Fostering the Development of Reading Skill Through Supplemental Instruction:

Results for Hispanic and Non-Hispanic Students

Barbara Gunn, Keith Smolkowski, Anthony Biglan, Carol Black, and **Jason Blair**
Oregon Research Institute

This article reports the effects of a 2-year supplemental reading program for students in kindergarten through third grade that focused on the development of decoding skills and reading fluency. Two hundred ninety-nine students were identified for participation and were randomly assigned to the supplemental instruction or to a no-treatment control group. Participants' reading ability was assessed in the fall, before the first year of the intervention, and again in the spring of Years 1, 2, 3, and 4. At the end of the 2-year intervention, students who received the supplemental instruction performed significantly better than their matched controls on measures of entry-level reading skills (i.e., letter–word identification and word attack), oral reading fluency, vocabulary, and comprehension. The benefits of the instruction were still clear 2 years after instruction had ended, with students in the supplemental-instruction condition still showing significantly greater growth on the measure of oral reading fluency. Hispanic students benefited from the supplemental reading instruction in English as much as or more than non-Hispanic students. Results support the value of supplemental instruction focused on the development of word recognition skills for helping students at risk for reading failure.

The long-term impact of reading failure on school success is well established (Cunningham & Stanovich, 1998; Juel, 1988; Slavin et al., 1996). So, too, is the relation between learning to read in the primary grades and the development of reading ability throughout elementary school (Francis, Shaywitz, Steubing, Shaywitz, & Fletcher, 1996; Juel, 1988). Reading acquisition is frequently viewed as a "bottom-up" process, based on the development of word recognition skills to promote fluency and comprehension (Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001). Within this framework, fluent word recognition allows the reader to allocate increased attention to key comprehension processes, such as making meaningful connections between sentences within a passage or relating text meaning to prior experiences and information (Fuchs, Fuchs, Hosp, & Jenkins, 2001). Thus, learning how to decode text provides a requisite foundation not only for reading fluency but also for higher level comprehension processes.

Evidence from 20 years of reading research points to the development of fluent word recognition skills as the biggest difficulty that students face in learning to read (Share & Stanovich, 1995). In particular, theories of word recognition (Ehri, 1998; Fuchs & Deno, 1991; Share & Stanovich, 1995) suggest that struggling readers have difficulty learning to recognize words as whole orthographic units or by using phonetic cues. Although we do not yet know the conditions required to prevent word recognition difficulties for all students, we do know that beginning readers benefit from systematic, explicit instruction in phonemic awareness and decoding skills (Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; Torgesen, Wagner, & Rashotte, 1997; Vellutino et al., 1996).

Reading Failure Among Hispanic Students

The number of children with limited English proficiency in U.S. public schools has risen dramatically in the past 20 years and continues to grow (August & Hakuta, 1997). These students make up about 5.5% of all public school students, with more than 70% speaking Spanish as their first language. Young Spanish-speaking students in U.S. schools have lower levels of reading achievement in English than other students (Fitzgerald, 1995) and are about twice as likely as non-Hispanic Whites to be reading below average for their age (Snow, Burns, & Griffin, 1998). Many Spanish-speaking students trail behind their classmates academically throughout elementary school and are referred in disproportionate numbers for special education services (Ortiz & Graves, 2001). Between 1976 and 1994, the percentage of Hispanic children identified as learning disabled increased from 24% to 51%. Given the low

Address: Barbara Gunn, Oregon Research Institute, 1715 Franklin Blvd., Eugene, OR 97403-1983; e-mail: barbarag@ori.org

levels of reading achievement among Spanish-speaking students, and the long-lasting, negative consequences of reading failure, ensuring that these students learn to read should be a high priority.

Reading and Behavior Problems

In addition to difficulties with word recognition, some children who struggle with reading have coexisting behavior problems. McGee, Prior, Williams, Smart, and Sanson (2002) found significant predictive pathways from early reading ability to later reading ability and from early reading ability to later school and attentional difficulties. Similarly, Hinshaw (1992) found a close relationship between later school difficulties, literacy, and behavior. Although the direction of the relationship is still unclear (Patterson, DeBaryshe, & Ramsey, 1989), there is some evidence that improving reading achievement may reduce aggressive behavior. Kellam, Mayer, Rebok, and Hawkins (1998) evaluated the effects of two preventive interventions with 1,196 first-grade students. One intervention focused on reducing aggressive behavior and one focused on improving reading achievement. Kellam et al. found that gains attributable to the reading intervention led to significant reduction in teachers' ratings of aggressive behavior. However, there was no increase in reading achievement due to the reduction of aggressive behavior attributable to the behavior intervention. Thus, effective reading instruction may be one element of an effort to prevent behavior problems.

Supplemental Reading Instruction

Despite the existing potential for ensuring reading success and significantly reducing the need for remedial services among struggling readers, practical difficulties often stand in the way of providing optimal reading instruction. Teachers are hard-pressed to find enough time in the day to teach the wide range of curricula required by districts and states. Reading instruction is complicated by children who enter school without the foundational literacy skills typically acquired in the preschool years (Biemiller, 1999; Whitehurst & Lonigan, 1998), by a burgeoning population of children who speak English as a second language, and by children with behavior problems. Coupled with the challenges and immense importance of teaching children to read in a relatively brief time frame, supplemental reading instruction is a promising approach for helping students at risk for reading difficulty develop essential literacy skills without missing important classroom instruction.

Efficacy of Supplemental Reading Instruction

Studies have documented the value of supplemental reading instruction for young English-speaking children at risk for reading difficulties (Foorman et al., 1998; O'Connor, 2000; Torgesen et al., 1997; Torgesen, Wagner, Rashotte, & Herron, 1999; Vellutino et al., 1996). Findings from these studies consistently indicate that children who received supplemental instruction in word-level reading skills and comprehension strategies in small, homogeneous groups improved their reading skills more than did children who began at similar skill levels but did not receive extra instruction.

Although we know less about the effectiveness of supplemental instruction for Spanish-speaking children, research indicates that English language learners and native English speakers follow similar paths in the development of early literacy skills (Lindsey, Manis, & Bailey, 2003). Further, findings indicate that English language learners can learn phonemic awareness and word identification skills in English at the same rate as native English speakers (Gersten & Geva, 2003). A small number of intervention studies have demonstrated that supplemental instruction can improve reading outcomes for Spanish-speaking children and that level of English oral language proficiency is not a factor in their ability to benefit from English reading instruction. For example, Linan-Thompson and Hickman-Davis (2002) found that low-SES second-grade Spanish-speaking children who received explicit and systematic supplemental reading instruction improved their English reading skills as much as native-English-speaking children. Similarly, Quiroga, Lemos-Britton, Mostafapour, Abbott, and Berninger (2002) found that Spanish-speaking first-grade children who received training in phonemic awareness and the alphabetic principle improved in English word-level reading beyond the level expected on the basis of their Spanish and English oral language skill. Based on these findings, supplemental reading instruction seems a viable approach to boost the reading achievement of Spanish-speaking students, and it does not need to wait until they become fluent in English. However, more evidence is needed to determine if supplemental instruction is equally effective across grade levels and in settings where children have a wide range of English language proficiency.

Purpose of the Study

The purpose of this study was to compare the effects of supplemental versus no supplemental instruction on the reading achievement of a diverse sample of students at risk for reading difficulty. Given the wide-ranging demographics and instructional needs of children in classrooms across the country, this study was designed to include a sample of children with the range of behavior and early literacy deficits that have been shown to affect reading outcomes. We included students with behavior problems because of their increasing numbers in the classroom (Walker, Ramsey, & Gresham, 2004) and because of the frequent coexistence of difficulties with reading and problem behavior (Hinshaw, 1992). Likewise, we included Hispanic students because of their high rates of academic failure and their increasing numbers in American classrooms. Al-

though the sources of reading difficulty differed, we hypothesized that supplemental reading instruction that used explicit instruction to develop word recognition skills, accompanied by clear feedback, active engagement, and cumulative review, would help students at risk for reading difficulties develop foundational reading skills. This article describes the effects of a supplemental reading program for a sample of kindergarten through Grade 3 (K–3) students who were at risk for reading difficulties. The target population included Hispanic and non-Hispanic students, and students with and without behavior problems. Instruction was part of the Schools and Homes in Partnership Project (SHIP), a community-based intervention that provided parents with parenting classes and early elementary students with reading instruction and social skills training. The study took place in four communities, three of which had large Mexican American populations. (See Smolkowski et al., 2005, for results related to the parenting skills training and social skills intervention.)

For the supplemental instruction, we used *Reading Mastery* (Engelmann & Bruner, 1988) and *Corrective Reading* (Engelmann, Carnine, & Johnson, 1988). We used these curricula for the intervention because they focus on the development of foundational word recognition skills identified as essential to skilled reading (Rayner et al., 2001) and because they incorporate the frequent opportunities for practice and review that help students learn and remember new skills. We hypothesized that the explicit instruction, clear feedback, active student engagement, and cumulative review that characterize *Reading Mastery* and *Corrective Reading* would help struggling readers by providing the scaffolding that would help them focus on important information and practice new skills. Both programs have been evaluated in whole-class and small-group conditions (G. L. Adams & Engelmann, 1996; Stahl & Miller, 1989). Students received 30 minutes of supplemental instruction daily for 2 years.

