 and 2 by cohort or baseline performance between students in standalone and embedded intervention schools.

Differences in outcomes between and within interventions for English learner and non–English learner students

This section describes the results of exploring differences in reading and language outcomes between English learner students in standalone and embedded intervention schools and differences in reading and language outcomes between English learner and non–English learner students in schools in the same intervention group. There were no differences in language outcomes in kindergarten or in reading or language outcomes in

In kindergarten the SESAT Word Reading outcome was higher in standalone intervention schools than in embedded intervention schools among students in both cohorts

Table 3. Early literacy intervention language outcomes among grade K–2 students, by grade and outcome measure, cohort, baseline score, and intervention group, 2013/14 and 2014/15

Grade and outcome measure, cohort, and baseline score	Sample size		Adjusted mean		Difference (standard error)	p value	Effect size	Improvement index[a]
	Standalone intervention	Embedded intervention	Standalone intervention (standard deviation)	Embedded intervention (standard deviation)				
Grade 1, FRA Following Directions outcome								
Students with a high FRA Word Reading baseline score[b]	534	564	453 (109)	435 (117)	18 (11)	.11	0.16	6
Students with a low FRA Word Reading baseline score[c]	534	564	433 (109)	445 (117)	–12 (10)	.27	–0.11	–4
Students with a high FRA Following Directions baseline score[b]	534	564	484 (109)	495 (117)	–11 (11)	.26	–0.10	–4
Students with a low FRA Following Directions baseline score[c]	534	564	402 (109)	385 (117)	–17 (11)	.09	0.15	6
Grade 2, FRA Sentence Comprehension outcome								
Cohort 1								
Students with a high FRA Vocabulary Pairs baseline score[b]	323	301	607 (99)	589 (101)	18 (11)	.12	0.18	7
Students with a low FRA Vocabulary Pairs baseline score[c]	323	301	597 (99)	559 (101)	38 (12)	.001*	0.38	15
Cohort 2								
Students with a high FRA Vocabulary Pairs baseline score[b]	295	369	616 (76)	603 (75)	13 (12)	.28	0.18	7
Students with a low FRA Vocabulary Pairs baseline score[c]	295	369	585 (76)	594 (75)	–9 (12)	.71	–0.12	–5

FRA is the Florida Center for Reading Research Reading Assessment.

* p-value is significant after applying the Benjamini–Hochberg Correction procedure (1995), where the identified p-value cutoff is $p \leq .00125$.

Note: A hierarchical linear model with students nested in small groups and small groups nested in schools was estimated for each of grades K–2. For each outcome measure the full subgroup model included all grade-specific baseline scores; several dichotomous indicators for region, cohort, and treatment; and several interactions, including baseline score by treatment, cohort by treatment, baseline score by cohort, and baseline score by cohort by treatment (see the appendix for model equation). Only outcomes with a significant interaction involving the treatment indicator (baseline score by treatment, cohort by treatment, or baseline score by cohort by treatment) was probed further and included in the table.

a. The expected change in percentile rank of an average student in an embedded intervention school had the student been in a standalone intervention school.

b. Refers to baseline scores that are one standard deviation above the mean.

c. Refers to baseline scores that are one standard deviation below the mean.

Source: Authors' analysis based on data from participating districts in Florida (see the appendix).

grades 1 and 2. However, there were three substantively important differences in reading outcomes in kindergarten.

In kindergarten the FRA Phonological Awareness outcome was higher in embedded intervention schools than in standalone intervention schools among English learner students, while the SESAT Word Reading outcome was higher in standalone intervention schools than in embedded intervention schools among non–English learner students. The average FRA Phonological Awareness outcome among kindergarten English learner students was 45 points higher in embedded intervention schools than in standalone intervention schools (table 4). The difference is equivalent to 0.32 standard deviation and 12 percentile rank points.

The SESAT Word Reading outcome was higher for the standalone intervention than for the embedded intervention among non–English learner students (see table 4). The difference was 11 points—equivalent to 0.31 standard deviation and 12 percentile rank points.

In kindergarten the SESAT Word Reading outcome was higher among English learner students in embedded intervention schools than among non–English learner students in embedded intervention schools. The 9-point difference is equivalent to 0.27 standard deviation (table 5). There were no differences in other reading outcomes between English learner students and non–English learner students in schools in the same intervention group.

The average FRA Phonological Awareness outcome among kindergarten English learner students was 45 points higher in embedded intervention schools than in standalone intervention schools

Table 4. Early literacy intervention reading outcomes among kindergarten students, by English learner status and intervention group, 2013/14 and 2014/15

Outcome measure and subgroup	Sample size		Adjusted mean		Difference (standard error)	p value	Effect size	Improvement index[a]
	Standalone intervention	Embedded intervention	Standalone intervention (standard deviation)	Embedded intervention (standard deviation)				
FRA Phonological Awareness								
Non–English learner students	343	297	435 (147)	439 (131)	−4 (15)	.79	−0.03	−1
English learner students	169	213	425 (146)	470 (138)	−45 (18)	.01†	−0.32	−12
FRA Word Reading								
Non–English learner students	343	297	327 (138)	354 (151)	−27 (17)	.11	−0.19	−7
English learner students	169	213	333 (129)	328 (146)	5 (14)	.69	0.04	1
SESAT Word Reading								
Non–English learner students	343	297	433 (37)	422 (34)	11 (5)	.04†	0.31	12
English learner students	169	213	429 (41)	431 (36)	−2 (6)	.76	−0.05	−2

FRA is the Florida Center for Reading Research Reading Assessment. SESAT is the Stanford Early Scholastic Achievement Test.

† *p*-value is not significant after applying the Benjamini–Hochberg Correction procedure (1995), where the identified *p*-value cutoff is $p \le .004$.

a. The expected change in percentile rank of an average student in an embedded intervention school had the student been in a standalone intervention school.

Source: Authors' analysis based on data from participating districts in Florida (see the appendix).

Table 5. Early literacy intervention reading outcomes among kindergarten students, by intervention group and English learner status, 2013/14 and 2014/15

Outcome measure and intervention group	Sample size		Adjusted mean				
	English learner students	Non English learner students	English learner students (standard deviation)	Non English learner students (standard deviation)	Difference (standard error)	p value	Effect size
FRA Phonological Awareness							
Standalone	169	343	425 (146)	435 (147)	–10 (15)	.50	–0.07
Embedded	213	297	470 (138)	439 (131)	31 (14)	.02†	0.23
FRA Word Reading							
Standalone	169	342	327 (138)	333 (129)	–6 (13)	.61	–0.04
Embedded	213	297	354 (151)	328 (146)	26 (12)	.03†	0.18
SESAT Word Reading							
Standalone	169	343	429 (41)	433 (37)	–4 (4)	.28	–0.10
Embedded	213	297	431 (36)	422 (34)	9 (3)	.006†	0.27

FRA is the Florida Center for Reading Research Reading Assessment. SESAT is the Stanford Early Scholastic Achievement Test.

† *p*-value is not significant after applying the Benjamini–Hochberg Correction procedure (1995), where the identified *p*-value cutoff is $p \leq .004$.

Source: Authors' analysis based on data from participating districts in Florida (see the appendix).

Implications of the study findings

This section discusses four major implications of the study findings.

Improvement was comparable in reading and language outcomes among at-risk students in schools in both intervention groups

On average, students in grades K–2 showed improvement in reading and language outcomes in both the standalone and embedded small-group tier 2 interventions. Students started the school year below the 10th percentile on the FRA Phonological Awareness and Word Reading measures and ended the year above the 20th percentile. Kindergarten students scored between the 20th and 26th percentile on the SESAT Word Reading and Sentence Reading outcomes at the end of the year across both cohorts. However, word reading skills were not sufficiently developed for students to achieve, on average, reading comprehension outcomes above the 15th percentile in grades 1 and 2. Starting intensive intervention in kindergarten and increasing the intensity in a multitiered system of support for students who fail to respond is one way to improve mastery of alphabetic skills to enable students to comprehend what they read (Gersten et al., 2009). However, the observed gains in literary skills should be interpreted cautiously because they might be due to regression to the mean (from students' very low baseline score), which occurs when an initially low or high score gravitates toward the mean on subsequent assessment. It is also possible that the observed gains in literacy skills are due to expected normative growth (solely from classroom instruction) rather than to an intervention.

Students started the school year below the 10th percentile on the FRA Phonological Awareness and Word Reading measures and ended the year above the 20th percentile

The largest change in average percentile points on the language outcomes was in FRA Sentence Comprehension in kindergarten and grade 1. The norms for this subtest are based on kindergarten students, which means that the percentile ranks for grade 1 reflect ability on a kindergarten scale. Students in kindergarten gained 24 percentile points across intervention

12

groups, and students in grade 1 gained 37 percentile points. The FRA Sentence Comprehension measure is a listening comprehension subtest in which students point to one of four pictures that corresponds to a sentence given by the computer (for example, "Point to the bird flying away from the nest"). This measure is similar to a task in the Comprehensive English Language Learning Assessment (Educational Testing Service, 2005) that was used in Florida at the time of this study to identify and designate students as English learner students. The lack of a significant interaction between the intervention and English learner status on the FRA Sentence Comprehension outcome suggests that once English learner and non–English learner students' FRA Sentence Comprehension baseline scores are taken into account, the two groups performed similarly on the FRA Sentence Comprehension outcome.

The two interventions had similar impacts on reading and language outcomes, except for spelling in grade 2

Reading and language outcomes were comparable in standalone and embedded intervention schools, except that the standalone intervention resulted in a significantly improved FRA Spelling outcome in grade 2, the only grade with a spelling outcome. Although the reading component of the standalone intervention (Sound Partners) required students to spell the words they learned to read in all three grades, spelling was measured only in grade 2. The reading component of the embedded intervention (Strategic Intervention) did not require students to spell the words they learned to read. By teaching students to encode (spell) as well as decode the words taught, Sound Partners is similar to other early reading interventions with significant impacts on reading outcomes (Foorman, Beyler, et al., 2016). However, in the current study a statistically significant difference for the standalone intervention relative to the embedded intervention in grade 2 was found only for spelling and not for other reading outcomes.

Inconsistent differences in intervention outcomes between students in standalone and embedded intervention schools by cohort and baseline scores suggest that the interventions had comparable effects on reading and language outcomes, except for spelling

Aside from the significantly improved spelling outcomes in grade 2 for students in standalone intervention schools relative to students in embedded intervention schools, the pattern of relative effects of the two interventions by cohort and baseline scores across all grades was inconsistent. Specifically, the standalone intervention resulted in significantly improved spelling outcomes relative to the embedded intervention among students with low baseline spelling scores across cohorts. The standalone intervention also had one significant effect relative to the embedded intervention on a language outcome in grade 2 in cohort 1 and two substantively important effects on the SESAT Word Reading outcome in kindergarten—one in cohort 1 and one in cohort 2. Inconsistent with these results is the finding that the embedded intervention resulted in a substantively improved FRA Word Reading outcome relative to the standalone intervention among students in cohort 1 schools in grade 1. The lack of a consistent pattern of effects across cohorts (except for spelling) implies that, on average, improvement was comparable among students in schools in both intervention groups.

The two interventions had similar impacts on reading and language outcomes by English learner status

There were no differences in reading and language outcomes in grades 1 and 2 or in language outcomes in kindergarten between English learner students and non–English

The standalone intervention resulted in significantly improved spelling outcomes relative to the embedded intervention among students with low baseline spelling scores across cohorts, but the embedded intervention resulted in a significantly improved FRA Word Reading outcome relative to the standalone intervention among students in cohort 1 schools in grade 1

13

learner students in schools in the same intervention group. However, there was a difference between kindergarten English learner students in schools in the two intervention groups. The FRA Phonological Awareness outcome among kindergarten English learner students was higher in embedded intervention schools than in standalone intervention schools.

Conversely, the SESAT Word Reading outcome among kindergarten non–English learner students was higher in standalone intervention schools than in embedded intervention schools.

Both interventions included instruction in phonological awareness, but the addition of comprehension activities in the embedded intervention may have helped scaffold English learner students' ability to segment sounds in speech. This finding is consistent with studies showing an advantage in phonological awareness tasks for bilingual students (for example, Bialystok, Majumder, & Martin, 2003). The fact that the non–English learner students in standalone intervention schools scored higher on the SESAT Word Reading outcome than did their peers in embedded intervention schools suggests that the decontextualized nature of alphabetic instruction in Sound Partners was sufficient to build their word reading skills.

The embedded intervention resulted in a significantly improved SESAT Word Reading outcome in kindergarten among English learner students relative to non–English learner students. These results underscore the value of emphasizing comprehension when building on English learner students' sensitivity to sounds in speech in order to connect to the sound-spelling patterns fundamental to reading.

The study also has implications for future research on early literacy interventions. Experiments could modify the standalone intervention in ways that might make it easier to implement. First, it was challenging for interventionists to decide how to remediate students on different skills and what to do with students who did not need remediation (see the appendix for a description of remediation). A version of the reading component of the standalone intervention that eliminates remediation could be contrasted with the current version to see whether student reading outcomes differed. Second, interventionists had to remember which day to teach vocabulary and which day to teach inferential language during the week. This was challenging because of the disruptions in school schedules that required interventionists to remember which language piece had to be rescheduled. An integrated version of the language component in the standalone intervention where vocabulary and inferential language are taught each day could be contrasted with the current version to see whether student language outcomes differed.