Earlier papers reported the effects of reading instruction for a subsample of the students included in the present report. One paper reported effects at the end of the intervention (Gunn, Biglan, Smolkowski, & Ary, 2000); a second article reported effects 1 year after the intervention ended (Gunn, Smolkowski, Biglan, & Black, 2002). At the end of the 2-year intervention, students who received the supplemental instruction performed better on measures of word attack, word identification, oral reading fluency, vocabulary, and reading comprehension. One year after the intervention, students in the supplemental instruction group still showed greater improvement in letter–word identification, word attack, and oral reading fluency than comparison students did. For reading comprehension, there was no overall effect for instruction but there was a significant interaction with ethnicity, indicating that effects on comprehension were still detectable for Hispanic students, but not for non-Hispanic students. This article goes beyond these reports in three ways. First, it reports results for a larger sample of students, as we added four schools to the study. Second, it examines the effects of instruction over a 4-year period, including 2 years after the end of instruction. Third, it provides a random coefficients analysis (Nich & Carroll, 1997; Singer & Willett, 2003) of the data so that we can examine growth in reading skill in a single analysis over 4 years.

Method

Design

We screened Hispanic and European American K–3 students in 13 schools across four Oregon communities on measures of reading skill and aggressive social behavior. We invited families to participate in the study if their child met criteria in a least one of these areas. Those who agreed to participate were randomly assigned to receive or not receive a comprehensive intervention that had three components: (a) 30 minutes daily of supplemental reading instruction, (b) parent training, and (c) a social skills intervention. We provided the intervention over a 2-year period, conducting a comprehensive assessment of students' reading skills and social behavior before the beginning of the intervention and in the spring of each school year for 4 years. Thus, the Time 1 (T_1) assessment occurred before intervention began, the Time 3 (T_3) assessment occurred immediately after intervention, and the Time 5 (T_5) assessment occurred 2 years following the end of the intervention.

Four communities participated in the project. In 1997, the school districts based in these communities served 3,722 students in Community A, 4,980 students in Community B, 3,647 students in Community C, and 11,227 students in Community D. In the fall of 1997, the proportion of Hispanics enrolled in each school district was 30.9% in A, 25.0% in B, 59.7% in C, and 4.6% in D. All the public elementary schools within each of the four communities were eligible to participate. One of 5 schools in Community A, 5 of 6 schools in Community B, 4 of 4 schools in Community C, and 4 of 16 schools in Community D participated.

Participants

Screening Procedures. We screened 4,004 students from a population of 4,508 K–3 students in the 14 participating elementary schools to identify students who showed reading deficits or aggressive social behavior. During the spring of the year before intervention, students in kindergarten and first and second grades received screening. In the fall of the first intervention year, new kindergarten students received screening, along with new students who had transferred into Grades 1, 2, or 3 in participating schools.

The sample included all K–3 students rated by their teachers as high in aggressive behavior or as performing below

grade level on the screening measures of early literacy skills. A total of 438 students met the screening criteria. From those meeting the criteria, we recruited 359 families. Of the families recruited, 28 moved or dropped out of the study before randomization, and another 32 moved, did not qualify for the reading intervention (i.e., they could read above grade level at pretest), or were dropped for other reasons (e.g., one control was contaminated), leaving 299 participants. Of these students, 159 (53.2%) were Hispanic and 140 (46.8%) were non-Hispanic. The study had 161 (53.8%) boys and 138 (46.2%) girls. At the outset of the study, the distribution of students across grades was K, 51 (17.1%); Grade 1, 87 (29.1%); Grade 2, 90 (30%); and Grade 3, 71 (23.7%). Table 1 presents the student characteristics.

School staff contacted the parents of students who met screening criteria to ask if they were interested in project staff talking to them about the project. Project staff called those who agreed and scheduled a home visit to describe the project in detail and to invite families to participate. Parents who agreed to participate signed an informed-consent form and provided demographic information. Parents also provided information about their ethnic identity, nativity, and language use. These data indicated that 94% of the Hispanic parents were of Mexican heritage, 5% were from Central America, and the remainder were from other Latin American countries. About 9% were born in the United States, 85% in Mexico, and the remainder in another Latin American country; 84% of Hispanic parents spoke only or mostly Spanish. The non-Hispanic parents were European American and spoke English.

Baseline Assessments and Randomization Procedures

Baseline assessments were used instead of the screening measures for randomization and for channeling students into specific intervention components. Prior to randomization, participating students were given the Word Attack and Word Identification subtests of the *Woodcock-Johnson Tests of Achievement* (WJ-R ACH; Woodcock & Mather, 1990) and a measure of oral reading fluency to measure their grade-level reading ability. Their teachers completed the *Walker-McConnell Test of Social Skills* (Walker & McConnell, 1988). All students who scored above 3.0 on the measure of social skills were grouped by grade and ethnicity (Hispanic or non-Hispanic), within community. They were rank ordered by their reading ability and then randomly assigned to condition, beginning with the poorest pair of readers. If single students remained from any group, they were matched across groups by reading score and randomly assigned to condition. The remaining students were matched by their scores on the *Walker-McConnell* and similarly assigned to condition. Students in the intervention condition who were below grade level on at least two of

TABLE 1. Number of Participants by Condition, Selection Criteria, Ethnicity, and Grade

Participant group/grade	Ethnicity by selection						
	Hispanic			Non-Hispanic			
	Reading	Aggression	All	Reading	Aggression	All	All
Intervention							
K	9	1	10	8	7	15	25
Grade 1	19	7	26	5	8	13	39
Grade 2	14	5	19	3	23	26	45
Grade 3	18	4	22	4	13	17	39
All	60	17	77	20	51	71	148
Control							
K	15	2	17	3	6	9	26
Grade 1	20	9	29	3	16	19	48
Grade 2	10	10	20	5	20	25	45
Grade 3	13	3	16	5	11	16	32
All	58	24	82	16	53	69	151
All							
K	24	3	27	11	13	24	51
Grade 1	39	16	55	8	24	32	87
Grade 2	24	15	39	8	43	51	90
Grade 3	31	7	38	9	24	33	71
All	118	41	159	36	104	140	299

the baseline reading measures were eligible to receive the supplemental reading instruction. On the Woodcock subtests, "below grade level" was defined as a grade-equivalent score below the average grade-level score in the norming sample. For oral reading fluency, "below grade level" was defined as students who read below the target rate for their grade level on the norms established by the *Dynamic Indicators of Basic Early Literacy Skills* (DIBELS; Good, Kaminski, Simmons, & Kame'enui, 2001).

Intervention Components

There were three intervention components: supplemental reading instruction, parent training, and social behavioral interventions.

Supplemental Reading Instruction. One hundred forty-eight students received the supplemental reading instruction. This sample included 80 students who were initially screened into the study on the basis of poor reading skills and 68 students who were initially screened into the study on the basis of aggressive social behavior and who also had poor reading skills (defined as below-grade-level performance on two or more of the reading baseline measures). Seventeen participants (6.6%) received special education services for reading, and 27 (10.5%) received Chapter services for reading.

The primary emphasis of the reading program in this study was the development of fluent word recognition skills through instruction in phonemic awareness and phonics, with practice reading decodable text. Students in the experimental condition received 6 to 7 months of supplemental reading instruction in their first year in the study and a full academic year (i.e., 9 months) of instruction in the second year. Instruction in the first year was shorter because of the time required to recruit families and conduct baseline assessments. During the summer between the first and second years of the intervention, 76 students (34 girls and 42 boys) attended a 5-week summer school in their community. They received 30 minutes of reading instruction 3 days a week using the same instructional methods and curriculum used during the school year.

Intervention students received the supplemental instruction as a pullout program during the school day at a time that their teachers determined would not interfere with key classroom instruction. All participants, treatment and control, also received daily reading instruction in their classrooms. Our interviews with the teachers revealed that their approaches to teaching reading varied widely. However, random assignment of treatment and control children within the same classroom should control for that variability and maintain the internal validity of comparisons between conditions.

Nine instructional assistants (IAs) hired from the project communities provided the supplemental instruction. The IAs were an asset to the project because they understood their community norms and were able to relate well to participants. Three assistants were certified teachers with 5 to 7 years of teaching experience. Seven of the nine had some previous experience working with elementary school students in small-group or tutorial settings. Two assistants spoke Spanish and English.

In the month before the intervention, the IAs received 10 hours of training on teaching lessons, motivating students, and managing students' behavior. The training also included an overview of the research findings on reading acquisition (M. J. Adams, 1990; Anderson, Hiebert, Scott, & Wilkinson, 1985; Snow, Burns, & Griffin, 1998; Stanovich, 1986). One of the authors served as the trainer.

Students were tested and placed in *Reading Mastery* if they were beginning readers in first or second grade. *Reading Mastery* initially teaches beginning readers phonemic awareness and sound–letter correspondence, and then teaches them how to sound out and blend words that they practice reading in decodable text. Third and fourth graders learned with *Corrective Reading*. Developers designed this program for older students who lack basic decoding skills or read below grade level. The program provides explicit instruction in sound–letter correspondence and spelling, with an emphasis on building fluency. New sounds are introduced at a slightly faster rate than in *Reading Mastery*, and the stories are geared to the interests of older students. About 5 to 7 minutes were spent daily on phonics and 10 to 15 minutes on word reading and spelling; and the remainder of the 30 minutes was spent on repeated reading of passages to build fluency and accuracy. In both programs, students received instruction in groups of two or three. Three students who could not participate in groups because of scheduling limitations received one-to-one instruction. Students in both programs usually completed one lesson a day, although the IAs spent more time, if needed, with the Spanish-speaking students to explain unfamiliar English vocabulary and to develop their background knowledge for comprehension.