An area of investigation for the embedded intervention is to verify its alignment to core classroom (tier 1) instruction and then to manipulate enhancements to both core classroom (tier 1) instruction and small-group (tier 2) instruction. To enhance and, thereby, achieve high implementation fidelity in the embedded intervention in the current study, REL Southeast staff developed an implementation manual that revealed the scope and sequence and established procedures for well trained interventionists to deliver daily small-group intervention in a consistent fashion to a diverse population of students. Once this enhanced implementation of the tier 2 embedded intervention is developed, the next step in studying modifications is to compare the current version of enhanced tier 2 and typical

Future research could contrast a version of the reading component of the standalone intervention that eliminates remediation with the current version to see whether student reading outcomes differed or an integrated version of the language component in the standalone intervention where vocabulary and inferential language are taught each day with the current version to see whether student language outcomes differed

tier 1 with a version where both are enhanced. Smith et al. (2016) found higher reading outcomes for at-risk students in the primary grades when they received enhanced tier 1 and 2 instruction compared with when they received the typical, nonenhanced tier 1 and 2 instruction. Tier 1 might be enhanced by making evidence-based elements more explicit and providing more scaffolding so that instruction is accessible to a broad range of students (for example, Smith et al., 2016). Additionally, Gersten et al. (in press) found that all effective reading interventions in the primary grades provided ongoing support to the adult delivering the tier 2 intervention.

Limitations of the study

The study has one main limitation: the lack of a control group (or business-as-usual group) that did not receive any intervention against which to compare the gains of the standalone and embedded intervention groups. But denying intervention to at-risk students is not an option in Florida schools, and business-as-usual differs across and even within schools and is constantly changing (Lemons, Fuchs, Gilbert, & Fuchs, 2014).

Appendix. Data, outcomes, intervention, and methodology

The appendix provides details on the study data; interventions and interventionists; implementation fidelity; measures of coverage and attendance; attrition; treatment of missing data; and methodology.

Data

The study used data provided by 55 schools in five districts in Florida. Two cohorts of schools were recruited to participate for one year. Participating schools did not overlap between the two cohorts, but two of the participating districts did overlap between the two cohorts. The first cohort of schools participated in the 2013/14 school year and included 27 schools across four districts: 16 schools in a large urban district in south Florida, 8 schools in a medium-size urban district in central Florida, and 3 schools in two small rural districts in north Florida. The second cohort of schools participated in the 2014/15 school year and included 28 schools across three districts: 16 schools in the same large urban district in south Florida as in cohort 1, 9 schools in the same medium-size urban district in central Florida as in cohort 1, and 3 schools from a different rural district in north Florida.

All participating schools were low-performing schools, as defined by the state's school grades system, which determines a grade on a scale of A (best) to F (worst) based on the percentage of students scoring at the proficient level and the percentage of students making learning gains on the state reading test. Districts requested that recruitment take place with schools that received a C or D and not with schools that received an F, which state accountability teams were involved in restructuring.

Each school was randomly assigned to implement a standalone intervention or embedded intervention in grades K–2. Random assignment was conducted within cohort and region (north, central, and south) by the study team. The random assignment process was conducted during the summer preceding the start of each school year using Microsoft Excel and consisted of three steps:

1. Assign a random number to each school by region within cohort.

2. Order schools in descending order within each region and cohort by the assigned random number.

3. Assign the first half within region and cohort[1] to the standalone intervention group and the second half to the embedded intervention group.

Of the 55 participating schools in the sample, 27 were randomly assigned to the standalone intervention (14 schools in cohort 1 and 13 schools in cohort 2), and 28 were randomly assigned to the embedded intervention (13 schools in cohort 1 and 15 schools in cohort 2).[2] The average percentage of students who qualified for the federal school lunch program (a proxy for low-income status) ranged from 72 percent to 78 percent for cohorts 1 and 2 combined across interventions and grades (table A1).

During September of each study year, students performing below the 30th percentile on one or more of three K–2 screening subtests from the Florida Center for Reading Research

Reading Assessment (FRA) were identified as eligible for study participation: Phonological Awareness (kindergarten only), Word Reading (grades 1 and 2), and Vocabulary Pairs (grades K–2). Students who were already receiving school services (for example, special education) were removed from the list of eligible students. In late September, school staff examined students' schedules to determine which of the remaining eligible students could be served in the daily 45-minute periods available in the bell schedule for small-group intervention and sent home parent consent forms with those students. School staff continued to send home parent consent forms with students who fit both the eligibility and scheduling criteria until the needed number of participants was achieved.

Across grades K–2, cohort 1 included 468–624 students (divided into 114–133 small groups), and cohort 2 included 592–685 students (divided into 138–143 small groups; table A2). On average, 4.11–5.00 students were in each small group, and each school contained 4.22–5.11 small groups per grade across cohorts 1 and 2. During the first 10 weeks, 6–15 percent of students across grades K–2 moved to another small group because their scores on skill mastery tests were more similar to the scores of students in another small group. Approximately 30–42 percent of participating students across cohorts and grades in schools in the two intervention groups were English learner students.

Figures A1–A3 provide details on enrollment, allocation, follow-up, and data collected and analyzed by grade, intervention group, and cohort.

Table A1. School-level percentage of English learner students and students eligible for the federal school lunch program, by grade, intervention group, and cohort, 2013/14 and 2014/15

Grade and school characteristic	Standalone intervention			Embedded intervention		
	Cohort 1 mean (standard deviation) (*N* = 14)	Cohort 2 mean (standard deviation) (*N* = 13)	Cohorts 1 and 2 combined mean (standard deviation) (*N* = 27)	Cohort 1 mean (standard deviation) (*N* = 13)	Cohort 2 mean (standard deviation) (*N* = 15)	Cohorts 1 and 2 combined mean (standard deviation) (*N* = 28)
Kindergarten						
Percentage of English learner students	18 (19)	27 (17)	22 (18)	22 (19)	31 (18)	26 (19)
Percentage of students eligible for the federal school lunch program	82 (14)	70 (20)	76 (18)	68 (19)	75 (25)	72 (22)
Grade 1						
Percentage of English learner students	20 (17)	25 (20)	22 (18)	19 (17)	30 (19)	25 (18)
Percentage of students eligible for the federal school lunch program	78 (17)	73 (15)	76 (16)	68 (19)	81 (18)	75 (19)
Grade 2						
Percentage of English learner students	16 (14)	23 (17)	20 (15)	19 (17)	28 (19)	24 (18)
Percentage of students eligible for the federal school lunch program	81 (14)	74 (15)	78 (15)	68 (21)	79 (19)	74 (20)

Source: Authors' analysis based on data from participating districts in Florida.

Table A2. Student demographic information by grade, intervention group, and cohort

Grade and student characteristic	Standalone intervention						Embedded intervention					
	Cohort 1		Cohort 2		Cohorts 1 and 2 combined		Cohort 1		Cohort 2		Cohorts 1 and 2 combined	
	Number of students	Mean percentage	Number of students	Mean percentage	Number of students	Mean percentage	Number of students	Mean percentage	Number of students	Mean percentage	Number of students	Mean percentage
Kindergarten												
Male	255	54	276	56	531	55	212	49	317	60	529	55
English learner students	254	26	258	40	512	33	211	45	289	40	510	42
Eligible for the federal school lunch program	254	88	241	88	495	88	212	78	299	88	501	84
Grade 1												
Male	267	51	267	57	534	54	237	54	325	54	562	54
English learner students	265	36	258	29	523	33	230	34	314	36	544	35
Eligible for the federal school lunch program	265	85	256	82	521	83	230	76	312	85	542	81
Grade 2												
Male	323	54	316	59	639	56	301	53	369	54	670	54
English learner students	323	25	308	35	631	30	300	27	347	42	647	35
Eligible for the federal school lunch program	323	89	308	81	631	85	300	73	345	86	645	80

Source: Authors' analysis based on data from participating districts in Florida.

Figure A1. Kindergarten student and school consolidated standards of reporting trials

Enrollment

Total participating schools: n = 55
- Cohort 1 (2013/14 school year): n = 27 schools from four districts (three regions)
- Cohort 2 (2014/15 school year): n = 28 schools from three districts (three regions)

Total eligible students with parent consent: n = 1,061
- Cohort 1: n = 468, Cohort 2: n = 593

Schools were randomly assigned to standalone or embedded intervention within cohort and region

Allocation

Standalone intervention

Schools: n = 27
(cohort 1: n = 14, cohort 2: n = 13)

Total students within schools: n = 531
(cohort 1: n = 255, cohort 2: n = 276)

Total small groups within schools: n = 131
(cohort 1: n = 63, cohort 2: n = 68)

Embedded intervention

Schools: n = 28
(cohort 1: n = 13, cohort 2: n = 15)

Total students within schools: n = 530
(cohort 1: n = 213, cohort 2: n = 317)

Total small groups within schools: n = 126
(cohort 1: n = 51, cohort 2: n = 75)

Follow-up

Missing outcome data:
- Schools: n = 0
- Students: n = 57
  (cohort 1: n = 30, cohort 2: n = 27)

Missing outcome data:
- Schools: n = 0
- Students: n = 56
  (cohort 1: n = 20, cohort 2: n = 36)

Data collected and analyzed

Analytic sample:
- Schools: n = 27
  (cohort 1: n = 14, cohort 2: n = 13)
- Students: n = 474
  (cohort 1: n = 225, cohort 2: n = 249)
- Groups: n = 131
  (cohort 1: n = 63, cohort 2: n = 68)

Analytic sample:
- Schools: n = 28
  (cohort 1: n = 13, cohort 2: n = 15)
- Students: n = 474
  (cohort 1: n = 193, cohort 2: n = 281)
- Groups: n = 126
  (cohort 1: n = 51, cohort 2: n = 75)

Baseline measures:
FRA Letter Sounds, Phonological Awareness, Vocabulary Pairs, Following Directions, and Sentence Comprehension

Outcome measures:
FRA Phonological Awareness, Word Reading, Vocabulary Pairs, Following Directions, and Sentence Comprehension and SESAT Word Reading and Sentence Reading

Students with missing data on outcome measures were included in the associated impact analyses.

All baseline measures were used as covariates for all outcome measures.

FRA is Florida Center for Reading Research Reading Assessment. SESAT is the Stanford Early Scholastic Achievement Test.

Source: Authors' analysis based on data from participating districts in Florida.

Figure A2. Grade 1 student and school consolidated standards of reporting trials

Enrollment

Total participating schools: *n* = 55
· Cohort 1 (2013/14 school year): *n* = 27 schools from four districts (three regions)
· Cohort 2 (2014/15 school year): *n* = 28 schools from three districts (three regions)

Total eligible students with parent consent: *n* = 1,098
· Cohort 1: *n* = 506, Cohort 2: *n* = 592

Schools were randomly assigned to standalone or embedded intervention within cohort and region

Allocation

Standalone intervention

Schools: *n* = 27
(cohort 1: *n* = 14, cohort 2: *n* = 13)

Total students within schools: *n* = 534
(cohort 1: *n* = 267, cohort 2: *n* = 267)

Total small groups within schools: *n* = 129
(cohort 1: *n* = 64, cohort 2: *n* = 65)

Embedded intervention

Schools: *n* = 28
(cohort 1: *n* = 13, cohort 2: *n* = 15)

Total students within schools: *n* = 564
(cohort 1: *n* = 239, cohort 2: *n* = 325)

Total small groups within schools: *n* = 137
(cohort 1: *n* = 59, cohort 2: *n* = 78)

Follow-up

Missing outcome data:
· Schools: *n* = 0
· Students: *n* = 63
  (cohort 1: *n* = 33, cohort 2: *n* = 30)

Missing outcome data:
· Schools: *n* = 0
· Students: *n* = 57
  (cohort 1: *n* = 25, cohort 2: *n* = 32)

Data collected and analyzed

Analytic sample:
· Schools: *n* = 27
  (cohort 1: *n* = 14, cohort 2: *n* = 13)
· Students: *n* = 471
  (cohort 1: *n* = 234, cohort 2: *n* = 237)
· Groups: *n* = 129
  (cohort 1: *n* = 64, cohort 2: *n* = 65)

Analytic sample:
· Schools: *n* = 28
  (cohort 1: *n* = 13, cohort 2: *n* = 15)
· Students: *n* = 507
  (cohort 1: *n* = 214, cohort 2: *n* = 293)
· Groups: *n* = 137
  (cohort 1: *n* = 59, cohort 2: *n* = 78)

Baseline measures:
FRA Word Reading, Vocabulary Pairs, Following Directions, and Sentence Comprehension

Outcome measures:
FRA Word Reading, Vocabulary Pairs, Following Directions, and Sentence Comprehension and SAT-10 Reading Comprehension

Students with missing data on outcome measures were included in the associated impact analyses.

All baseline measures were used as covariates for all outcome measures.

FRA is Florida Center for Reading Research Reading Assessment. SAT-10 is the Stanford Achievement Test, 10th edition.

Source: Authors' analysis based on data from participating districts in Florida.