To document fidelity of implementation, we observed the instructional assistants weekly during the first month of supplemental instruction and twice a month thereafter. One of the researchers or an assistant (both former teachers of reading) observed the instruction with a copy of the lesson plan and documented how closely the IAs followed the lesson plans. Observers also kept a tally of student errors and teachers' corrective feedback. Across observations, lessons were followed with 90% to 100% fidelity. The supervisors met individually with the IAs after the lesson to give them feedback on their instruction and to discuss questions or concerns with particular students. The observers taught a subsequent lesson, if needed, to demonstrate a particular instructional approach. Supervisors and IAs also met as a group twice a month to practice and refine instructional approaches and to discuss the progress of individual students.

Parent Training. All parents were offered the *Incredible Years* parent training program (Webster-Stratton, 1992a).

The program was provided in 12 to 16 sessions. Groups of 5 to 14 parents met weekly for 2.25 hours with two facilitators. During each session, parents viewed videotaped vignettes of parent–child interactions, discussed effective parenting methods, and role-played preferred strategies. Assignments to practice new skills were given each week. Childcare and dinner were provided. Groups were conducted in Spanish or English, depending on the parents' language preferences. Sixty-two percent of parents attended one or more sessions, but only 43% attended six or more sessions. Participants completed an average of 5.88 parent-training sessions (SD = 6.18).

Social Behavior Interventions. Two programs were used to teach students how to manage their behavior in classrooms and in interactions with peers outside of classes. Twenty-seven students received *Contingencies for Learning Academic and Social Skills* (CLASS; Hops & Walker, 1988). This program is designed to reduce acting-out behaviors of young children by teaching and reinforcing appropriate classroom behaviors to individual students. In Phase 1 (1–5 days), a trained consultant worked with the target child by awarding points and praise for target behavior. In Phase 2, the teacher took over from the consultant. Because CLASS requires direct teacher involvement for successful maintenance of the skills introduced by the consultant, some teachers chose not to implement this intervention with participating students. Students of those teachers (n = 99) received the *Dina Dinosaur's Social Skills and Problem-Solving Curriculum* (Webster-Stratton, 1992b) from a trained project interventionist. The program uses puppets and videotaped modeling to teach appropriate classroom and social behavior to small groups of children. Dinosaur School met after school for approximately 20 sessions. The groups ranged in size from 4 to 10 children. Each group had two adult leaders. Thirty-three percent of participants did not attend Dinosaur School. Out of approximately 30 sessions, 63% of participants attended fewer than 15.

Measures

Screening. We asked teachers to rate the frequency of specific behaviors in the previous 2 months for each of their students with the 25-item Aggression scale of the Teacher Rating Form (TRF) of the *Child Behavior Checklist* (CBCL; Achenbach, 1991). The TRF did not require training to complete, but teachers received payment to compensate them for their time. We measured early literacy skill with tests geared to the students' age and grade at the time of screening. In kindergarten, the screening measures were the letter naming fluency (LNF), phoneme segmentation, and onset fluency tasks from the DIBELS (Good, Kaminski, Laimon, & Johnson, 1992). In Grade 1, the measures were phoneme segmentation fluency and two 1-minute samples of oral reading fluency, and in Grade 2, the measures were three 1-minute samples of oral reading fluency. Averaging the raw scores on the screening measures created a composite score for reading skill for students in each grade.

A two-stage procedure used these screening data to determine eligibility for the study. First, children were eligible for the study if they were above the 95th percentile (T score of 67) on the Aggression scale of the CBCL. Second, if students did not meet this criterion on aggressive behavior, they were eligible based on the reading tests, if they scored in the bottom 5% of the distribution in their grade level in their community.

Annual Assessments. Assessors who were unaware of the child's intervention status administered all tests. At baseline and at the end of the first year of intervention, we measured students' growth in word-level reading skills. At the end of the second year of intervention and at each remaining assessment, we also measured their growth in vocabulary and comprehension. We used four subtests of the *Woodcock-Johnson Revised Tests of Achievement* (WJ-R; Woodcock & Mather, 1989) to measure growth in reading skill. The Letter–Word Identification subtest measured the ability to identify letter names and read irregular words. Word Attack measured the ability to use phonic and structural analysis to decode words. The Passage Comprehension subtest required students to read passages and identify key words that were missing. The Reading Vocabulary subtest required students to read words and supply appropriate meanings. We analyzed and reported the WJ-R scores as W scores. According to the examiner's manual for the WJ-R (Woodcock & Mather, 1989, 1990), W scores are based on an equal interval scale that is centered on a value of 500. In addition, Table 2 presents the means and standard deviations for T_1, T_3, and T_5 as standard scores.

Researchers measured oral reading fluency by having students read three grade-level passages (Markell & Deno, 1997), using the same procedure used for the initial screening. Oral reading fluency is defined as the speed at which students read connected text. It is expressed as the number of words read correctly per minute. We measured reading fluency at each time point because several theoretical frameworks suggest that oral reading fluency is a prerequisite skill for the development of comprehension and serves as an indicator of overall reading ability (Fuchs et al., 2001; LaBerge & Samuels, 1974; Stanovich, 1980).

For example, Fuchs and colleagues (Fuchs, Fuchs, & Maxwell, 1988) found that a measure of oral reading fluency correlated .91 with the Reading Comprehension subtest of the *Stanford Achievement Test* (Harcourt Brace, 1996) among middle school students with a reading disability. Similarly, Jenkins, Fuchs, Espin, van den Broek, and Deno (2000) found that measures of oral reading fluency and comprehension were highly correlated.

Language Proficiency. Project assessors who spoke Spanish and English assessed Spanish and English language

TABLE 2. Means and Standard Deviations, Standard Scores, for T_1, T_3, and T_5

Dependent variable	Non-Hispanic			Hispanic		
	T_1	T_3	T_5	T_1	T_3	T_5
Woodcock-Johnson Letter–Word ID Standard Score						
Control group						
M	80.62	93.17	90.42	63.25	77.14	89.23
SD	19.88	19.06	19.66	15.81	22.11	16.40
n	66	59	36	79	65	60
Intervention group						
M	83.61	97.33	94.12	61.99	87.61	94.78
SD	22.62	24.07	19.16	19.48	25.16	19.42
n	70	55	41	75	66	58
Woodcock-Johnson Word Attack Standard Score						
Control group						
M	91.88	97.10	95.03	85.89	88.11	93.18
SD	11.97	16.93	20.50	10.37	13.71	16.20
n	66	59	36	79	65	60
Intervention group						
M	94.53	106.04	98.20	82.29	97.15	100.31
SD	14.98	19.26	20.55	11.85	16.77	19.83
n	70	55	41	75	66	58
Oral Reading Fluency Correct Words Per Minute						
Control group						
M	22.25	60.80	73.60	6.11	31.14	69.32
SD	31.67	42.63	35.39	15.55	26.27	33.49
n	66	59	38	79	61	59
Intervention group						
M	26.03	68.93	80.03	6.66	43.02	79.45
SD	36.53	46.87	35.68	16.02	34.65	40.37
n	70	54	41	76	65	59
Woodcock-Johnson Vocabulary Standard Score						
Control group						
M	—	92.00	85.36	—	75.40	76.76
SD	—	19.32	16.37	—	14.12	12.73
n	—	49	36	—	57	59
Intervention group						
M	—	95.59	91.80	—	78.54	77.17
SD	—	19.87	18.85	—	15.77	16.14
n	—	46	41	—	59	58
Woodcock-Johnson Comprehension Standard Score						
Control group						
M	—	91.53	88.56	—	75.50	82.44
SD	—	17.55	15.95	—	17.40	12.72
n	—	49	36	—	56	59
Intervention group						
M	—	95.33	92.46	—	82.02	84.47
SD	—	23.44	20.62	—	21.63	16.58
n	—	46	41	—	59	58

Note. Woodcock-Johnson = *Woodcock-Johnson Revised Tests of Achievement* (Woodcock & Mather, 1989).

proficiency among the Hispanic students. Before baseline assessments, teachers identified the participating Hispanic children who spoke some Spanish. The assessors, who were native Spanish speakers, spoke with the children in Spanish and in English to determine which language they understood and preferred to use. The conversations, which focused on familiar topics, such as the child's everyday activities and interests, helped the assessors to confirm the teachers' information and to determine the child's proficiency with Spanish and English. Children who spoke only in Spanish with the assessors and who did not appear to understand any English were identified as Spanish speakers. Children who spoke only Spanish were assigned to bilingual IAs for the supplemental instruction. The IAs spent more time with the students, as needed, to explain unfamiliar English vocabulary and develop their background knowledge. The information was recorded on the child's baseline reading test profile and entered into the database. At the outset of the intervention, 17 of the Hispanic children spoke only Spanish and the rest spoke Spanish and English. All of the non-Hispanic students spoke only English.

Sample Maintenance

Parents received payment for their participation in the study. At each assessment, they received $30 for completing a parent questionnaire and a $10 gift certificate if they returned the questionnaire within 10 days. They also received $15 for providing data about social behavior in a series of three brief phone interviews at each time point. We paid teachers $30 for completing the screening measures for all students in their class and $5 for each child they rated on the *Walker-McConnell*.

Project staff made extensive efforts to maintain families in the sample. Although students began the study in 14 schools, they had dispersed to 78 schools by the T_5 assessment. During the nonassessment phases of each study year (September through March), staff mailed families a newsletter with a gift certificate to a grocery store in their community. In April, they received another newsletter with a reminder of the approaching assessment. This letter also told families how many years they had participated in the project. All mailings, including newsletters and birthday cards, included a toll-free number that participating families were encouraged to use to update their address and phone number. As an incentive, we gave families $10 each time they sent us new information.

At each assessment, parents were asked to give us telephone numbers for family or friends whom we could contact if their information was no longer current and we could not reach them. This information was used if mailings were returned with no forwarding address or if we received a card from the post office indicating that a family had moved. Once a family was contacted, we asked them for written permission to test their child at the new school. This was followed by a contact with the school principal to explain the study and ask permission to schedule an assessment with the student's teacher.