Figure A3. Grade 2 student and school consolidated standards of reporting trials

Enrollment

Total participating schools: $n = 55$
- Cohort 1 (2013/14 school year): $n = 27$ schools from four districts (three regions)
- Cohort 2 (2014/15 school year): $n = 28$ schools from three districts (three regions)

Total eligible students with parent consent: $n = 1,309$
- Cohort 1: $n = 624$, Cohort 2: $n = 685$

Schools were randomly assigned to standalone or embedded intervention within cohort and region

Allocation

Standalone intervention

Schools: $n = 27$
(cohort 1: $n = 14$, cohort 2: $n = 13$)

Total students within schools: $n = 639$
(cohort 1: $n = 323$, cohort 2: $n = 316$)

Total small groups within schools: $n = 134$
(cohort 1: $n = 70$, cohort 2: $n = 64$)

Embedded intervention

Schools: $n = 28$
(cohort 1: $n = 13$, cohort 2: $n = 15$)

Total students within schools: $n = 670$
(cohort 1: $n = 301$, cohort 2: $n = 369$)

Total small groups within schools: $n = 137$
(cohort 1: $n = 63$, cohort 2: $n = 74$)

Follow-up

Withdrew from study:
- Schools: $n = 1$ (from cohort 2, causing withdrawal of 21 students)[a]

Missing outcome data:
- Students: $n = 70$
(cohort 1: $n = 34$, cohort 2: $n = 36$)

Missing outcome data:
- Schools: $n = 0$
- Students: $n = 85$
(cohort 1: $n = 38$, cohort 2: $n = 47$)

Data collected and analyzed

Analytic sample:
- Schools: $n = 26$
(cohort 1: $n = 14$, cohort 2: $n = 12$)
- Students: $n = 548$
(cohort 1: $n = 289$, cohort 2: $n = 259$)
- Groups: $n = 130$
(cohort 1: $n = 70$, cohort 2: $n = 60$)

Analytic sample:
- Schools: $n = 28$
(cohort 1: $n = 13$, cohort 2: $n = 15$)
- Students: $n = 585$
(cohort 1: $n = 263$, cohort 2: $n = 322$)
- Groups: $n = 137$
(cohort 1: $n = 63$, cohort 2: $n = 74$)

Baseline measures:
FRA Word Reading, Vocabulary Pairs, Following Directions, and Spelling

Outcome measures:
FRA Word Reading, Vocabulary Pairs, Following Directions, Sentence Comprehension, and Spelling and SAT-10 Reading Comprehension

Students with missing data on outcome measures were included in the associated impact analyses.

FRA is Florida Center for Reading Research Reading Assessment. SAT-10 is the Stanford Achievement Test, 10th edition.

a. One of the standalone intervention schools in cohort 2 was excluded from the grade 2 analyses because scheduling conflicts resulted in the withdrawal of the 21 participating grade 2 students at that school.

Source: Authors' analysis based on data from participating districts in Florida.

Intervention and interventionists

An independent subcontractor reviewed the levels of evidence on the What Works Clearinghouse (WWC) website for reading interventions that had been studied with at-risk students in grades K–2 and implemented in small groups. The reading intervention program that met these criteria and had the strongest levels of evidence in alphabetics, fluency, and comprehension was Sound Partners (Vadasy & Sanders, 2012; Vadasy, Sanders, & Abbott, 2008; Vadasy, Sanders, & Peyton, 2006; Vadasy et al., 2004). Sound Partners consists of a kindergarten book and a combined book for grades 1 and 2. No academic language intervention programs for at-risk students in grades K–2 have been rated by the WWC, so a vocabulary program with good clinical evidence, Bridge of Vocabulary (Montgomery, 2007), and an inferential language program with evidence of efficacy,

Language in Motion (Phillips, 2014), were added to Sound Partners to create the standalone intervention. The reading and language components of the standalone intervention each had an implementation manual developed by the authors.

The most widely adopted core reading program in Florida at the start of this study was Houghton Mifflin Harcourt (HMH) Journeys, and this was the core curriculum used in all study schools. Therefore, the embedded intervention in this study consisted of the tier 2 program Strategic Intervention and the supplementary vocabulary piece Curious about Words, both of which are part of HMH Journeys. Because Strategic Intervention and Curious about Words came to the schools in shrink-wrapped packages without implementation manuals, Regional Educational Laboratory (REL) Southeast staff developed a manual for each, which included information about scope and sequence and instructional procedures.

Sound Partners is similar to Strategic Intervention in the alphabetic skills taught, but Strategic Intervention also includes vocabulary and comprehension skills. Another difference is that in Sound Partners the progression of lessons depends on students' mastery of content, as reflected in the corresponding skill assessments, and remediation targets lessons that included concepts students had not mastered, whereas HMH Journeys has no specific provision for remediation, though weaknesses in students' mastery of content based on skills assessment are noted, and instruction on those skills is emphasized in future lessons. Bridge of Vocabulary focuses on building oral vocabulary and concepts using manipulatives and discussion, and Language in Motion uses science-based manipulatives to build oral language components of syntax, inferential language, and listening comprehension (see figure 1 in the main text). In contrast, Curious about Words is based on Beck, McKeown, and Kucan's (2013) strategies for teaching vocabulary words embedded in challenging text read aloud by a teacher.

Both interventions were taught daily from mid-October to the end of May for 45 minutes and consisted of a 25–30 minute reading component and a 15-minute oral language component. In standalone intervention schools, Sound Partners (reading component) was taught daily, Bridge of Vocabulary (oral language component) was taught three times a week, and Language in Motion (oral language component) twice a week. Both interventions were taught in groups of four students in kindergarten and grade 1 and in groups of five students in grade 2.

As an incentive to participate in the study, REL Southeast hired two to three interventionists per school but encouraged school leaders to contribute paraprofessionals as interventionists in order to serve more at-risk students and build capacity at the school for intervention to continue after the study ended. For cohort 1, REL Southeast provided 66 interventionists, schools provided 17 paraprofessionals, and together they served 370 small groups; 32 percent of the interventionists were certified teachers. For cohort 2, REL Southeast provided 64 interventionists (42 percent of whom had been interventionists for cohort 1 schools), schools provided 25 paraprofessionals, and together they served 424 small groups; 37 percent of the interventionists were certified teachers. On average, each school had three to four interventionists, who each served four to six small groups.

Interventionists had some experience working with young children in education settings. Each year REL Southeast staff trained the interventionists over a two-day period during

late September and sent them home with the manuals and instructional materials they would be using to familiarize themselves with the strategies, materials, and corresponding skill assessments. During early October the interventionists visited their assigned school to meet the grade K–2 teachers and school staff and set up materials in their intervention space. Once the intervention started in mid-October, REL Southeast staff visited each interventionist to answer questions and to provide additional training, if needed. A lead interventionist was designated at each school to communicate with school leadership and REL Southeast staff. In addition, interventionists audio-recorded one week of lessons each month for periodic review by REL Southeast staff. The audio-recordings were referred to occasionally in discussions of student behavior.

Implementation fidelity, coverage, and student attendance

REL Southeast staff observed all small groups twice a year, once in the fall and once in the spring, and completed a fidelity checklist. Fidelity is defined as the percentage of the lesson in which instruction followed the lesson sequence and script for each of the skills taught. REL Southeast staff members were trained to achieve better than 80 percent reliability on the checklist. Inter-rater reliability was evaluated during the observations on 15 percent of the checklists.

Separate fidelity ratings were calculated for the reading and oral language components in both the fall and spring, resulting in four fidelity ratings for each small group (fall reading, fall oral language, spring reading, and spring oral language). For each small group the fall and spring fidelity ratings were then averaged to create separate overall fidelity ratings for the reading and oral language components. Fidelity was considered high if it was 80 percent or higher.

In both the standalone and embedded interventions, interventionists implemented instruction with high fidelity (table A3). In grades K–2, 72–91 percent of small groups in cohorts 1 and 2 combined demonstrated at least 80 percent fidelity on the reading and oral language components in the two intervention groups. The median overall fidelity across interventions was 96 percent in kindergarten, 94 percent in grade 1, and 96 percent in grade 2.

Across grades K–2, interventionists in standalone intervention schools covered 55–80 percent of the reading curricula (80 percent in kindergarten, 55 percent in grade 1, and 62 percent in grade 2) and 77–79 percent of the oral language curricula (79 percent in kindergarten, 78 percent in grade 1, and 77 percent in grade 2) for cohorts 1 and 2 combined (table A4). Interventionists in embedded intervention schools covered 86–88 percent of the reading and oral language curricula for cohorts 1 and 2 combined. By cohort, interventionists covered 53–84 percent of the reading and oral language curricula in standalone intervention schools (72–84 percent for the oral language curricula alone) and 83–90 percent of the reading and oral language curricula in embedded intervention schools (table A4). In grades 1 and 2 coverage of the reading component in the standalone intervention was 15–23 percentage points lower than coverage of the oral language component. The difference is likely due to the requirements for skill mastery in Sound Partners. Intervention groups across grades 1 and 2 required, on average, remediation on 8–11 out of a possible 30 skill assessments in Sound Partners. When remediation occurred, it was because an average of 45–59 percent of the intervention group had not demonstrated

Table A3. Implementation fidelity, by grade, component, intervention group, and cohort, 2013/14 and 2014/15

Grade and intervention component	Standalone intervention[a]							Embedded intervention						
	Cohort 1		Cohort 2		Cohorts 1 and 2 combined			Cohort 1		Cohort 2		Cohorts 1 and 2 combined		
	Number of small groups	Mean percentage (standard deviation)	Number of small groups	Mean percentage (standard deviation)	Number of small groups	Mean percentage (standard deviation)	Percentage of small groups with 80 percent or higher fidelity	Number of small groups	Mean percentage (standard deviation)	Number of small groups	Mean percentage (standard deviation)	Number of small groups	Mean percentage (standard deviation)	Percentage of small groups with 80 percent or higher fidelity
Kindergarten														
Reading	59	93 (9)	68	88 (16)	127	91 (13)	87	51	97 (7)	76	88 (13)	127	92 (11)	87
Oral language	59	86 (16)	68	94 (9)	127	90 (13)	82	51	87 (26)	76	89 (23)	127	88 (24)	78
Bridge of Vocabulary	59	85 (21)	68	94 (11)	127	90 (17)	86	na	na	na	na	na	na	na
Language in Motion	59	87 (21)	68	94 (13)	127	91 (17)	80	na	na	na	na	na	na	na
Grade 1														
Reading	67	94 (8)	65	91 (14)	132	93 (11)	89	58	96 (8)	77	89 (12)	135	92 (11)	89
Oral language	67	87 (16)	65	95 (8)	132	91 (14)	84	58	87 (27)	77	94 (15)	135	91 (22)	83
Bridge of Vocabulary	67	87 (17)	65	96 (10)	132	91 (14)	83	na	na	na	na	na	na	na
Language in Motion	67	87 (20)	65	95 (11)	132	91 (16)	79	na	na	na	na	na	na	na
Grade 2														
Reading	69	93 (12)	63	89 (17)	132	92 (15)	87	64	97 (7)	73	87 (12)	137	92 (11)	91
Oral language	69	89 (13)	63	93 (11)	132	91 (12)	81	64	84 (29)	73	85 (24)	137	85 (26)	72
Bridge of Vocabulary	69	90 (13)	63	93 (12)	132	91 (13)	83	na	na	na	na	na	na	na
Language in Motion	69	89 (18)	63	93 (15)	132	91 (17)	81	na	na	na	na	na	na	na

na is not applicable.

a. The standalone intervention included two oral language components (Bridge of Vocabulary and Language in Motion) that were observed on separate days in both the fall and spring. The ratings were then averaged to determine overall oral language fidelity for the standalone intervention.

Source: Authors' analysis of data collected by Regional Educational Laboratory Southeast staff during fidelity observations.

Table A4. Percentage of reading and oral language content covered, by grade, intervention group, and cohort, 2013/14 and 2014/15

Grade and intervention component	Standalone intervention[a]						Embedded intervention					
	Cohort 1		Cohort 2		Cohorts 1 and 2 combined		Cohort 1		Cohort 2		Cohorts 1 and 2 combined	
	Number of groups	Mean percentage (standard deviation)	Number of groups	Mean percentage (standard deviation)	Number of groups	Mean percentage (standard deviation)	Number of groups	Mean percentage (standard deviation)	Number of groups	Mean percentage (standard deviation)	Number of groups	Mean percentage (standard deviation)
Kindergarten												
Reading	63	80 (16)	68	81 (15)	131	80 (16)	51	87 (3)	75	89 (10)	126	88 (8)
Oral language	63	73 (9)	68	84 (7)	131	79 (10)	51	86 (3)	75	88 (10)	126	87 (8)
Bridge of Vocabulary	63	73 (9)	68	84 (7)	131	79 (10)	na	na	na	na	na	na
Language in Motion	63	72 (9)	68	84 (7)	131	79 (10)	na	na	na	na	na	na
Grade 1												
Reading	64	58 (12)	65	53 (7)	129	55 (10)	59	86 (3)	78	89 (10)	137	88 (8)
Oral Language	64	72 (7)	65	84 (7)	129	78 (9)	59	86 (3)	78	89 (10)	137	88 (8)
Bridge of Vocabulary	64	72 (7)	65	84 (7)	129	78 (9)	na	na	na	na	na	na
Language in Motion	64	72 (7)	65	84 (7)	129	78 (9)	na	na	na	na	na	na
Grade 2												
Reading	67	66 (10)	63	58 (10)	133	62 (11)	63	83 (15)	74	90 (8)	137	87 (12)
Oral language	67	73 (7)	63	81 (13)	133	77 (11)	63	83 (15)	74	89 (8)	137	86 (12)
Bridge of Vocabulary	67	73 (8)	63	81 (13)	133	77 (11)	na	na	na	na	na	na
Language in Motion	67	73 (7)	63	81 (13)	133	77 (11)	na	na	na	na	na	na

na is not applicable.

a. The standalone intervention included two oral language components (Bridge of Vocabulary and Language in Motion) that were observed on separate days in both the fall and spring. The ratings were then averaged to determine overall oral language fidelity for the standalone intervention.