Attrition

Reading data were available from 190 students at all five time points. Table 2 presents the number of students tested at baseline (T_1), at the end of the intervention (T_3), and 2 years after the intervention (T_5). This distribution of missing data did not differ by intervention condition. A series of chi-square analyses of the relationship between missing data and condition at each time point showed that the number of missing cases did not differ by condition at any time point.

We then examined whether there was an interaction between condition and the number of time points at which the student had missing data, for each of the T_1 reading measures (letter–word identification, word attack, and oral reading fluency). A significant interaction would indicate a systematic difference between conditions in the skill level of the students who did not provide data and would threaten the internal validity of the study. These analyses did not indicate any significant interactions.

Next, we tested for interactions between condition and missing data on T_1 reading scores at each individual time point. That is, we analyzed whether those who had missing data at any given assessment differed between conditions in their T_1 scores on any reading measure. Here, too, significant results would threaten internal validity. There were no significant interactions for those missing data at T_2 or T_3. At T_4, there was one significant interaction for letter–word identification, $F(1, 230) = 8.098$, $p = .005$. It showed that, among the students who were missing data at this time point, the intervention condition had more students than the control condition who had scored low on this measure at T_1. The same interaction was found for letter–word identification at T_5, $F(1, 230) = 7.515$, $p = .005$.

Overview of Analysis

To model appropriately the repeated assessments nested within individuals, and individuals nested within treatment condition, we conducted a random coefficients analysis, also known as a random-effects regression, linear mixed model, or multilevel model (Kreft & de Leeuw, 1998; Murray, 1998; Nich & Carroll, 1997; Singer & Willett, 2003; Wallace & Green, 2002). The data were analyzed with SAS PROC MIXED (Littell, Milliken, Stroup, & Wolfinger, 1996; SAS Institute, 1999; Singer, 1998, 2002). The random coefficients model can be represented by two sets of equations, one that models within-person assessments and one that models individual students. This overview describes the standard multilevel model (Kreft & de Leeuw, 1998; Singer & Willett, 2003) and then discusses several extensions used in the analyses reported below.

The first equation represents the within-person model:

$$Y_{ij} = \pi_{0j} + \pi_{1j}T_{ij} + \boldsymbol{r}_{ij}$$

The terms Y_{ij}, T_{ij}, and r_{ij} represent the dependent variable, the effect of time, and random error for each assessment occasion *i* within each individual *j*. From this model, we obtain estimates of π_{0j} and π_{1j}, the intercept and slope, for the *j*th individual's trajectory across time. The specification of T_{ij} determines the type of slope and the placement of the intercept. We typically code T_{ij} to model linear growth and to place the intercept at the first measurement occasion. That is, we set T_{1j} to 0 for every participant *j* and increment T_{ij} equally at each assessment occasion *i* thereafter, which means that we set T_{2j} to 1, T_{3j} to 2, and so on. With this specification, then, the slope parameter, π_{1j}, estimates the average increase in the dependent variable per measurement occasion for individual *j*.

The next two equations represent the between-person intercept and slope, respectively, and allow us to test the effect of condition. We coded $C_j = 0$ for control students and 1 for intervention students:

$$\pi_{0j} = \beta_{00} + \beta_{01}C_j + \boldsymbol{u}_{0j}$$
$$\pi_{1j} = \beta_{10} + \beta_{11}C_j + \boldsymbol{u}_{1j}$$

The first equation models the average control-group intercept, β_{00}, and the deviation from that control-group average due to condition, β_{01}. The second equation estimates the average control-group growth, β_{10}, and the deviation from that normative growth due to the intervention, β_{11}. Thus, the magnitude and statistical significance of the β_{01} and β_{11} estimates represent the effect of supplemental reading instruction on the intercept and slope of the reading measures; these terms represent our primary hypotheses. The equations above also include individual-level random variation around the intercept, $\boldsymbol{u}_{0j}$, and slope, $\boldsymbol{u}_{1j}$.

Substituting the second two equations into the first and rearranging terms, we obtain the following model, with the fixed and random terms grouped, which we estimate with SAS PROC MIXED (Littell et al., 1996; Singer, 1998, 2002; Singer & Willett, 2003):

$$Y_{ij} = (\beta_{00} + \beta_{01}C_j + \beta_{10}T_{1j} + \beta_{11}C_j\,T_{1j}) + (\boldsymbol{u}_{0j} + \boldsymbol{u}_{1j}T_{ij} + \boldsymbol{r}_{ij})$$

The model presented so far places the intercept at the initial assessment occasion, T_1, before intervention. When the intercept is set at this point, differences between conditions on the intercept are not expected, because students were randomized to condition. However, one can set intercept at other time points (Singer & Willett, 2003). By setting it at T_3, we test for effects of the intervention at the end of intervention. By setting it at T_5 (by setting T_{1j} to −4, T_{2j} to −3, T_{3j} to −2, T_{4j} to −1, and T_{5j} to 0), we test for effects 2 years following the intervention. In what follows, we report results with the intercept at each of these time points, as each provides unique information about the effect of the intervention.

We also included a term for quadratic growth, T_{ij}^2, and its interaction with condition because we expected that a linear term might not adequately model the underlying data. When we found statistically significant quadratic growth, we also examined both intercept and slope differences at pretest, with intercept at T_1, and immediately after intervention, with intercept at T_3, as well as at T_5.

It is important to note that changes in placement of the intercept do not change the model. That is, two linear growth models that have different intercepts but are otherwise identical provide the same set of curves and the same model fit statistics. Only the intercept estimates differ. With a quadratic effect, changing the intercept will affect estimates of the slope, but different estimates still describe the same set of curves. They just do so from different reference points, namely, the slope at the given point of intercept. Thus, there is some redundancy in the testing of effects when we vary the intercept. These are not independent tests, but they allow us to pinpoint the effects of the intervention in time.

We examined a number of additions or modifications to this basic random coefficients model. The nature of this sample is complicated, with participants recruited by different criteria, approximately half our sample consisting of Hispanic families, and so on. The analysis included dichotomous terms for key background influences: gender, selection criteria, either poor reading or aggressive behavior, ethnicity, and grade level. We expected main effects and interactions with time for these factors, especially selection criteria, ethnicity, and grade level, but we were unsure about their interactions with treatment effects. For each of these background variables, then, we tested its main effect and its interactions with time, condition, and time by condition. The inclusion of these terms also reduces the impact of potential confounding effects, such as with ethnicity and selection criteria.

We removed nonsignificant effects from each model. We fixed the denominator degrees of freedom, however, to that which we specified for the full model to maintain unbiased *p* value estimates (Harrell, 2001): 268 for letter–word identification, word attack, and oral reading fluency, and 179 for passage comprehension and vocabulary. This conservative approach uses 1 degree of freedom for each variable or interaction, whether the variable is retained in the analysis or not. This is, in essence, a penalty for exploring each relationship and should lead to more robust findings.

As is the case with all longitudinal studies, some participants failed to provide data for one or more of the assessments. Maximum likelihood models with time as a random variable, such as the random coefficients model employed here, allow the use of all available data from all assessments, reducing bias and increasing power (Laird, 1988; Nich & Carroll, 1997). We assumed that the missing data were missing at random or ignorable. Random coefficients analyses generate appropriate models with such missing data—when the missing data mechanism does not depend on unobserved determinants (Little, 1995; Little & Rubin, 1987; Singer & Willett, 2003).

Results

Letter–Word Identification

Figure 1 depicts the results for letter–word identification *W* scores. As noted above, the presentation of results begins with analyses that placed the intercept at T_1 and at T_3, immediately after intervention, and then presents results for the intercept set at T_5. Results for T_1 and T_3 are not entirely new; they parallel those previously reported (Gunn, Biglan, Smolkowski, & Ary, 2000; Gunn, Smolkowski, Biglan, & Black, 2002). The present analysis, however, includes a larger sample of students, data from the final assessment, T_5, and a more sophisticated analysis method.

With intercept at T_1, we found a statistically significant slope by condition effect ($t = 2.82$, $p = .0052$). Intervention students gained faster than controls on letter–word identification. As expected, intervention students did not differ from controls at T_1 ($t = -0.79$, $p = .4331$). There was a significant quadratic effect ($t = -4.32$, $p < .0001$), which did not quite differ according to condition ($t = -1.93$, $p = .0545$). Because this latter effect could have mitigated against the value of the intervention and because the p value was very close to .05, we included it in the model.

We did not find a statistically significant condition effect with the intercept placed at T_3, the end of the intervention ($t = 1.71$, $p = .0887$). There was, however, a difference in slope, with letter–word identification growing faster among intervention students than among controls ($t = 2.62$, $p = .0092$).