Source: Authors' analysis of data collected by Regional Educational Laboratory Southeast staff during daily intervention implementation.

mastery on the skill assessment. This means that in groups of four or five students, one or two students received potentially unnecessary remediation. It is likely that group remediation disadvantages some students while benefiting others.

Interventionists recorded student attendance daily. Attendance reflects the total number of intervention sessions a student attended. If a student was present at school but did not attend intervention for any reason, the student was marked absent from intervention. In total, students could have attended approximately 134 days of instruction. Across grades K–2 the average number of days of intervention attended for cohorts 1 and 2 combined was 92–95 among students in standalone intervention schools and 96–98 among students in embedded intervention schools (table A5). By cohort, students in standalone intervention schools attended on average 89–100 days of intervention, and students in embedded intervention schools attended 94–99 days of intervention (table A5). The higher average attendance rates observed for cohort 2 across grades is likely due to increased flexibility in intervention scheduling around school events and holidays.

Measures

The study included reading and language measures from the FRA, the Stanford Early Scholastic Achievement Test (SESAT), and the Stanford Achievement Test, 10th edition (SAT-10). Reading outcomes included the Phonological Awareness (kindergarten only), Word Reading, and Spelling (grade 2 only) subtests from the FRA (table A6) and the Word Reading subtest from SESAT in kindergarten. Language outcomes included the Vocabulary Pairs, Following Directions, and Sentence Comprehension subtests from the FRA (see table A6), the Sentence Reading subtest from the SESAT in kindergarten, and the Reading Comprehension subtest from the SAT-10 in grades 1 and 2. Although the FRA Sentence Comprehension subtest was administered to K–2 students, it is a kindergarten-normed subtest, which means that the percentile ranks for all grades reflect ability on a kindergarten scale.

All measures were assessed at baseline except FRA Word Reading in kindergarten, Sentence Comprehension in grade 2, SESAT Word Reading and Sentence Reading in kindergarten, and SAT-10 Reading Comprehension in grades 1 and 2. In addition, FRA Letter Sounds was assessed only at baseline in kindergarten (see table A6).

The FRA is a computer-adaptive screening assessment of reading and language for students in K–2. The FRA was developed under federal grants to Florida State University (Foorman, Petscher, & Schatschneider, 2015) and normed on Florida students. In all of the FRA subtests, students receive five items at grade level and then the system adapts up or down based on performance to reach a precise estimate of a student's ability. The marginal reliability (Sireci, Thissen, & Wainer, 1991) for the FRA subtests based on the normative sample ranges from .85 to .96 across grades K–2. Students are given a developmental ability score on each subtest that has a mean of 500 and a standard deviation of 100.

The SESAT and SAT-10 are norm-referenced reading tests. Reliability is .85 for the SESAT Word Reading subtest and .88 for the SESAT Sentence Reading subtest. Reliability for the SAT-10 Reading Comprehension subtest is .91 for grades 1 and 2. Scaled scores from the SESAT and SAT-10 were used in all analyses.

Table A5. Number of intervention days attended by grade, intervention group, and cohort, 2013/14 and 2014/15

	Standalone intervention						Embedded intervention					
	Cohort 1		Cohort 2		Cohorts 1 and 2 combined		Cohort 1		Cohort 2		Cohorts 1 and 2 combined	
Grade	Number of students	Mean number of days (standard deviation)	Number of students	Mean number of days (standard deviation)	Number of students	Mean number of days (standard deviation)	Number of students	Mean number of days (standard deviation)	Number of students	Mean number of days (standard deviation)	Number of students	Mean number of days (standard deviation)
Kindergarten	255	89 (20)	276	100 (20)	531	95 (21)	213	97 (18)	317	98 (23)	530	97 (21)
Grade 1	267	90 (19)	267	100 (19)	534	95 (20)	239	96 (17)	325	99 (22)	564	98 (20)
Grade 2	323	89 (21)	295	95 (24)	639	92 (22)	301	94 (21)	369	98 (23)	670	96 (22)

Note: Attendance represents the total number of days a student was present for intervention. If a student was present at school but did not participate in intervention, the student was marked absent from intervention.

Source: Authors' analysis of data collected by Regional Educational Laboratory Southeast staff during daily intervention implementation.

Table A6. Florida Center for Reading Research Reading Assessment subtests, by grade and assessment period

Subtest	Kindergarten		Grades 1 and 2		Subtest description
	Baseline	Outcome	Baseline	Outcome	
Phonological Awareness	✔	✔			Students listen to a word that has been broken into parts and then blend them back together to reproduce the word.
Letter Sounds	✔				A letter is presented on the monitor in upper and lower case and students provide the sound it makes.
Vocabulary Pairs	✔	✔	✔	✔	Three words appear on the monitor and are pronounced by the computer. The student selects the two words that go together best (for example, dark, night, swim).
Following Directions	✔	✔	✔	✔	Students listen and then click and drag objects in response to the computer's directions (for example, put the square in front of the chair and then put the circle behind the chair).
Sentence Comprehension[a]	✔	✔	✔	✔	Students listen to a sentence given by a computer (for example, click on the picture of the bird flying towards the nest) and then select the one picture out of the four presented on the monitor that depicts the sentence.
Word Reading		✔	✔	✔	Words of varying difficulty are presented on the monitor one at a time and students read them aloud.
Spelling[b]			✔	✔	The computer provides each word and uses it in a sentence. Students respond by using the computer keyboard to spell the word.

a. Administered at baseline only to kindergarten and grade 1 students.

b. Administered only to grade 2 students.

Note: Tasks were administered to individual students. Baseline testing occurred in September or October; outcome testing occurred in April or May.

Source: Authors' compilation based on tasks included in the computer-adaptive K–2 Florida Center for Reading Research Reading Assessment.

Student-level FRA baseline scores by grade were used to create small group–level and school-level baseline scores that were used in the analyses as small group–level and school-level covariates. School-level and student-level differences in baseline scores between standalone and embedded intervention schools for the baseline sample by grade and cohort were estimated for all baseline measures (tables A7–A10). The majority of differences in baseline scores between the two interventions for the baseline sample were determined to be nonsignificant across grades and cohorts at the school and student levels, except for FRA Word Reading in grade 1 cohort 1, grade 1 full sample, and grade 2 cohort 1; FRA Following Directions in kindergarten cohort 1; and FRA Vocabulary Pairs in grade 2 cohort 2 (tables A8 and A10). FRA baseline scores at the small group level are reported in table A11 by grade, intervention group, and cohort, but differences in baseline scores were not estimated at the small group level because this level did not serve as the unit of assignment or analysis.

FRA baseline percentile rank, FRA outcome percentile rank, and difference between baseline and outcome as well as percentile ranks for the SESAT and SAT-10 outcomes for the analytic sample are reported in table A12.

Table A7. Preintervention school-level sample sizes and characteristics for the baseline and analytic samples, by grade, cohort, and intervention group, 2013/14 and 2014/15

| Grade, cohort and sample, and baseline measure | Standalone intervention | | | | Embedded intervention | | | |
| | Sample size | | Sample characteristics | | Sample size | | Sample characteristics | |
	Unit of assignment	Unit of analysis	Mean (adjusted mean)	Standard deviation	Unit of assignment	Unit of analysis	Mean (adjusted mean)	Standard deviation
Kindergarten								
Cohort 1, baseline and analytic samples								
FRA Letter Sounds	14	14	337 (338)	32	13	13	355 (354)	57
FRA Phonological Awareness	14	14	313 (314)	31	13	13	293 (293)	36
FRA Vocabulary Pairs	14	14	356 (356)	25	13	13	356 (356)	48
FRA Following Directions	14	14	284 (283)	46	13	13	238 (239)	42
FRA Sentence Comprehension	14	14	408 (408)	33	13	13	394 (395)	29
Cohort 2, baseline and analytic samples								
FRA Letter Sounds	13	13	275 (274)	42	15	15	267 (268)	29
FRA Phonological Awareness	13	13	244 (245)	32	15	15	251 (251)	23
FRA Vocabulary Pairs	13	13	330 (331)	30	15	15	331 (330)	21
FRA Following Directions	13	13	229 (231)	43	15	15	231 (229)	56
FRA Sentence Comprehension	13	13	389 (390)	29	15	15	397 (396)	32
Cohorts 1 and 2 combined, baseline and analytic samples								
FRA Letter Sounds	27	27	307 (307)	48	28	28	308 (308)	62
FRA Phonological Awareness	27	27	280 (280)	47	28	28	271 (271)	36
FRA Vocabulary Pairs	27	27	340 (344)	30	28	28	343 (343)	38
FRA Following Directions	27	27	258 (258)	52	28	28	234 (234)	50
FRA Sentence Comprehension	27	27	399 (399)	32	28	28	396 (396)	30
Grade 1								
Cohort 1, baseline and analytic samples								
FRA Word Reading	14	14	259 (257)	60	13	13	341 (343)	95
FRA Vocabulary Pairs	14	14	413 (413)	16	13	13	423 (424)	18
FRA Following Directions	14	14	384 (383)	64	13	13	413 (414)	57
FRA Sentence Comprehension	14	14	464 (464)	41	13	13	450 (450)	31

(continued)

Table A7. Preintervention school-level sample sizes and characteristics for the baseline and analytic samples, by grade, cohort, and intervention group, 2013/14 and 2014/15 *(continued)*

Grade, cohort and sample, and baseline measure	Standalone intervention				Embedded intervention			
	Sample size		Sample characteristics		Sample size		Sample characteristics	
	Unit of assignment	Unit of analysis	Mean (adjusted mean)	Standard deviation	Unit of assignment	Unit of analysis	Mean (adjusted mean)	Standard deviation
Cohort 2, baseline and analytic samples								
FRA Word Reading	13	13	215 (220)	76	15	15	247 (243)	62
FRA Vocabulary Pairs	13	13	401 (403)	27	15	15	390 (389)	34
FRA Following Directions	13	13	385 (389)	55	15	15	376 (373)	57
FRA Sentence Comprehension	13	13	461 (462)	25	15	15	462 (461)	31
Cohorts 1 and 2 combined, baseline and analytic samples								
FRA Word Reading	27	27	238 (238)	71	28	28	291 (290)	91
FRA Vocabulary Pairs	27	27	408 (408)	23	28	28	408 (405)	32
FRA Following Directions	27	27	385 (385)	59	28	28	393 (393)	59
FRA Sentence Comprehension	27	27	462 (462)	33	28	28	456 (456)	31
Grade 2								
Cohort 1, baseline and analytic samples								
FRA Word Reading	14	14	471 (470)	36	13	13	520 (522)	73
FRA Spelling	14	14	351 (351)	34	13	13	356 (357)	46
FRA Vocabulary Pairs	14	14	497 (497)	25	13	13	502 (502)	35
FRA Following Directions	14	14	529 (528)	30	13	13	527 (528)	47
Cohort 2, baseline sample								
FRA Word Reading	13	13	442 (444)	46	15	15	440 (438)	51
FRA Spelling	13	13	310 (313)	48	15	15	319 (317)	46
FRA Vocabulary Pairs	13	13	496 (498)	19	15	15	468 (467)	32
FRA Following Directions	13	13	476 (479)	56	15	15	467 (464)	65
Cohort 2, analytic sample								
FRA Word Reading	13	12[a]	447 (449)	44	15	15	440 (439)	51
FRA Spelling	13	12[a]	317 (319)	43	15	15	319 (318)	46
FRA Vocabulary Pairs	13	12[a]	498 (499)	19	15	15	468 (468)	32
FRA Following Directions	13	12[a]	481 (483)	55	15	15	467 (465)	65
Cohorts 1 and 2 combined, baseline sample								
FRA Word Reading	27	27	457 (457)	43	28	28	477 (477)	73
FRA Spelling	27	27	332 (332)	46	28	28	336 (336)	49
FRA Vocabulary Pairs	27	27	497 (497)	22	28	28	484 (484)	37
FRA Following Directions	27	27	503 (504)	51	28	28	495 (495)	64
Cohorts 1 and 2 combined, analytic sample								
FRA Word Reading	27	26[a]	460 (460)	41	28	28	477 (478)	73
FRA Spelling	27	26[a]	336 (336)	41	28	28	336 (337)	49
FRA Vocabulary Pairs	27	26[a]	497 (497)	22	28	28	484 (484)	37
FRA Following Directions	27	26[a]	507 (507)	49	28	28	495 (495)	64

FRA is Florida Center for Reading Research Reading Assessment.

a. Does not include the 21 students from the one standalone intervention school that removed all grade 2 students from the intervention.

Note: A regression model with a dichotomous indicator for treatment (the embedded intervention served as the referent) was used to test for baseline equivalence between students in standalone and embedded intervention schools. Because random assignment was conducted separately within geographic region, the model also included region as a covariate when estimating the adjusted means.

Source: Authors' analysis based on data from participating districts in Florida.