To estimate the long-term effects, we placed the intercept at T_5, 2 years after the end of intervention. There was a statistically significant difference between conditions at the T_5 intercept ($t = 2.12$, $p = .0346$). This intercept difference at T_5 represented an effect size, Cohen's *d*, of 0.25 standard deviations (Cohen, 1987; Rosenthal & Rosnow, 1991). Due to the quadratic effect, the intervention students were no longer increasing at a rate greater than controls ($t = -0.60$, $p = .5461$). That is, intervention- and control-group trajectories, by T_5,

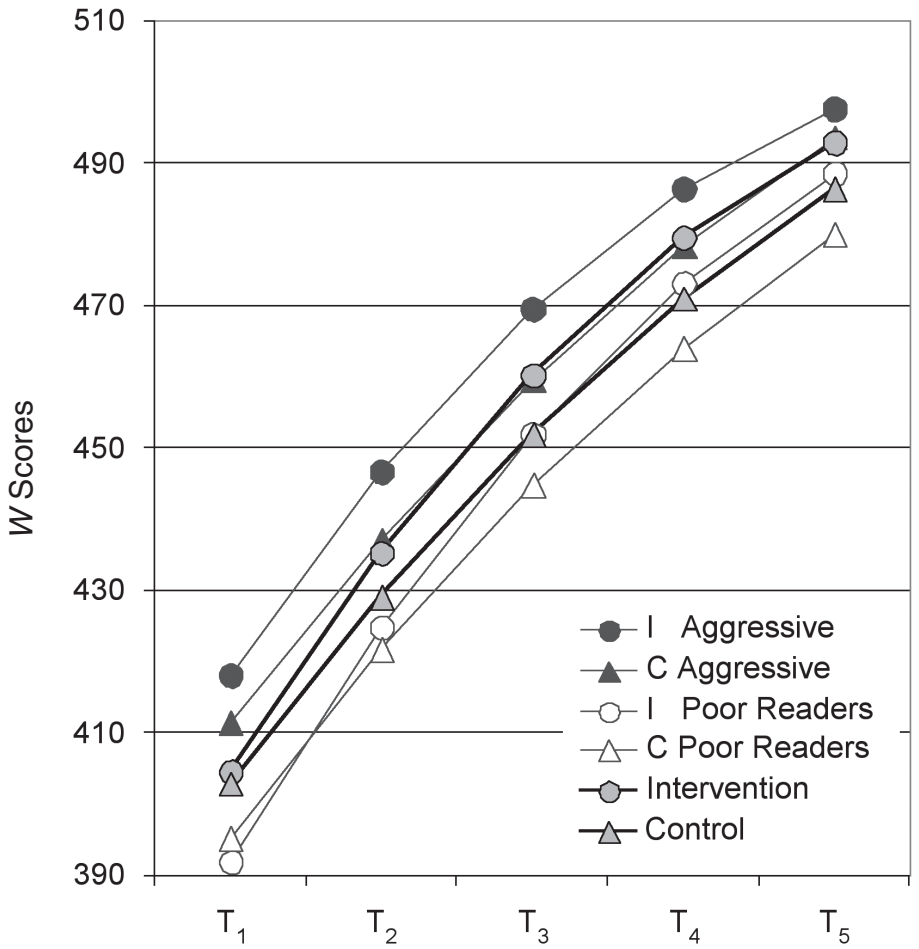

FIGURE 1. Growth curves for letter–word identification *W* score within each condition (dark lines) and for condition by selection, aggressive behavior or poor reading skills.

TABLE 3. Random Coefficients Model Estimate for Letter–Word Identification *W* Score with Intercept at T5

Effect	Estimate	Effect size	*t* or *Z* value[a]	*p* value
Fixed				
Intercept	477.05		144.17	< .0001
Condition[b]	8.52	0.245	2.12	.0346
Linear time	18.36	1.070	9.25	< .0001
Linear × Condition	−1.52	0.069	−0.60	.5461
Quadratic time	−1.77	0.500	−4.32	< .0001
Quadratic × Condition	−1.12	0.223	−1.93	.0545
Selection[c]	13.54	0.348	3.01	.0029
Condition × Selection	−4.45	0.083	−0.72	.4699
Linear × selection	−0.64	0.059	−0.51	.6128
Linear × Condition × Selection	−3.62	0.245	−2.12	.0349
Ethnicity[d]	2.68	0.090	0.78	.4342
Linear × Ethnicity	4.06	0.494	4.27	< .0001
Grade level[e]	2.82	0.106	0.92	.3561
Linear × Grade level	−12.15	1.655	−14.31	< .0001
Random				
Intercept	289.79		5.30	< .0001
Linear time	8.68		0.21	.8299
Quadratic time	1.14		0.50	.6187
Residual	254.67		14.60	< .0001
Covariance between intercept and time	−174.18		−5.50	< .0001
Covariance between intercept and quadratic	−46.61		−6.37	< .0001
Covariance between time and quadratic	1.39		0.15	.8810

[a] *t* value for fixed effects, Wald *Z* values for random effects. [b]Condition coded 1 for intervention, 0 for control. [c]Selection coded 1 for aggression, 0 for reading. [d]Ethnicity coded 1 for Hispanic, 0 for non-Hispanic. [e]Grade level coded 1 for Grades 2 and 3, 0 for kindergarten and Grade 1.

had become nearly parallel. See Tables 3 through 7 for model estimates.

The model included several additional statistically significant terms. We found a statistically significant time by condition by selection criterion term ($t = -2.12$, $p = .0349$) as well as a statistically significant selection criteria term ($t = 3.01$, $p = .0029$). The condition by selection and time by selection terms, however, were not significantly different. As indicated in Figure 1, the students selected by their aggressive behavior began higher than those students selected for their reading difficulties did. The intervention improved the slope of letter–word identification scores for poor readers but not for the aggressive students. By T_5, however, the intercepts differed between conditions for both groups, as the condition by selection interaction was not significant. We also found statistically significant effects at the T_5 intercept and by time for ethnicity and for grade level, but neither interacted with condition.

Word Attack

Figure 2 shows the slopes estimated by the random coefficients analysis for word attack *W* scores. As expected, at T_1 the groups did not differ ($t = 0.73$, $p = .4680$). However, students who received supplemental instruction had a higher slope than controls from their T_1 intercept ($t = 4.18$, $p < .0001$). The model estimated a statistically significant condition by quadratic slope ($t = -4.95$, $p < .0001$), which represents a concave-down shape for intervention participants.

Moving the intercept to T_3, we found statistically significant differences between conditions at that point ($t = 3.25$, $p = .0013$). Word attack scores from both groups of students continued to increase over time, with statistically significant growth for controls ($t = 16.83$, $p < .0001$) but no difference between conditions ($t = -0.28$, $p = .7761$). With the change in intercept, from T_1 to T_3, the quadratic effects remain identical to those presented above.

By the final assessment, 2 years after intervention, the treatment groups no longer differed ($t = 0.22$, $p = .8274$). Control group scores were still increasing ($t = 8.40$, $p < .0001$), but the scores from the intervention students, which were higher than controls at T_3, were increasing at a slower rate by T_5 ($t = -4.23$, $p < .0001$).

To summarize, word attack skills improved at a greater rate from T_1 for those who received instruction. Their greater

TABLE 4. Random Coefficients Model Estimate for Word Attack *W* Score

Effect	Estimate	Effect size	*t* or *Z* value[a]	*p* value
Fixed				
Intercept	489.27		150.68	< .0001
Condition[b]	0.80	0.025	0.22	0.8274
Linear time	11.45	0.972	8.40	< .0001
Linear time × Condition	−7.80	0.489	−4.23	< .0001
Quadratic time	−0.41	0.175	−1.51	.1319
Quadratic × Condition	−1.88	0.573	−4.95	< .0001
Selection[c]	8.71	0.549	4.75	< .0001
Ethnicity[d]	−2.67	0.088	−0.76	.4485
Condition × Ethnicity	7.46	0.180	1.56	.1205
Linear time × Ethnicity	0.79	0.102	0.88	.3776
Linear time × Condition × Ethnicity	2.86	.265	2.29	.0228
Grade level[e]	2.43	0.118	1.02	.3089
Linear time × Grade level	−5.09	0.951	−8.22	< .0001
Random				
Intercept	221.88		6.51	< .0001
Linear time	5.04		0.28	.7792
Quadratic time	0.32		0.33	.7447
Residual	109.02		14.51	< .0001
Covariance between intercept and time	−24.12		−1.41	.1581
Intercept and time				
Covariance between intercept and quadratic	−13.93		−3.73	.0002
Covariance between time and quadratic	0.01		0.00	.9975

[a]*t* value for fixed effects, Wald *Z* values for random effects. [b]Condition coded 1 for intervention, 0 for control. [c]Selection coded 1 for aggression, 0 for reading. [d]Ethnicity coded 1 for Hispanic, 0 for non-Hispanic. [e]Grade level coded 1 for Grades 2 and 3, 0 for Kindergarten and Grade 1.

rate of growth in word-attack skills leveled off at T_3, but their earlier, higher rate of growth in reading skill left them at a higher level than their controls at T_3. At T_5, however, growth in word attack had diminished for intervention students, so that control and intervention students no longer differed in mean word attack.

The analysis of word attack scores also involved a statistically significant slope by condition by ethnicity effect, and we found a simple main effect for ethnicity at T_1. As shown in Figure 2, Hispanic control students' scores began at a lower level than those of their non-Hispanic, control-group classmates ($t = -2.22$, $p = .0276$), and both Hispanic and non-Hispanic control groups improved at the same rate ($t = 0.88$, $p = .3776$). As expected, Hispanic students in the intervention condition did not differ at T_1 from Hispanic controls ($t = -1.13$, $p = .2591$). Non-Hispanic students in the intervention condition improved at a greater rate than their controls ($t = 4.18$, $p < .0001$), and Hispanic intervention students improved even more quickly ($t = 2.29$, $p = .0228$). With the intercept at T_3, we found that, on average, ethnicity no longer accounted for a difference in word attack *W* scores ($t = -1.67$, $p = .0954$), nor did it interact with condition ($t = 0.52$, $p = .6043$), but as reported above, the intervention conditions significantly differed. We did not find ethnicity main effects at T_5 ($t = -0.76$, $p = .4485$) or an interaction with condition ($t = 1.56$, $p = .1205$), but at T_5, we no longer found a difference between conditions.

For word attack scores, we also found effects for selection criteria and grade level. For selection criteria, we found only main effect ($t = 4.75$, $p < .0001$). For grade level, we found that older students started higher ($t = 12.82$, $p < .0001$), as we expected, but they improved at a slower rate ($t = -8.22$, $p < .0001$). By T_5, older students no longer differed from younger students ($t = 1.02$, $p = .3089$). Neither selection criteria nor grade level differed by condition.

Oral Reading Fluency

Figure 3 shows the growth curves for oral reading fluency. With the intercept at T_1, oral reading fluency scores and correct words per minute (CWPM) did not differ by condition ($t = 0.69$, $p = .4890$), and there was no quadratic trend. However, scores increased at a significantly faster rate for intervention than for control students ($t = 2.50$, $p = .0129$). Placing the intercept at T_3, we found significant group differences for condition ($t = 2.10$, $p = .0365$), and the difference had widened

TABLE 5. Random Coefficients Model Estimate for Oral Reading Fluency

Effect	Estimate	Effect size	*t* or *Z* value[a]	*p* value
Fixed				
Intercept	66.05		10.29	< .0001
Condition[b]	13.79	0.285	2.46	.0144
Linear time	17.03	1.614	13.95	< .0001
Linear × Condition	2.96	0.289	2.50	.0129
Selection[c]	24.10	0.478	4.13	< .0001
Linear × Selection	2.33	0.227	1.96	.0505
Ethnicity[d]	−10.94	0.396	−3.42	.0007
Gender[e]	−14.46	0.298	−2.58	.0104
Linear × Gender	−2.91	0.285	−2.46	.0145
Grade level[f]	27.43	1.119	9.67	< .0001
Random				
Intercept	1,984.46		10.16	< .0001
Covariance between intercept and linear time	326.01	8.63	< .0001	
Linear time	68.40		8.29	< .0001
Residual	157.35		17.57	< .0001

[a]*t* value for fixed effects, Wald *Z* values for random effects. [b]Condition coded 1 for intervention, 0 for control. [c]Selection coded 1 for aggression, 0 for reading. [d]Ethnicity coded 1 for Hispanic, 0 for non-Hispanic. [e]Gender coded 1 for boys, 0 for girls. [f]Grade level coded 1 for Grades 2 and 3, 0 for kindergarten and Grade 1.

TABLE 6. Random Coefficients Model Estimate for Passage Comprehension *W* Score

Effect	Estimate	Effect size	*t* or *Z* value[a]	*p* value
Fixed	475.26		235.37	< .0001
Intercept				
Condition[b]	4.38	0.288	2.09	0.0383
Linear time	9.56	0.578	4.19	< .0001
Linear × Condition	−1.28	0.124	−0.90	0.3703
Quadratic	−5.88	0.864	−6.26	< .0001
Selection[c]	13.22	0.850	6.16	< .0001
Linear × Selection	−3.09	0.295	−2.14	0.0340
Grade level[d]	2.23	0.146	1.06	0.2907
Linear × Grade level	−13.02	1.256	−9.10	< .0001
Random				
Intercept	146.72		5.68	< .0001
Covariance between intercept and time	−56.20		−4.19	< .0001
Linear time	45.97		3.52	0.0004
Residual	105.18		8.80	< .0001

[a]*t* value for fixed effects, Wald *Z* values for random effects. [b]Condition coded 1 for intervention, 0 for control. [c]Selection coded 1 for aggression, 0 for reading. [d]Grade level coded 1 for Grades 2 and 3, 0 for kindergarten and Grade 1.

by T_5 (t = 2.46, $p = .0144$). From the estimated curves, intervention students at T_1 read less than 2 CWPM faster than controls, but by T_5, they read almost 14 CWPM faster.

For oral reading fluency, we found no interactions with condition other than slope. However, we did find significant main effect differences between Hispanics and non-Hispanics ($t = -3.42$, $p = .0007$) and between older and younger students ($t = 9.67$, $p < .0001$). In addition, we found slope by gender ($t = -2.46$, $p = .0145$) and slope by selection criteria ($t = 1.96$, $p = .0505$) interactions, where female students and students selected because of reading difficulty grew faster than males and students selected because of their aggressive behavior.

TABLE 7. Random Coefficients Model Estimate for Vocabulary *W* Score

Effect	Estimate	Effect size	*t* or *Z* value[a]	*p* value
Fixed				
Intercept	478.86		194.79	< .0001
Condition[b]	3.39	0.247	1.79	.0751
Linear time	6.74	0.636	4.61	< .0001
Linear × Condition	–0.43	0.068	–0.49	.6232
Quadratic	–2.98	0.642	–4.65	< .0001
Selection[c]	10.78	0.737	5.34	< .0001
Ethnicity[d]	–6.84	0.468	–3.39	.0009
Grade level[e]	5.00	0.362	2.62	.0095
Linear × Grade level	–6.96	1.101	–7.98	< .0001
Random				
Intercept	156.90		7.51	< .0001
Covariance between intercept and time	–8.31		–1.16	.2449
Linear time	8.31		1.72	.0854
Residual	50.17		9.18	< .0001

[a]*t* value for fixed effects, Wald *Z* values for random effects. [b]Condition coded 1 for intervention, 0 for control. [c]Selection coded 1 for aggression, 0 for reading. [d]Ethnicity coded 1 for Hispanic, 0 for non-Hispanic. [e]Grade level coded 1 for Grades 2 and 3, 0 for kindergarten and Grade 1.

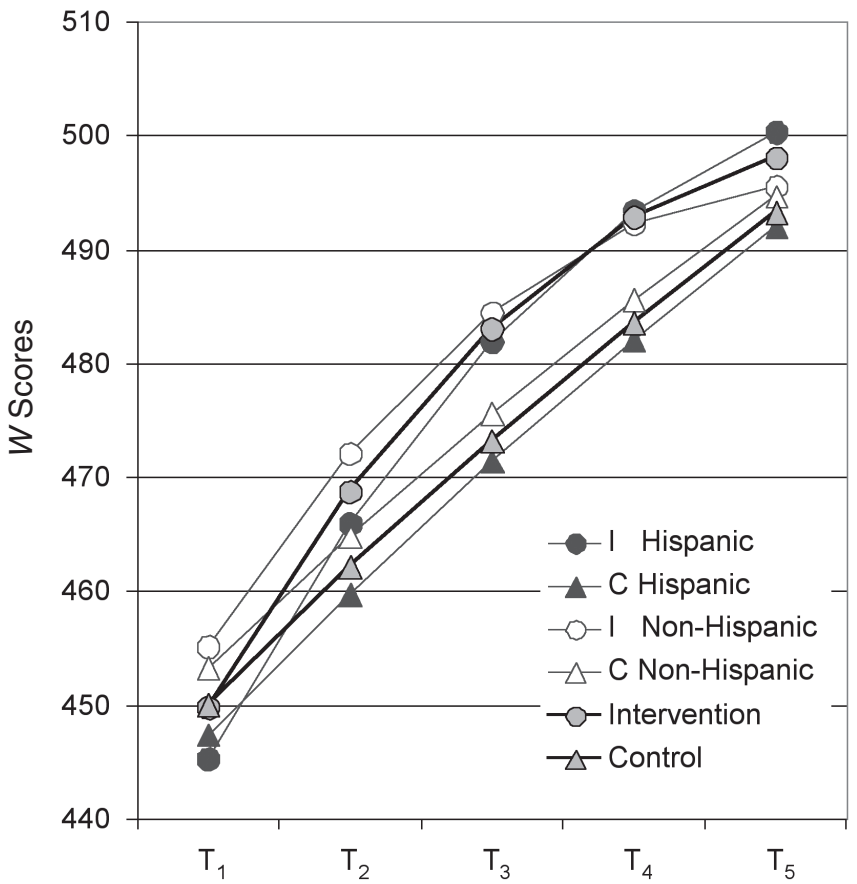

FIGURE 2. Growth curves for word attack *W* score within each condition (dark lines) and for condition by ethnicity.

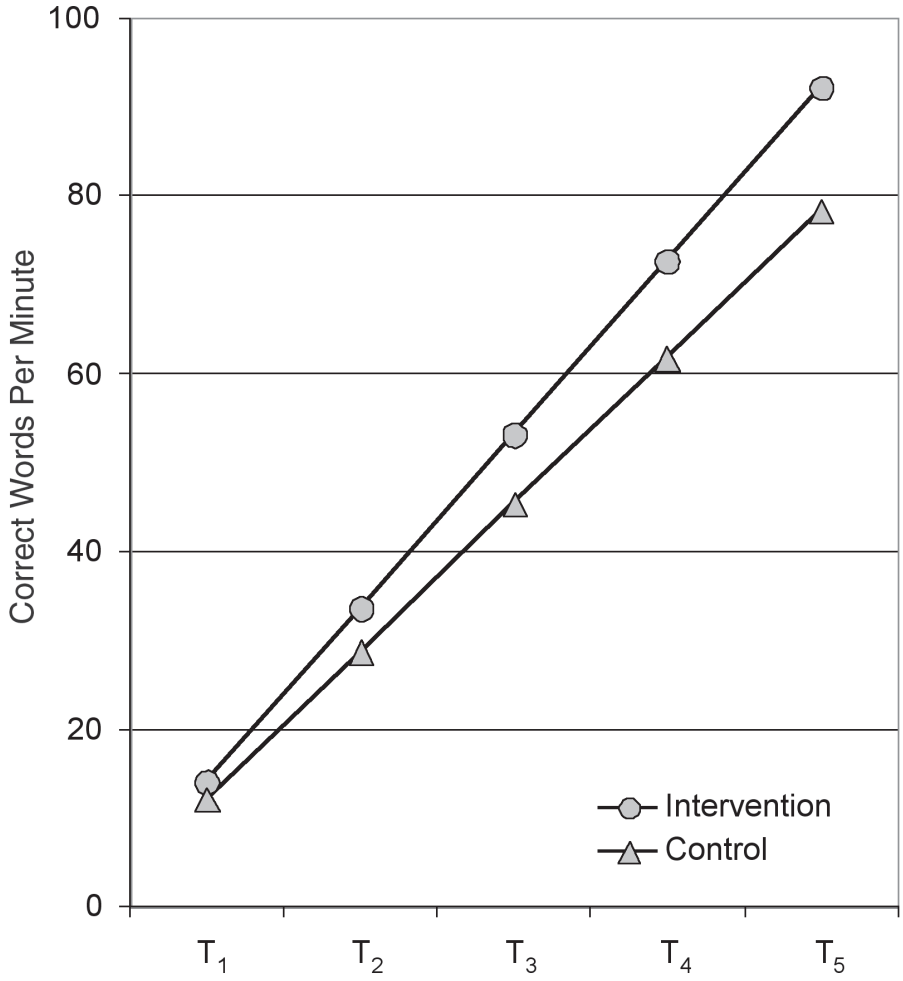

FIGURE 3. Growth curves for oral reading fluency within each condition.