Table A8. School-level baseline scores for the baseline and analytic samples, by grade, cohort, and intervention group, 2013/14 and 2014/15

Grade, cohort and sample, and baseline measure	Adjusted mean		Difference (standard error)	p value	Effect size	School sample size
	Standalone intervention (standard deviation)	Embedded intervention (standard deviation)				
Kindergarten						
Cohort 1, baseline and analytic samples						
FRA Letter Sounds	338 (32)	354 (57)	–16 (17)	.37	–0.34	27
FRA Phonological Awareness	314 (31)	293 (36)	21 (13)	.13	0.61	27
FRA Vocabulary Pairs	356 (25)	356 (48)	0 (15)	.99	0.00	27
FRA Following Directions	283 (46)	239 (42)	44 (17)	.01	0.97	27
FRA Sentence Comprehension	408 (33)	395 (29)	13 (12)	.30	0.40	27
Cohort 2, baseline and analytic samples						
FRA Letter Sounds	274 (42)	268 (29)	6 (13)	.63	0.16	28
FRA Phonological Awareness	245 (32)	251 (23)	–6 (11)	.59	–0.21	28
FRA Vocabulary Pairs	331 (30)	330 (21)	1 (10)	.96	0.04	28
FRA Following Directions	231 (43)	229 (56)	2 (18)	.91	0.04	28
FRA Sentence Comprehension	390 (29)	396 (32)	–6 (11)	.63	–0.19	28
Cohorts 1 and 2 combined, baseline and analytic samples						
FRA Letter Sounds	307 (48)	308 (62)	–1 (15)	.97	–0.02	55
FRA Phonological Awareness	280 (47)	271 (36)	9 (11)	.38	0.21	55
FRA Vocabulary Pairs	344 (30)	343 (38)	1 (9)	.91	0.03	55
FRA Following Directions	258 (52)	234 (50)	24 (13)	.07	0.46	55
FRA Sentence Comprehension	399 (32)	396 (30)	3 (8)	.68	0.10	55
Grade 1						
Cohort 1, baseline and analytic samples						
FRA Word Reading	257 (60)	343 (95)	–86 (28)	.006	–1.0	27
FRA Vocabulary Pairs	413 (16)	424 (18)	–11 (7)	.14	–0.63	27
FRA Following Directions	383 (64)	414 (57)	–31 (24)	.22	–0.49	27
FRA Sentence Comprehension	464 (41)	450 (31)	14 (14)	.36	0.37	27
Cohort 2, baseline and analytic samples						
FRA Word Reading	220 (76)	243 (62)	–23 (19)	.23	–0.32	28
FRA Vocabulary Pairs	403 (27)	389 (34)	14 (11)	.20	0.44	28
FRA Following Directions	389 (55)	373 (57)	16 (17)	.37	0.28	28
FRA Sentence Comprehension	462 (25)	461 (31)	1 (10)	.91	0.03	28
Cohorts 1 and 2 combined, baseline and analytic samples						
FRA Word Reading	238 (71)	290 (91)	–52 (20)	.01	–0.63	55
FRA Vocabulary Pairs	408 (23)	405 (32)	3 (7)	.76	0.11	55
FRA Following Directions	385 (59)	393 (59)	–8 (15)	.59	–0.13	55
FRA Sentence Comprehension	462 (33)	456 (31)	6 (8)	.47	0.18	55
Grade 2						
Cohort 1, baseline and analytic samples						
FRA Word Reading	470 (36)	522 (73)	–52 (21)	.02	–0.89	27
FRA Spelling	351 (34)	357 (46)	–6 (15)	.67	–0.14	27
FRA Vocabulary Pairs	497 (25)	502 (35)	–5 (12)	.65	–0.16	27
FRA Following Directions	528 (30)	528 (47)	0 (14)	.98	0.00	27

(continued)

Table A8. School-level baseline scores for the baseline and analytic samples, by grade, cohort, and intervention group, 2013/14 and 2014/15 (continued)

Grade, cohort and sample, and baseline measure	Adjusted mean		Difference (standard error)	p value	Effect size	School sample size
	Standalone intervention (standard deviation)	Embedded intervention (standard deviation)				
Cohort 2, baseline sample						
FRA Word Reading	444 (46)	438 (51)	6 (16)	.70	0.12	28
FRA Spelling	313 (48)	317 (46)	−4 (16)	.79	−0.08	28
FRA Vocabulary Pairs	498 (19)	467 (32)	31 (9)	.003	1.12	28
FRA Following Directions	479 (56)	464 (65)	15 (20)	.44	0.24	28
Cohort 2, analytic sample						
FRA Word Reading	449 (44)	439 (51)	10 (16)	.56	0.20	27
FRA Spelling	319 (43)	318 (46)	1 (15)	.94	0.02	27
FRA Vocabulary Pairs	499 (19)	468 (32)	31 (10)	.003	1.11	27
FRA Following Directions	483 (55)	465 (65)	18 (20)	.37	0.29	27
Cohorts 1 and 2 combined, baseline sample						
FRA Word Reading	457 (43)	477 (73)	−20 (15)	.20	−0.33	55
FRA Spelling	332 (46)	336 (49)	−4 (12)	.72	−0.08	55
FRA Vocabulary Pairs	497 (22)	484 (37)	13 (8)	.12	0.42	55
FRA Following Directions	504 (51)	495 (64)	9 (14)	.53	0.15	55
Cohorts 1 and 2 combined, analytic sample						
FRA Word Reading	460 (41)	478 (73)	−18 (16)	.26	−0.30	54
FRA Spelling	336 (41)	337 (49)	−1 (12)	.94	−0.02	54
FRA Vocabulary Pairs	497 (22)	484 (37)	13 (8)	.11	0.42	54
FRA Following Directions	507 (49)	495 (64)	12 (14)	.42	0.21	54

FRA is Florida Center for Reading Research Reading Assessment.

Note: A regression model with a dichotomous indicator for treatment (the embedded intervention served as the referent) was used to test for baseline equivalence between students in standalone and embedded intervention schools. Because random assignment was conducted separately within geographic region, the model also included region as a covariate when estimating the adjusted means.

Source: Authors' analysis based on data from participating districts in Florida.

Table A9. Preintervention student-level sample sizes and characteristics for the baseline and analytic samples, by grade, cohort, and intervention group, 2013/14 and 2014/15

Grade, cohort and sample, and baseline measure	Standalone intervention				Embedded intervention			
	Sample size		Sample characteristics		Sample size		Sample characteristics	
	Unit of assignment	Unit of analysis	Mean (adjusted mean)	Standard deviation	Unit of assignment	Unit of analysis	Mean (adjusted mean)	Standard deviation
Kindergarten								
Cohort 1, baseline and analytic samples								
FRA Letter Sounds	255	255	336 (335)	91	213	213	350 (347)	106
FRA Phonological Awareness	255	255	312 (312)	113	213	213	292 (292)	122
FRA Vocabulary Pairs	255	255	363 (364)	57	213	213	367 (366)	60
FRA Following Directions	255	255	287 (286)	127	213	213	241 (243)	156
FRA Sentence Comprehension	255	255	409 (409)	90	213	213	397 (397)	92
Cohort 2, baseline and analytic samples								
FRA Letter Sounds	276	276	272 (272)	99	317	317	268 (267)	102
FRA Phonological Awareness	276	276	244 (246)	84	317	317	254 (253)	86
FRA Vocabulary Pairs	276	276	338 (339)	65	317	317	335 (335)	65
FRA Following Directions	276	276	236 (237)	154	317	317	234 (233)	149
FRA Sentence Comprehension	276	276	396 (398)	85	317	317	399 (399)	82
Cohorts 1 and 2 combined, baseline and analytic samples								
FRA Letter Sounds	531	531	303 (304)	100	530	530	301 (303)	111
FRA Phonological Awareness	531	531	277 (279)	105	530	530	269 (271)	103
FRA Vocabulary Pairs	531	531	350 (351)	63	530	530	348 (349)	64
FRA Following Directions	531	531	260 (261)	144	530	530	237 (237)	152
FRA Sentence Comprehension	531	531	402 (403)	87	530	530	398 (399)	86
Grade 1								
Cohort 1, baseline and analytic samples								
FRA Word Reading	267	267	251 (257)	186	239	239	342 (342)	205
FRA Vocabulary Pairs	267	267	412 (412)	62	239	239	425 (425)	65
FRA Following Directions	267	267	408 (408)	115	239	239	431 (431)	119
FRA Sentence Comprehension	267	267	461 (462)	129	239	239	452 (451)	125
Cohort 2, baseline and analytic samples								
FRA Word Reading	267	267	227 (231)	139	325	325	254 (253)	148
FRA Vocabulary Pairs	267	267	407 (409)	71	325	325	400 (399)	74
FRA Following Directions	267	267	394 (398)	116	325	325	387 (386)	124
FRA Sentence Comprehension	267	267	464 (465)	70	325	325	464 (464)	77
Cohorts 1 and 2 combined, baseline and analytic samples								
FRA Word Reading	534	534	239 (243)	165	564	564	291 (293)	179
FRA Vocabulary Pairs	534	534	410 (411)	67	564	564	410 (410)	71
FRA Following Directions	534	534	401 (402)	115	564	564	406 (407)	124
FRA Sentence Comprehension	534	534	462 (463)	104	564	564	459 (458)	100
Grade 2								
Cohort 1, baseline and analytic samples								
FRA Word Reading	323	323	476 (478)	72	301	301	525 (521)	94
FRA Spelling	323	323	350 (351)	106	301	301	357 (357)	107
FRA Vocabulary Pairs	323	323	496 (497)	82	301	301	504 (503)	87
FRA Following Directions	323	323	533 (535)	113	301	301	536 (533)	103

(continued)

Table A9. Preintervention student-level sample sizes and characteristics for the baseline and analytic samples, by grade, cohort, and intervention group, 2013/14 and 2014/15 (continued)

| Grade, cohort and sample, and baseline measure | Standalone intervention | | | | Embedded intervention | | | |
| | Sample size | | Sample characteristics | | Sample size | | Sample characteristics | |
	Unit of assignment	Unit of analysis	Mean (adjusted mean)	Standard deviation	Unit of assignment	Unit of analysis	Mean (adjusted mean)	Standard deviation
Cohort 2, baseline sample								
FRA Word Reading	316	316	452 (455)	103	369	369	454 (452)	103
FRA Spelling	316	316	316 (318)	98	369	369	323 (321)	99
FRA Vocabulary Pairs	316	316	498 (499)	68	369	369	476 (475)	72
FRA Following Directions	316	316	484 (486)	119	369	369	476 (475)	127
Cohort 2, analytic sample[a]								
FRA Word Reading	295	295	455 (457)	102	369	369	454 (453)	103
FRA Spelling	295	295	323 (324)	96	369	369	323 (322)	99
FRA Vocabulary Pairs	295	295	499 (500)	67	369	369	476 (475)	72
FRA Following Directions	295	295	489 (490)	119	369	369	476 (476)	127
Cohorts 1 and 2 combined, baseline sample								
FRA Word Reading	639	639	464 (467)	90	670	670	486 (484)	105
FRA Spelling	639	639	333 (335)	103	670	670	339 (338)	104
FRA Vocabulary Pairs	639	639	497 (498)	75	670	670	489 (488)	81
FRA Following Directions	639	639	509 (511)	118	670	670	503 (501)	120
Cohorts 1 and 2 combined, analytic sample[a]								
FRA Word Reading	618	618	464 (469)	90	670	670	486 (484)	105
FRA Spelling	618	618	333 (338)	103	670	670	339 (338)	104
FRA Vocabulary Pairs	618	618	497 (499)	75	670	670	489 (488)	81
FRA Following Directions	618	618	509 (514)	118	670	670	503 (502)	120

FRA is Florida Center for Reading Research Reading Assessment.

a. The sample size for unit of assignment for standalone intervention schools does not include the 21 students from the one stand-alone intervention school that removed all grade 2 students from the intervention.

Note: A hierarchical linear model with students nested in schools and a dichotomous indicator for treatment (the embedded intervention served as the referent) was used to test for baseline equivalence between the standalone and embedded interventions. Because random assignment was conducted within geographic region, the model included region as a covariate when estimating the adjusted means.

Source: Authors' analysis based on data from participating districts in Florida.

Table A10. Student-level baseline scores for the baseline and analytic samples, by grade, cohort, and intervention group, 2013/14 and 2014/15

| Grade, cohort and sample, and baseline measure | Adjusted mean | | Difference (standard error) | p value | Effect size | Student sample size |
	Standalone intervention (standard deviation)	Embedded intervention (standard deviation)				
Kindergarten						
Cohort 1, baseline and analytic samples						
FRA Letter Sounds	335 (91)	347 (106)	–12 (15)	.44	–0.12	468
FRA Phonological Awareness	312 (113)	292 (122)	20 (13)	.13	0.17	468
FRA Vocabulary Pairs	364 (57)	366 (60)	–2 (8)	.80	–0.03	468
FRA Following Directions	286 (127)	243 (156)	43 (17)	.01	0.30	468
FRA Sentence Comprehension	409 (90)	397 (92)	12 (11)	.29	0.13	468
Cohort 2, baseline and analytic samples						
FRA Letter Sounds	272 (99)	267 (102)	5 (13)	.72	–0.94	593
FRA Phonological Awareness	246 (84)	253 (86)	–7 (11)	.46	–0.08	593
FRA Vocabulary Pairs	339 (65)	335 (65)	4 (8)	.62	0.06	593
FRA Following Directions	237 (154)	233 (149)	4 (18)	.82	0.03	593
FRA Sentence Comprehension	398 (85)	399 (82)	–1 (10)	.88	–0.01	593
Cohorts 1 and 2 combined, baseline and analytic samples						
FRA Letter Sounds	304 (100)	303 (111)	1 (14)	.91	0.01	1,061
FRA Phonological Awareness	279 (105)	271 (103)	8 (11)	.45	0.08	1,061
FRA Vocabulary Pairs	351 (63)	349 (64)	2 (7)	.75	0.03	1,061
FRA Following Directions	261 (144)	237 (152)	24 (13)	.07	0.16	1,061
FRA Sentence Comprehension	403 (87)	399 (86)	4 (8)	.55	0.05	1,061
Grade 1						
Cohort 1, baseline and analytic samples						
FRA Word Reading	257 (186)	342 (205)	–85 (28)	.003	–0.43	506
FRA Vocabulary Pairs	412 (62)	425 (65)	–13 (6)	.03	–0.20	506
FRA Following Directions	408 (115)	431 (119)	–23 (16)	.14	–0.20	506
FRA Sentence Comprehension	462 (129)	451 (125)	11 (12)	.40	0.09	506
Cohort 2, baseline and analytic samples						
FRA Word Reading	231 (139)	253 (148)	–22 (18)	.23	–0.15	592
FRA Vocabulary Pairs	409 (71)	399 (74)	10 (8)	.25	0.14	592
FRA Following Directions	398 (116)	386 (124)	12 (17)	.49	0.09	592
FRA Sentence Comprehension	465 (70)	464 (77)	1 (10)	.94	0.01	592
Cohorts 1 and 2 combined, baseline and analytic samples						
FRA Word Reading	243 (165)	293 (179)	–50 (19)	.01	–0.29	1,098
FRA Vocabulary Pairs	411 (67)	410 (71)	1 (6)	.99	0.02	1,098
FRA Following Directions	402 (115)	407 (124)	–5 (13)	.69	–0.04	1,098
FRA Sentence Comprehension	463 (104)	458 (100)	5 (8)	.51	0.05	1,098
Grade 2						
Cohort 1, baseline and analytic samples						
FRA Word Reading	478 (72)	521 (94)	–43 (15)	.005	–0.52	624
FRA Spelling	351 (106)	357 (107)	–6 (15)	.70	–0.06	624
FRA Vocabulary Pairs	497 (82)	503 (87)	–6 (11)	.57	–0.07	624
FRA Following Directions	535 (113)	533 (103)	2 (13)	.85	0.02	624

(continued)

Table A10. Student-level baseline scores for the baseline and analytic samples, by grade, cohort, and intervention group, 2013/14 and 2014/15 *(continued)*

Grade, cohort and sample, and baseline measure	Adjusted mean		Difference (standard error)	*p* value	Effect size	Student sample size
	Standalone intervention (standard deviation)	Embedded intervention (standard deviation)				
Cohort 2, baseline sample						
FRA Word Reading	455 (103)	452 (103)	3 (13)	.81	0.03	685
FRA Spelling	318 (98)	321 (99)	−3 (16)	.83	−0.03	685
FRA Vocabulary Pairs	499 (68)	475 (72)	24 (8)	.003	0.34	685
FRA Following Directions	486 (119)	475 (127)	11 (17)	.49	0.09	685
Cohort 2, analytic sample[a]						
FRA Word Reading	457 (102)	453 (103)	4 (13)	.72	0.04	664
FRA Spelling	324 (96)	322 (99)	2 (15)	.89	0.02	664
FRA Vocabulary Pairs	500 (67)	475 (72)	25 (8)	.003	0.36	664
FRA Following Directions	490 (119)	476 (127)	14 (17)	.39	0.11	664
Cohorts 1 and 2 combined, baseline sample						
FRA Word Reading	467 (90)	484 (105)	−17 (12)	.17	−0.20	1,309
FRA Spelling	335 (103)	338 (104)	−3 (12)	.78	−0.03	1,309
FRA Vocabulary Pairs	498 (75)	488 (81)	10 (7)	.16	0.13	1,309
FRA Following Directions	511 (118)	501 (120)	10 (13)	.47	0.08	1,309
Cohorts 1 and 2 combined, analytic sample[a]						
FRA Word Reading	469 (90)	484 (105)	−15 (12)	.21	−0.15	1,288
FRA Spelling	338 (103)	338 (104)	0 (11)	.99	0.00	1,288
FRA Vocabulary Pairs	499 (75)	488 (81)	11 (7)	.14	0.15	1,288
FRA Following Directions	514 (118)	502 (120)	12 (13)	.35	0.10	1,288

FRA is Florida Center for Reading Research Reading Assessment.

a. The sample size for unit of assignment for standalone intervention schools does not include the 21 students from the one standalone intervention school that removed all grade 2 students from the intervention.

Note: A hierarchical linear model with students nested in schools and a dichotomous indicator for treatment (the embedded intervention served as the referent) was used to test for baseline equivalence between the standalone and embedded interventions. Because random assignment was conducted within geographic region, the model included region as a covariate when estimating the adjusted means.

Source: Authors' analysis based on data from participating districts in Florida.

Table A11. Small-group baseline scores on Florida Center for Reading Research Reading Assessment (FRA) subtests for the analytic sample, grade, intervention group, and cohort, 2013/14 and 2014/15

	Standalone intervention						Embedded intervention					
	Cohort 1		Cohort 2		Cohorts 1 and 2 combined		Cohort 1		Cohort 2		Cohorts 1 and 2 combined	
Grade and outcome measure	Number of small groups	Mean (standard deviation)	Number of small groups	Mean (standard deviation)	Number of small groups	Mean (standard deviation)	Number of small groups	Mean (standard deviation)	Number of small groups	Mean (standard deviation)	Number of small groups	Mean (standard deviation)
Kindergarten												
FRA Letter Sounds	63	340 (56)	68	271 (57)	131	304 (65)	51	359 (77)	75	269 (50)	126	305 (76)
FRA Phonological Awareness	63	314 (62)	68	244 (55)	131	277 (68)	51	292 (60)	75	252 (56)	126	269 (61)
FRA Vocabulary Pairs	63	355 (44)	68	333 (49)	131	343 (48)	51	361 (59)	75	332 (38)	126	344 (49)
FRA Following Directions	63	288 (75)	68	236 (98)	131	261 (91)	51	236 (88)	75	236 (94)	126	236 (91)
FRA Sentence Comprehension	63	409 (56)	68	392 (55)	131	400 (56)	51	395 (44)	75	398 (47)	126	397 (46)
Grade 1												
FRA Word Reading	64	248 (99)	65	229 (103)	129	238 (101)	59	342 (133)	78	254 (107)	137	292 (126)
FRA Vocabulary Pairs	64	412 (30)	65	404 (50)	129	408 (41)	59	422 (36)	78	394 (52)	137	406 (48)
FRA Following Directions	64	386 (92)	65	393 (73)	129	390 (83)	59	412 (84)	78	382 (80)	137	395 (83)
FRA Sentence Comprehension	64	460 (74)	65	464 (42)	129	462 (60)	59	451 (72)	78	464 (46)	137	458 (59)
Grade 2												
FRA Word Reading	70	469 (58)	60	443 (86)	130	457 (74)	63	529 (83)	74	446 (85)	137	484 (94)
FRA Vocabulary Pairs	70	493 (52)	60	498 (36)	130	496 (45)	63	505 (52)	74	470 (47)	137	486 (52)
FRA Following Directions	70	528 (63)	60	483 (71)	130	507 (71)	63	536 (68)	74	473 (84)	137	502 (83)
FRA Spelling	70	348 (58)	60	316 (64)	130	333 (63)	63	359 (61)	74	324 (64)	137	340 (65)

Source: Authors' analysis based on data from participating districts in Florida.

Table A12. Average student baseline and outcome percentile rank for the analytic sample, by grade, cohort, and intervention group, 2013/14 and 2014/15

Grade and outcome measure	Cohort 1						Cohort 2					
	Standalone intervention			Embedded intervention			Standalone intervention			Embedded intervention		
	Baseline	Outcome	Difference	Baseline	Outcome	Difference	Baseline	Outcome	Difference	Baseline	Outcome	Difference
Kindergarten												
FRA Phonological Awareness	1	19	18	1	28	27	1	23	22	1	25	24
FRA Word Reading	na	46	na	na	46	na	na	20	na	na	20	na
SESAT Word Reading	na	29	na	na	22	na	na	23	na	na	20	na
FRA Vocabulary Pairs	31	34	3	33	36	3	21	34	13	19	33	14
FRA Following Directions	11	33	22	5	26	21	5	23	18	5	26	21
FRA Sentence Comprehension	12	50	38	9	38	29	9	26	17	10	29	19
SESAT Sentence Reading	na	25	na	na	24	na	na	21	na	na	22	na
Grade 1												
FRA Word Reading	1	30	29	4	47	43	1	17	16	1	15	14
FRA Vocabulary Pairs	13	16	3	16	18	2	12	21	9	10	18	8
FRA Following Directions	11	17	6	17	27	10	9	21	12	7	17	10
FRA Sentence Comprehension	29	66	37	25	72	47	29	63	34	29	62	33
SAT-10 Reading Comprehension	na	16	na	na	18	na	na	11	na	na	10	na
Grade 2												
FRA Word Reading	7	28	21	18	37	19	4	20	16	4	18	14
FRA Spelling	9	25	16	6	21	15	2	20	18	3	15	12
FRA Vocabulary Pairs	12	23	11	12	23	11	12	22	10	7	14	7
FRA Following Directions	22	38	16	23	39	16	9	23	14	8	17	9
FRA Sentence Comprehension	na	88	na	na	82	na	58	86	28	57	82	25
SAT-10 Reading Comprehension	na	16	na	na	19	na	na	13	na	na	11	na

FRA is the Florida Center for Reading Research Reading Assessment. SESAT is the Stanford Early Scholastic Achievement Test. SAT-10 is the Stanford Achievement Test, 10th edition. na is not applicable.

Note: Percentiles are based on winter norms.

Source: Authors' analysis of the K–2 FRA data and SESAT/SAT-10 data, 2013–15.

Attrition

Attrition occurs when study participants initially assigned to intervention groups are missing outcome data. The level of attrition is determined by a combination of overall attrition (calculated across both interventions) and differential attrition (calculated as the difference in attrition rates between intervention groups). U.S. Department of Education (2014) provides a table for determining the level of attrition based on overall and differential attrition. High levels of attrition can lead to biased estimates of an intervention's effectiveness. Therefore, it is important to determine the level of attrition, based on What Works Clearinghouse (WWC) criteria, within the current study.

In the current cluster-level randomized controlled trial, attrition is evaluated at the school and student levels to ensure that the estimates of effectiveness for the standalone intervention are not biased. Using the WWC liberal boundary for attrition (U.S. Department of

Table A13. Overall and differential attrition estimates, by grade, school and student level, and cohort, 2013/14 and 2014/15

Grade and sample	Baseline sample size		Analytic sample size		Attrition		
	Standalone intervention	Embedded intervention	Standalone intervention	Embedded intervention	Overall	Differential	Level[a]
Kindergarten							
School level							
Cohort 1	14	13	14	13	0	0	Low
Cohort 2	13	15	13	15	0	0	Low
Cohorts 1 and 2 combined	27	28	27	28	0	0	Low
Student level							
Cohort 1	255	213	225	193	10.7	2.4	Low
Cohort 2	276	317	249	281	10.6	1.6	Low
Cohorts 1 and 2 combined	531	530	474	474	10.7	0.2	Low
Grade 1							
School level							
Cohort 1	14	13	14	13	0	0	Low
Cohort 2	13	15	13	15	0	0	Low
Cohorts 1 and 2 combined	27	28	27	28	0	0	Low
Student level							
Cohort 1	267	239	234	214	11.5	1.9	Low
Cohort 2	267	325	237	293	10.5	1.4	Low
Cohorts 1 and 2 combined	534	564	471	507	10.9	2.7	Low
Grade 2							
School level							
Cohort 1	14	13	14	13	0	0	Low
Cohort 2	13	15	12	15	3.6	7.7	Low
Cohorts 1 and 2 combined	27	28	26	28	1.8	3.7	Low
Student level							
Cohort 1	323	301	289	263	11.5	2.1	Low
Cohort 2	295	369	259	322	12.5	0.5	Low
Cohorts 1 and 2 combined	618	670	548	585	12.0	1.4	Low

a. Based on the liberal boundary determined by What Works Clearinghouse criteria (U.S. Department of Education, 2014).

b. This excludes the 21 students attending the cohort 2 school that withdrew from the study because of scheduling conflicts.

Source: Authors' analysis based on data from participating districts in Florida.

Education, 2014), school- and student-level attrition was determined to be low for all grades and by cohort (table A13).

Treatment of missing data

Multiple imputation for clustered data sets (Mistler, 2013) was used by grade, cohort, and intervention group to account for missing outcome data. The multiple imputation procedure was conducted using a multilevel multiple imputation macro in SAS (Mistler, 2013) that takes into account the nested structure of the data. In the imputation procedure, several variables, including baseline, outcome, and student-level demographics (gender, eligibility for the federal lunch program, English learner status, and race/ethnicity), were used to inform the imputations. One thousand imputed files per grade, cohort, and intervention group were created and aggregated for use in all analyses.

All eligible students with parent consent participated in an intervention. The proportion of students across grades K–2 that did not complete outcome testing ranged from approximately 11 percent to 13 percent. However, attrition rates were determined to be low based on the liberal boundary determined by WWC criteria, and multiple imputation for clustered data was used to account for missing outcome data. Therefore, all baseline students (eligible students with parent consent) were included in all analyses.

Methodology

Prior to data analysis, descriptive analyses were conducted to identify the presence of outliers and to verify that the data were normally distributed. Corrections for outliers were made during this data cleaning process, and all measures demonstrated normality. Outliers were identified using the median plus or minus two interquartile ranges, such that any value that exceeded this range was considered an outlier and scores were changed to reflect the appropriate bound.

Analytic approach and statistical adjustments. A three-level hierarchical linear model (HLM) with students nested in small groups, nested in schools was used to estimate treatment effects by grade using the MIXED procedure in SAS (version 9.4). Prior to estimating any models, an unconditional model was estimated for each outcome to calculate the intraclass correlation (the proportion of variance in an outcome that is accounted for by differences between students, between small groups, and between schools) for each level modeled in the estimated three-level HLM (table A14).