Passage Comprehension

We collected all passage comprehension data after the end of supplemental reading instruction. Control students' scores increased from T_3 to T_5, and intervention students improved at the same rate ($t = -.90$, $p = .3703$). With the intercept at T_3, the students who received supplemental instruction scored higher than their controls ($t = 2.12$, $p = .0357$), as they did at T_5 ($t = 2.09$, $p = .0383$). For passage comprehension, we also found a quadratic effect ($t = -6.26$, $p < .0001$), giving curves a concave-down shape, but there was no interaction with condition. Thus, students' scores, in general, grew faster from T_3 to T_4, and leveled off from T_4 to T_5, as shown in Figure 4. Intervention students, however, scored higher across all time points.

Scores differed by grade level and selection criteria, but those differences did not involve condition. With the intercept at T_3, older students performed considerably better ($t = 8.51$, $p < .0001$) but increased at a slower pace ($t = -9.10$, $p < .0001$). Students selected into the study for their aggressive behavior also performed better ($t = 5.85$, $p < .0001$) and increased at a slower rate ($t = -2.14$, $p = .0340$).

Reading Vocabulary

The estimated growth curves for vocabulary scores looked very similar to those for passage comprehension. Students' scores increased ($t = 12.56$, $p < .0001$) with the intercept at T_3. The model included a significant quadratic term ($t = 4.65$, $p < .0001$), but no quadratic interaction with condition. The curves for control students, then, increased across time with a concave-down shape, showing a lesser rate of increase by T_5 ($t = 4.61$, $p < .0001$). Intervention participants scored higher at T_3 ($t = 2.02$, $p = .0446$), but not quite so high at T_5 ($t = 1.79$, $p = .0751$). Grade level, selection criteria, and ethnicity all had an influence on vocabulary scores, but none interacted with condition. As with comprehension, older students scored higher at T_3 ($t = 8.88$, $p < .0001$) but increased at a slower rate ($t = -7.98$, $p < .0001$). Students selected into the study due to

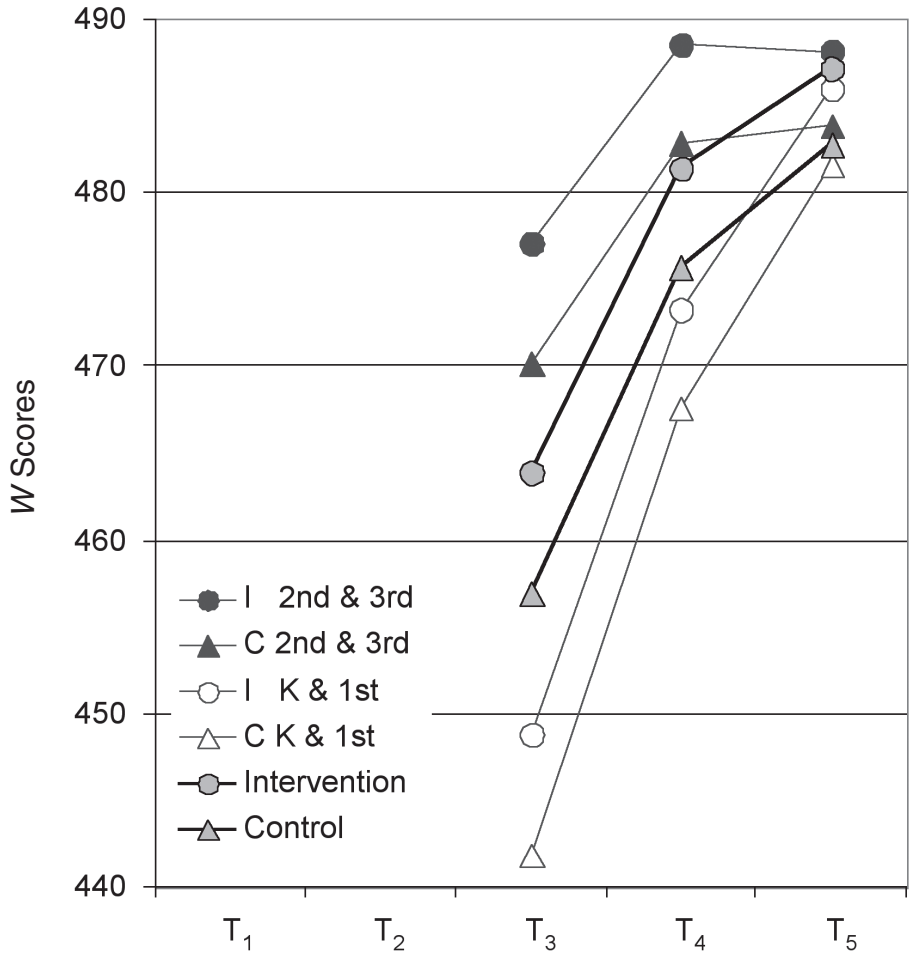

FIGURE 4. Growth curves for passage comprehension W score by each condition (dark lines) and separately for older and younger students each by condition.

their aggressive behavior performed better on average (t = 5.34, $p < .0001$) than those selected by their poor reading skills. Hispanic children performed worse, in general, than non-Hispanic children ($t = -3.39$, $p = .0009$).

Scoring Methods

The *Woodcock-Johnson Tests of Achievement* provide several different scoring methods, including raw scores, *W* scores, normal curve equivalent (NCE) scores, *T* scores, stanine scores, standardized scores, percentile ranks, and extended percentile scores. The literature is not entirely clear about the most appropriate scoring method to use for research. We have analyzed our data with raw scores, NCE scores, and *W* scores and presented the latter. The results of the analyses with raw scores and NCE scores, however, differed only in the details from those presented here. They did not provide substantively different findings. For example, the time by condition effect for word attack with intercept at T_1, reported above, gave a t value of 4.18 for *W* scores. The same analysis with NCE scores included all the same factors in the model and gave a t value of 3.77 for the time by condition effect. In the analysis with raw scores, the same effect returned a t value of 4.11. These effects are all very similar; this particular effect translates into effect sizes (Cohen, 1987) of $d = 0.51$, $d = 0.46$, and $d = 0.50$, respectively. Thus, the metric used to measure effects in a randomized control trial is clearly much less important than the size of the effects themselves.

For the oral reading fluency measure, we used grade-level passages at each assessment. The actual passage that students read varied from year to year. This raises questions about comparisons across time and between groups. Many measures of academic skills, performance, or aptitude also change across time. Items are carefully chosen, however, to make valid comparisons from grade to grade. The reading passages for our oral reading fluency measure were chosen with similar care.

For comparisons between groups, the use of different passages could also have created analytical problems, but only

in a quasi-experimental design, such as when comparing only pretest with posttest without a control group or for comparisons with a nonrandomized comparison group. In a randomized trial, such as the present study, measurement error associated with variation in reading passages, at worst, should obscure the effects due to treatment. We have just shown, however, that this variation is minimal. Randomization of children within the same classrooms also maintains the internal validity of comparisons. Finally, alternative methods for collecting oral reading fluency data, such as using the same set of reading passages over time, have similarly challenging problems.

Summary of Results

From the beginning of the intervention, we found improvements in slope due to condition for the three measures collected at T_1 and T_2. Although the effects for letter–word identification were limited to poor readers, the analyses supported intervention effects on slopes for all students with the word attack and reading comprehension measures. For word attack, both Hispanic and non-Hispanic intervention students' slopes improved over controls, and Hispanic students' slopes grew at a significantly greater rate.

At the end of the intervention, at T_3, we found statistically significant difference between conditions on slopes for oral reading fluency and, again only for poor readers, letter–word identification. At T_3, the analyses also provided evidence for mean differences on word attack (d = 0.38), oral reading fluency (d = 0.24), reading comprehension (d = 0.29), and vocabulary (d = 0.28). (See Rosenthal & Rosnow, 1991, p. 302, for a formula to convert a t value to Cohen's d; Cohen, 1988.) Thus, we found evidence of intervention effects on every measure by the end of intervention.

By the final assessment, 2 years after intervention, students differed by condition on letter–word identification (d = 0.25), oral reading fluency (d = 0.29), and reading comprehension (d = 0.29). The effects for vocabulary, however, fell just under the chosen .05 alpha level (d = 0.25). Conditions differed on slopes for oral reading fluency and word attack. For word attack, the slopes had started to converge (see Figure 2), possibly demonstrating the limits of the intervention 2 years after its conclusion. Oral reading fluency scores, however, continued to improve for intervention students.