Table A14. Percentage of variance in each outcome that is accounted for by differences between students, between small groups, and between schools, by grade, 2013/14 and 2014/15

Grade and level	Stanford Early Scholastic Achievement Test/Stanford Achievement Test, 10th edition			Florida Center for Reading Research Reading Assessment					
	Word Reading	Sentence Reading	Reading Comprehension	Phonological Awareness	Word Reading	Vocabulary Pairs	Following Directions	Sentence Comprehension	Spelling
Kindergarten									
Student	72	75	na	91	68	93	90	87	na
Small group	8	12	na	3	8	5	5	1	na
School	20	13	na	6	24	2	5	12	na
Grade 1									
Student	na	na	74	na	73	89	87	88	na
Small group	na	na	10	na	10	1	1	5	na
School	na	na	16	na	18	10	12	7	na
Grade 2									
Student	na	na	81	na	83	91	87	94	82
Small group	na	na	6	na	5	1	0	0	10
School	na	na	13	na	12	8	13	6	8

na is not applicable.

Source: Authors' analysis based on data from participating districts in Florida.

Research question 2 was addressed using the following full sample HLM equation by grade:

Level 1 (student)

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}(Baseline)_{ijk} + e_{ijk}$$

Level 2 (small-group)

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}(Group\ Baseline)_{jk} + r_{0jk}$$
$$\pi_{1jk} = \beta_{10k}$$

Level 3 (school)

$$\beta_{00k} = \gamma_{000} + \gamma_{001}(School\ Baseline)_k + \gamma_{002}(Region)_k + \gamma_{003}(Treatment)_k + u_{00k}$$
$$\beta_{01k} = \gamma_{010}$$
$$\beta_{10k} = \gamma_{100}$$

Mixed model

$$Y_{ijk} = \gamma_{000} + \gamma_{001}(School\ Baseline)_k + \gamma_{002}(Region)_k + \gamma_{003}(Treatment)_k + \gamma_{010}(Group\ Baseline)_{jk} + \gamma_{100}(Baseline)_{ijk} + u_{00k} + r_{0jk} + e_{ijk}$$

where i denotes a student, j denotes a small group, k denotes a school, Y is the outcome variable being studied, Baseline is a vector of FRA baseline scores for each grade, *Group Baseline* is a vector of small-group aggregated FRA baseline scores for each grade, *School Baseline* is a vector of school aggregated FRA baseline scores for each grade, *Region* is the stratifying variable used for random assignment, and *Treatment* is a dichotomous variable indicating student's intervention (embedded serves as the referent). All FRA baseline scores were included as covariates at the student, group, and school level for all outcomes, including FRA and SESAT/SAT-10 outcomes. All continuous predictors and region were grand mean centered. Relative impacts of the two interventions for the full sample model by grade and outcome are reported in table A15.

Table A15. Relative impact of the standalone and embedded interventions for the full sample, by grade and outcome, 2013/14 and 2014/15

Grade and outcome measure	Sample size		Adjusted mean		Difference (standard error)	*p* value	Effect size	Improvement index[a]
	Standalone intervention	Embedded intervention	Standalone intervention (standard deviation)	Embedded intervention (standard deviation)				
Kindergarten								
FRA Phonological Awareness	531	530	434 (147)	452 (134)	–18 (13)	.18	–0.13	–5
FRA Word Reading	531	530	332 (134)	337 (149)	–5 (17)	.79	–0.04	–1
SESAT Word Reading	531	530	433 (38)	426 (35)	7 (5)	.18	0.19	8
FRA Vocabulary Pairs	531	530	369 (77)	372 (76)	–3 (5)	.56	–0.04	–2
FRA Following Directions	531	530	358 (115)	363 (105)	–5 (7)	.56	–0.05	–2
FRA Sentence Comprehension	531	530	472 (81)	476 (77)	–4 (6)	.58	–0.05	–2
SESAT Sentence Reading	531	530	459 (50)	460 (46)	–1 (6)	.80	–0.02	–1
Grade 1								
FRA Word Reading	534	564	448 (105)	436 (123)	12 (13)	.37	0.10	4
FRA Vocabulary Pairs	534	564	435 (79)	428 (86)	7 (7)	.35	0.08	3
FRA Following Directions	534	564	442 (109)	440 (117)	2 (9)	.80	0.02	1
FRA Sentence Comprehension	534	564	542 (87)	542 (87)	0 (6)	.96	0.00	0
SAT-10 Reading Comprehension	534	564	519 (39)	514 (42)	5 (5)	.28	0.12	5
Grade 2								
FRA Word Reading	618	670	546 (82)	541 (100)	5 (8)	.52	0.05	2
FRA Spelling	618	670	434 (88)	417 (98)	17 (6)	.009*	0.18	7
FRA Vocabulary Pairs	618	670	526 (80)	519 (81)	8 (6)	.23	0.09	3
FRA Following Directions	618	670	556 (122)	548 (125)	8 (9)	.31	0.05	2
FRA Sentence Comprehension	618	670	601 (89)	589 (88)	12 (7)	.06	0.10	4
SAT-10 Reading Comprehension	618	670	565 (31)	565 (32)	0 (3)	.88	0.00	0

FRA is the Florida Center for Reading Research Reading Assessment. SESAT is the Stanford Early Scholastic Achievement Test. SAT-10 is Stanford Achievement Test, 10th edition.

* *p*-value is significant after applying the Benjamini-Hochberg Correction procedure (1995) where the identified *p*-value cut-off for reading outcomes is $p \leq .025$.

a. The expected change in percentile rank of an average student in an embedded intervention school had the student been in a standalone intervention school.

Source: Authors' analysis based on data from participating districts in Florida.

A top-down approach was used to answer the second part of research question 2, in which the full subgroup model that includes all of the covariates, treatment indicator, and interactions (vectors of baseline score by treatment, cohort by treatment, baseline score by cohort, and baseline score by cohort by treatment interactions), is estimated first, and then nonsignificant predictors (interactions and cohort) are removed iteratively in subsequent models (West, Welch, & Galecki, 2007). The full subgroup HLM by grade is represented by:

Level 1 (student)

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}(Baseline)_{ijk} + e_{ijk}$$

Level 2 (small-group)

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}(Group\ Baseline)_{jk} + r_{0jk}$$
$$\pi_{1jk} = \beta_{10k}$$

Level 3 (school)

$$\beta_{00k} = \gamma_{000} + \gamma_{001}(School\ Baseline)_k + \gamma_{002}(Region)_k + \gamma_{003}(Cohort)_k + \gamma_{004}(Treatment)_k +$$
$$\gamma_{005}(Cohort * Treatment)_k + u_{00k}$$
$$\beta_{01k} = \gamma_{010}$$

$$\beta_{01k} = \gamma_{100} + \gamma_{101}(Cohort)_k + \gamma_{102}(Treatment)_k + \gamma_{103}(Cohort * Treatment)_k$$

Mixed model

$$Y_{ijk} = \gamma_{000} + \gamma_{001}(School\ Baseline)_k + \gamma_{002}(Region)_k + \gamma_{003}(Cohort)_k +$$
$$\gamma_{004}(Treatment)_k + \gamma_{005}(Cohort * Treatment)_k + \gamma_{010}(Group\ Baseline)_{jk} +$$
$$\gamma_{100}(Baseline)_{ijk} + \gamma_{101}(Baseline)_{ijk}(Cohort)_k + \gamma_{102}(Baseline)_{ijk}(Treatment)_k +$$
$$\gamma_{103}(Baseline)_{ijk}(Cohort * Treatment)_k + u_{00k} + r_{0jk} + e_{ijk}$$

where i denotes a student, j denotes a small group, k denotes a school, Y is the outcome variable being studied, *Baseline* is a vector of FRA baseline scores for each grade, *Group Baseline* is a vector of small-group aggregated FRA baseline scores for each grade, *School Baseline* is a vector of school aggregated FRA baseline scores for each grade, *Region* is the stratifying variable used for random assignment, *Cohort* is a dichotomous variable indicating cohort (cohort 1 serves as the referent), and *Treatment* is a dichotomous variable indicating student's intervention (embedded serves as the referent). All FRA baseline scores were included as covariates at the student, small group, and school levels for all FRA, SESAT, and SAT-10 outcomes. All continuous predictors and region were grand mean centered.

The removal of nonsignificant predictors from the full subgroup model followed a systematic process, such that the three-way interactions (baseline score by cohort by treatment interactions) were removed first, then two-way interactions (baseline score by treatment, cohort by treatment, and baseline score by cohort interactions), and finally cohort. Baseline covariates at all levels were retained regardless of significance to increase the precision of the treatment effect. If the final subgroup model included a significant treatment interaction, the highest level interaction (the three-way or two-way interaction) involving the treatment variable was explored further. A significant three-way interaction among treatment, cohort, and baseline score was explored further by testing treatment differences within each cohort when baseline score was either one standard deviation above the mean

or one standard deviation below the mean. A significant two-way interaction between treatment and cohort was explored further by testing treatment differences within each cohort. Finally, a significant two-way interaction between treatment and baseline score was explored further by testing treatment differences when baseline was either one standard deviation above or below the mean (see tables 2 and 3 in the main text for differences in reading and language outcomes between the standalone and embedded interventions by grade for outcomes that included a significant interaction involving the treatment indicator).

The final subgroup HLM did not include a significant interaction involving the treatment indicator for several reading and language outcomes across grades K–2. When this occurred, adjusted means for the standalone and embedded interventions were estimated for the full sample and are reported in table A16. The adjusted means reported in table A16 may differ slightly from those reported in table A15 because of the inclusion of cohort in the full subgroup model. If the cohort indicator in the final subgroup model was significant, it was retained.

The final subgroup model from the primary impact analysis for each outcome by grade was then used as the base model when estimating treatment differences for English and non–English learner students. At a minimum, two variables were added to the English learner status base model: Student level *English learner status* and the cross-level *English learner status* by *Treatment* interaction. The top-down approach described above was used to iteratively remove nonsignificant predictors from the English learner status base models.

Effect sizes (Hedges's *g*) were calculated by dividing effect estimates by the unadjusted pooled within-group standard deviation. The improvement index was calculated using the approach outlined in U.S. Department of Education (2014).

Multiple hypothesis testing. Multiple hypothesis tests were included by grade within the reading and language outcomes. Reading outcomes included the Phonological Awareness (kindergarten only), Word Reading, and Spelling (grade 2 only) subtests from the FRA (see table A6) and the Word Reading subtest from SESAT in kindergarten. Language outcomes included the Vocabulary Pairs, Following Directions, and Sentence Comprehension subtests from the FRA (see table A6), the Sentence Reading subtest from the SESAT in kindergarten, and the Reading Comprehension subtest from the SAT-10 in grades 1 and 2. The estimation of multiple hypothesis tests can increase the probability of falsely detecting a statistically significant treatment effect. Therefore, a correction to all significant treatment effects must be applied to reduce the false discovery rate. The Benjamini–Hochberg linear step-up procedure (Benjamini & Hochberg, 1995) was used by research question, grade, and outcome type (that is, reading and language) to correct for multiple hypothesis testing (table A17) following procedures outlined in U.S. Department of Education (2014).

The Benjamini–Hochberg (1995) procedure is conducted in three steps. First, the *p*-values associated with statistically significant treatment effects within an outcome type are ranked in ascending order. Second, a critical *p*-value is computed for each ranked *p*-value by multiplying the rank by 0.05 and dividing the product by the total number of significant and nonsignificant treatment effects within the research question, grade, and outcome type. Third, a *p*-value cutoff is identified by finding the largest rank that is associated with a model-estimated *p*-value that is less than or equal to the critical *p*-value. This *p*-value

Table A16. Relative impact of the standalone and embedded interventions for reading and language outcomes with no significant subgroup interactions in the final subgroup hierarchical linear model, by grade and outcome, 2013/14 and 2014/15

Grade, outcome type, and outcome measure	Sample size		Adjusted mean		Difference (standard error)	p value	Effect size	Improvement index[a]
	Standalone intervention	Embedded intervention	Standalone intervention (standard deviation)	Embedded intervention (standard deviation)				
Kindergarten								
Reading outcomes								
FRA Phonological Awareness	531	530	434 (147)	452 (133)	−18 (14)	.18	−0.13	−5
FRA Word Reading	531	530	330 (133)	335 (149)	−5 (13)	.71	−0.03	−1
Language outcomes								
FRA Vocabulary Pairs	531	530	369 (77)	372 (76)	−3 (5)	.56	−0.04	−2
FRA Following Directions	531	530	359 (115)	362 (105)	−3 (8)	.66	−0.03	−1
FRA Sentence Comprehension	531	530	472 (81)	476 (77)	−4 (6)	.46	−0.05	−2
SESAT Sentence Reading	531	530	459 (46)	460 (50)	−1 (6)	.80	−0.02	−1
Grade 1								
Language outcomes								
FRA Vocabulary Pairs	534	564	438 (79)	426 (86)	11 (7)	.08	0.14	5
FRA Sentence Comprehension	534	564	541 (87)	543 (87)	−2 (6)	.78	−0.02	−1
SAT-10 Reading Comprehension	534	564	519 (39)	514 (42)	5 (5)	.28	0.13	5
Grade 2								
Reading outcomes								
FRA Word Reading	618	670	545 (82)	542 (100)	3 (8)	.71	0.03	1
Language outcomes								
FRA Vocabulary Pairs	618	670	526 (80)	518 (81)	8 (6)	.22	0.10	4
FRA Following Directions	618	670	555 (122)	549 (125)	6 (8)	.51	0.05	2
SAT-10 Reading Comprehension	618	670	564 (31)	565 (32)	−1 (3)	.74	−0.04	−2

FRA is the Florida Center for Reading Research Reading Assessment. SESAT is the Stanford Early Scholastic Achievement Test. SAT-10 is Stanford Achievement Test, 10th edition.

Note: A hierarchical linear model with students nested in small groups and small groups nested in schools was estimated for each of grades K–2. For each outcome measure the full subgroup model (see the appendix for model equation) included all grade specific baseline scores, several dichotomous indicators for region, cohort, and treatment, and several interactions including baseline score by treatment, cohort by treatment, baseline score by cohort, and baseline score by cohort by treatment. If the final subgroup model for an outcome did not include any significant interactions involving the treatment indicator it is included in this table and the sample size reflects the full sample by grade.

a. The expected change in percentile rank of an average student in an embedded intervention school had the student been in a standalone intervention school.

Source: Authors' analysis based on data from participating districts in Florida.

cutoff becomes the threshold for identifying a significant treatment effect. In other words, if the model-estimated *p*-value is less than or equal to the identified *p*-value cutoff, the treatment effect is considered significant after the Benjamini–Hochberg correction. Conversely, if the model estimated *p*-value is greater than the identified *p*-value cutoff, the treatment is no longer considered significant after the Benjamini–Hochberg correction. When identifying the *p*-value cutoff, it is possible for a model-estimated *p*-value with a rank lower than the one identified in step three to exceed the critical *p*-value.

Table A17. Benjamini–Hochberg linear step-up procedure applied to the significant treatment effects by research question, grade, and outcome type

Samples compared	Outcome	Model p value	Rank	Total effects	Critical p value	Significant after correction
Research question 2 full sample						
Grade 2, reading outcomes						
Standalone compared to embedded intervention for full sample	FRA Spelling	.009	1	2	.025	Yes
Standalone compared to embedded intervention for full sample	FRA Word Reading	.52	2	2	.05	No
Research question 2 subgroup						
Grade 2, reading outcomes						
Standalone compared to embedded intervention for a subset of full sample with low FRA spelling baseline scores	FRA Spelling	.001	1	20	.0025	Yes
Standalone compared to embedded intervention for a subset of full sample with high FRA spelling baseline scores	FRA Spelling	.29	2	20	.005	No
Standalone compared to embedded intervention for full sample	FRA Word Reading	.71	3	20	.0075	No
Grade 2, language outcomes						
Standalone compared to embedded intervention for a subset of cohort 1 with low FRA vocabulary pairs baseline scores	FRA Sentence Comprehension	.001	1	40	.00125	Yes
Standalone compared to embedded intervention for a subset of cohort 1 with high FRA vocabulary pairs baseline scores	FRA Sentence Comprehension	.12	2	40	.0025	No
Standalone compared to embedded intervention for full sample	FRA Vocabulary Pairs	.22	3	40	.00375	No
Standalone compared to embedded intervention for a subset of cohort 2 with high FRA vocabulary pairs baseline scores	FRA Sentence Comprehension	.28	4	40	.005	No
Standalone compared to embedded intervention for full sample	FRA Following Directions	.51	5	40	.00625	No
Standalone compared to embedded intervention for a subset of cohort 2 with low FRA vocabulary pairs baseline scores	FRA Sentence Comprehension	.71	6	40	.0075	No
Standalone compared to embedded for full sample	SAT-10 Reading Comprehension	.74	7	40	.00875	No
Research question 3						
Kindergarten, reading outcomes						
Non–English learner students compared to English learner students in the embedded intervention	SESAT Word Reading	.006	1	12	.004	No
Standalone compared to embedded intervention for English learner students	FRA Phonological Awareness	.01	2	12	.008	No
Non–English learner students compared to English learner students in the embedded intervention	FRA Phonological Awareness	.02	3	12	.013	No

(continued)

Table A17. Benjamini–Hochberg linear step-up procedure applied to the significant treatment effects by research question, grade, and outcome type *(continued)*

Samples compared	Outcome	Model *p* value	Rank	Total effects	Critical *p* value	Significant after correction
Non–English learner students compared to English learner students in the embedded intervention	FRA Word Reading	.03	4	12	.017	No
Standalone compared to embedded intervention for non–English learner students	SESAT Word Reading	.04	5	12	.02	No
Standalone compared to embedded intervention for non–English learner students	FRA Word Reading	.11	6	12	.025	No
Non–English learner students compared to English learner students in the standalone intervention	SESAT Word Reading	.28	7	12	.029	No
Non–English learner students compared to English learner students in the standalone intervention	FRA Phonological Awareness	.50	8	12	.03	No
Non–English learner students compared to English learner students in the standalone intervention	FRA Word Reading	.61	9	12	.038	No
Standalone compared to embedded intervention for English learner students	FRA Word Reading	.69	10	12	.04	No
Standalone compared to embedded intervention for English learner students	SESAT Word Reading	.76	11	12	.046	No
Standalone compared to embedded intervention for non–English learner students	FRA Phonological Awareness	.79	12	12	.05	No

FRA is Florida Center for Reading Research Reading Assessment. SESAT is the Stanford Early Scholastic Achievement Test. SAT-10 is Stanford Achievement Test, 10th edition.

Note: Total effects are the total number of significant and nonsignificant treatment effects within the research question, grade, and outcome type following procedures outlined in U.S. Department of Education (2014).

Source: Authors' analysis based on data from participating districts in Florida.

Notes

1. Across cohorts 1 and 2 the north Florida region included an odd number of participating schools (that is, three schools in each cohort). Therefore, two schools were randomly assigned to the standalone intervention in cohort 1 and two schools were randomly assigned to the embedded intervention in cohort 2. In cohort 2 the central Florida region also included an odd number of participating schools. In this case five schools were randomly assigned to the embedded intervention and four schools were assigned to the standalone intervention.

2. One of the standalone intervention schools in cohort 2 was excluded from the grade 2 analyses because scheduling conflicts resulted in the withdrawal of the 21 participating grade 2 students at that school.

References

Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., et al. (2014). *Teaching academic content and literacy to English learners in elementary and middle school* (NCEE No. 2014–4012). National Center for Education Evaluation and Regional Assistance Working Paper. Washington, DC: U.S. Department of Education. http:// eric.ed.gov/?id=ED544783

Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., & Gersten, R. (2015). *Evaluation of Response to Intervention practices for elementary school reading* (NCEE 2016–4000). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. http://eric. ed.gov/?id=ED560820

Beck, I., McKeown, M., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. New York, NY: Guilford Press.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B Methodological, 57*(1), 289–300. Retrieved January 5, 2015, from http://www.jstor.org/ stable/2346101.

Bialystok, E., Majumder, S., & Martin, M. M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics*, *24*(1), 27–44. http://eric. ed.gov/?id=EJ664619

Dombek, J. L., Foorman, B. R., Garcia, M., & Smith, K. G. (2016). *Early literacy interventions self-study guide for implementation* (REL 2016–129). Washington, DC: U.S. Department of Education, Institute of Education Sciences. National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. http://eric.ed.gov/?id=ED565630

Educational Testing Service. (2005). *Comprehensive English Language Learning Assessment (CELLA): Technical summary report.* Princeton, NJ: Authors. Retrieved January 5, 2015, from http://www.accountabilityworks.org/photos/CELLA_Technical_Summary_ Report.pdf.

Foorman, B. R., & Al Otaiba, S. (2009). Reading remediation: State of the art. In K. Pugh & P. McCardle (Eds.), *How children learn to read: Current issues and new directions in the integration of cognition, neurobiology and genetics of reading and dyslexia research and practice* (pp. 257–274). New York, NY: Psychology Press.

Foorman, B., Beyler, N., Borradaile, K., Coyne, M., Denton, C. A., Dimino, J., et al. (2016). *Foundational skills to support reading for understanding in kindergarten through 3rd grade* (NCEE No. 4008). National Center for Education Evaluation and Regional Assistance Working Paper. Washington, DC: U.S. Department of Education. http://eric. ed.gov/?id=ED566956

Foorman, B. R., Breier, J. I., & Fletcher, J. M. (2003). Interventions aimed at improving reading success: An evidence-based approach. *Developmental Neuropsychology, 24*(2/3), 613–639.

Foorman, B. R., & Connor, C. (2011). Primary reading. In M. Kamil, P. D. Pearson, & E. Moje (Eds.), *Handbook on Reading Research,* vol. IV (pp. 136–156). New York, NY: Taylor & Francis.

Foorman, B. R., Dombek, J., & Smith, K. (2016). Seven elements important to successful implementation of early literacy intervention. *New Directions for Child and Adolescent Development, 154,* 49–65.

Foorman, B. F., Herrera, S., Petscher, Y., Mitchell, A., & Truckenmiller, A. (2015). The structure of oral language and reading and their relation to comprehension in kindergarten through grade 2. *Reading and Writing, 28*(5), 655–681. http://eric.ed.gov/?id=EJ1057505

Foorman, B., Petscher, Y., & Schatschneider, C. (2015). *Florida Center for Reading Research (FCRR) reading assessments: Technical manuals.* Tallahassee, FL: Florida State University. Retrieved October 5, 2015, from http://www.fcrr.org/for-researchers/fra.asp.

Foorman, B. R., & Torgesen, J. K. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research and Practice, 16*(4), 202–211. http://eric.ed.gov/?id=EJ637166

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., et al. (2009). *Assisting students struggling with reading: Response to intervention and multi-tier intervention for reading in the primary grades. A practice guide* (NCEE No. 2009–4045). National Center for Education Evaluation and Regional Assistance Working Paper. Washington, DC: U.S. Department of Education. http://eric.ed.gov/?id=ED504264

Gersten, R., Newman-Gonchar, R., Haymond, K. S., & Dimino, J. (in press). *What is the evidence base for response to intervention in reading for grades 1–3?* (REL 2017–271). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast.

Lemons, C. J., Fuchs, D., Gilbert, J., & Fuchs, L. S. (2014). Evidence-based practices in a changing world: Reconsidering the counterfactual in education research. *Educational Researcher, 43*(5), 242–252. http://eric.ed.gov/?id=EJ1032986

Mehta, P., Foorman, B. R., Branum-Martin, L., & Taylor, W. P. (2005). Literacy as a unidimensional multilevel construct: Validation, sources of influence, and implications in a longitudinal study of grades 1–4. *Scientific Studies of Reading, 9*(2), 85–116. http://eric.ed.gov/?id=EJ683144

Mistler, S. A. (2013). *A SAS macro for applying multiple imputation to multilevel data* (SAS Global Forum Working Paper No. 438–2013). Cary, NC: SAS Institute Inc.

Montgomery, J. (2007). *Bridge of vocabulary*. New York, NY: Pearson.

National Institute of Child Health and Human Development. (2000). *National reading panel—Teaching children to read: Reports of the subgroups* (NIH Pub. No. 00–4754). Washington, DC: U.S. Department of Health and Human Services. Retrieved January 5, 2015, from https://www.nichd.nih.gov/publications/pubs/nrp/Documents/report.pdf.

Phillips, B. (2014). Promotion of syntactical development and oral comprehension: Development and initial evaluation of a small-group intervention. *Child Language Teaching and Therapy, 30*(1), 63–77. http://eric.ed.gov/?id=EJ1019082

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2*(2), 31–74.

Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement, 28*(3), 237–247. http://eric.ed.gov/?id=EJ435193

Smith, J. L, Nelson, N. J., Smolkowski, K., Baker, S. K., Fien, H., & Kosty, D. (2016). Examining the efficacy of a multitiered intervention for at-risk readers in grade 1. *The Elementary School Journal, 116*(4), 549–573. http://eric.ed.gov/?id=EJ1103958

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Research Council.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2014). *Procedures and standards handbook version 3.0*. Washington, DC: Institute for Education Sciences. Retrieved January 5, 2015, from http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19.

Vadasy, P., & Sanders, E. (2012). Two-year follow-up of a kindergarten phonics intervention for English learners and native English speakers: Contextualizing treatment impacts by classroom literacy instruction. *Journal of Educational Psychology, 104*(4), 987–1005. http://eric.ed.gov/?id=EJ994096

Vadasy, P., Sanders, E., & Abbott, R. (2008). Effects of supplemental early reading intervention at 2-year follow up: Reading skill growth patterns and predictors. *Scientific Studies of Reading, 12*(1), 51–89. http://eric.ed.gov/?id=EJ785486

Vadasy, P., Sanders, E., & Peyton, J. (2006). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trial with paraeducator implementers. *Journal of Educational Psychology, 98*(3), 508–528. http://eric.ed.gov/?id=EJ742197

Vadasy, P., Wayne, S., O'Connor, R., Jenkins, J., Firebaugh, M., & Peyton, J. (2004). *Sound partners*. Denver, CO: Sopris West.

West, B. T., Welch, K. B., & Galecki, A. T. (2007). *Linear mixed models: Practical guide using statistical software*. Boca Raton, FL: Chapman & Hall.

The Regional Educational Laboratory Program produces 7 types of reports

Making Connections
Studies of correlational relationships

Making an Impact
Studies of cause and effect

What's Happening
Descriptions of policies, programs, implementation status, or data trends

What's Known
Summaries of previous research

Stated Briefly
Summaries of research findings for specific audiences

Applied Research Methods
Research methods for educational settings

Tools
Help for planning, gathering, analyzing, or reporting data or research