Discussion

These results support the value of supplemental instruction in decoding skills for improving the reading achievement of K–3 students at risk for reading difficulty. Findings are consistent with other evaluations of supplemental instruction (Foorman et al., 1998; Linan-Thompson & Hickman-Davis, 2002; O'Connor, 2000; Quiroga et al., 2002; Torgesen et al., 1997; Torgesen et al., 1999; Vellutino et al., 1996). It appears that the emphasis on developing word recognition skills, through explicit instruction in phonemic awareness and phonics, accompanied by practice reading decodable text, contributed to improvements in reading ability. Indeed, students in the intervention condition performed significantly better than their controls on measures of entry-level reading skills (i.e., letter word identification and word attack) and on measures of more advanced literacy skills (i.e., oral reading fluency, vocabulary, and comprehension). The benefits of instruction were still clear 2 years after the intervention ended.

Ethnicity

As a subgroup, the Hispanic students had lower baseline scores on the measures of word attack, word identification, and oral reading fluency. This is not surprising, given that these students had varying degrees of familiarity with English and came from homes where 84% of the parents reported speaking only or mostly Spanish. Yet, individually and as a group, the Hispanic students benefited from the supplemental reading instruction in English as much as or more than did the non-Hispanic students. Although their greater gains may be because they began the study with less proficiency in English, it is worth noting that the instruction improved their reading outcomes in comparison with the Hispanic children in the control condition. It is also worth noting that their initial level of English oral language proficiency was not a factor in their ability to benefit from instruction. The present findings are consistent with studies indicating that Spanish-speaking students can benefit from supplemental instruction in reading English (Linan-Thompson & Hickman-Davis, 2002; Quiroga et al., 2002). Results also suggest that rather than delaying such instruction until Spanish-speaking students have developed English oral language skills, schools can help these students succeed in school by teaching them to read English as early as first grade.

Behavior

There were a few differences in performance on the reading measures between those identified as only poor readers and the students who were screened into the study based on aggressive social behavior and who were also below grade level in reading skill on the baseline reading measures. For word identification, the students selected on the behavior criterion had higher initial scores than students selected on the reading criterion. However, by T_5 the intercepts differed between conditions for both groups, as the condition by selection interaction was not significant. For oral reading fluency, there was slope by selection criteria (t = 1.96, p = .0505), where students selected because of reading difficulty grew faster than students selected because of their behavior. Nonetheless, intervention students selected on the behavior criterion made greater gains on the reading measures than their controls, suggesting that the supplemental instruction had an impact on their reading skills.

Grade Level

Students who received supplemental instruction beginning in Grades 2 and 3 benefited to the same extent as students who began instruction in kindergarten and Grade 1. This implies that the tendency of students who are poor readers at the end of Grade 1 to continue to be poor readers (e.g., Juel, 1988) is not an inherent function of their inability to learn to read in later grades. Rather, it seems likely that the lack of subsequent growth or slower growth in reading ability after first grade is due to the absence of continued high-quality instruction in the key skills that some students have not yet acquired, coupled with increasing academic demands and decreasing motivation on the part of the child. Although these factors make it increasingly difficult for older poor readers to catch up as time goes on, our findings suggest that educators can help these students.

In no analyses did we find differences by grade and intervention. That is, the intervention had an impact on the reading skills of older and younger students similarly. We suspect that the tendency of the younger students to catch up or perform better on the WJ-R subtests than the older students was because the older students were more "selected" in the sense that they had already had 1 or 2 years of reading instruction, yet were still performing below grade level. So, among older students, we may have identified those who were more difficult to remediate because of phonological processing or language deficits. However, among the younger students (who had received little or no instruction prior to entry in the study), we may have included students lacking easily taught beginning reading skills who were thus better able to gain from the supplemental instruction.

Fidelity of Implementation

The main premise of the study was that explicit supplemental reading instruction to develop word recognition skills, which was delivered with clear feedback, active engagement, and cumulative review, would be of value for the range of students in elementary classrooms who are at risk for reading difficulty. *Reading Mastery* and *Corrective Reading* were chosen for the intervention because the programs are designed to teach students to decode words and read connected text, and because they give clear guidance to teachers on how to help students master new content and skills. Although the content and instructional design features of the program were essential to the success of the intervention, it was critical that the IAs received training and ongoing coaching to implement the program well. With such support, the IAs were a valuable and cost-effective resource for helping at-risk students learn to read.

Limitations

Although supplemental instruction had clear benefits, examination of the *W* scores on the WJ-R subtests at T_5 indicates that there was still substantial room for improvement on most measures. Intervention students approached the national average on word identification (42nd percentile) and exceeded the national average on word attack (53rd percentile), compared with the control students' average of 30th percentile and 39th percentile, respectively. These results are in keeping with the fact that instruction concentrated on these basic skills. However, averages for vocabulary were 18th percentile for intervention and 12th percentile for control students. For comprehension, the means were the 25th and 18th percentiles, respectively. Thus, with the exception of word attack, even when they received supplemental instruction, students were still performing below the national averages for their grade-level peers. This suggests to us that all students, regardless of grade, probably needed more direct attention to developing their language skills and that the intervention should have included more emphasis on vocabulary development and comprehension strategies (Biemiller, 1999). Although students received supplemental instruction daily, they met for only 30 minutes, instead of the 40-minute sessions recommended by the program authors. So, the shorter duration of each session may have been a factor. It is also possible that the quality of the classroom reading instruction students received after the intervention did not give them the continued instruction they needed to become grade-level readers.

Another limitation of the present study is that the intervention included parenting skills and social skills components. Therefore, we cannot state unequivocally that the improvements in reading skill were due solely to reading instruction. There was evidence that the complete intervention affected parent daily reports of antisocial behavior and parents' use of coercive discipline with boys (see Smolkowski et al., in press, for a complete discussion). Considering the finding of Kellam et al. (1998) that there was no increase in achievement due to improvements in the aggressive behavior in their behavioral intervention, it is possible, though it seems to us unlikely, that these changes contributed to improved reading skill. Moreover, the inference that it was the reading instruction that affected reading skill is bolstered by the specificity of its focus on those skills and on considerable evidence from other studies that such instruction affects reading skill (Linan-Thompson & Hickman-Davis, 2002; O'Connor, 2000; Torgesen, 2000).

It is also possible that intervention students improved simply because of the extra 30 minutes of instruction they received each day and not because of the specific skills they learned. However, given the consistency of our findings with other supplemental interventions focused on explicit instruction in word-level skills, it seems likely that the content of the program, not simply additional time, contributed to reading outcomes.

It remains to be seen whether children who received supplemental instruction will continue to make adequate progress without continued support. Among previous intervention studies, Linan-Thompson and Hickman-Davis (2002) found sig-

nificant effects 4 months postintervention for Hispanic students, and Torgesen (2000) noted that two other studies continue to follow reading development of monolingual children. However, at this time, little is known about the long-term effects of supplemental instruction after intervention. Thus, the results reported here extend the findings of previous studies on the effectiveness of supplemental reading instruction by reporting effects 2 years after the intervention ended.

In conclusion, supplemental reading that used explicit instruction to develop word recognition skills, accompanied by clear feedback, active engagement, and cumulative review, helped students at risk for reading difficulty develop foundational reading skills. Evidence that the benefits of the instruction provided in this study persisted 2 years after the instruction ended attests to the long-term effectiveness of the intervention. In particular, the growth that students made and maintained in decoding skills is encouraging, for skilled reading cannot proceed without fluent word recognition. At the same time, longer term continued instruction that includes more vocabulary development and comprehension strategies would provide even greater benefit in helping children develop the skills they need to be successful readers.

AUTHORS' NOTES

1. Grant DA 09678 from the National Institute on Drug Abuse supported the preparation of this manuscript.
2. The authors thank Christine Cody and Sylvia Gillings for their assistance in the preparation of this manuscript.

REFERENCES

Achenbach, T. M. (1991). *Manual for the teacher's report form and 1991 profile*. Burlington: University of Vermont, Department of Psychiatry.

Adams, G. L., & Engelmann, S. (1996). *Research on direct instruction: 25 years beyond DISTAR*. Seattle, WA: Educational Achievement Systems.

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. G. (1985). *Becoming a nation of readers: The report of the Commission on Reading*. Washington, DC: U. S. Department of Education, National Institute of Education.

August, D., & Hakuta, K. (1997). *Improving schooling for language-minority children: A research agenda*. Washington, DC: National Academy Press.

Biemiller, A. (1999). *Language and reading success*. Cambridge, MA: Brookline.

Cohen, J. (1987). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cunningham, A. E., & Stanovich, K. E. (1998). The impact of print exposure on word recognition. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 235–262). Mahwah, NJ: Erlbaum.

Ehri, L. C. (1998). Grapheme–phoneme knowledge is essential for learning to read words in English. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 3–40). Hillsdale, NJ: Erlbaum.

Engelmann, S., & Bruner, E. C. (1988). *Reading mastery*. Chicago: Science Research.

Engelmann, S., Carnine, L., & Johnson, G. (1988). *Corrective reading: Word-attack basics. Decoding A*. Chicago: Science Research.

Fitzgerald, J. (1995). English-as-a-second language reading instruction in the United States: A research review. *Journal of Reading Behavior, 27*(2), 115–152.

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk students. *Journal of Educational Psychology, 90,* 37–55.

Francis, D. J., Shaywitz, S. E., Steubing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88,* 3–17.

Fuchs, L. S., & Deno, S. L. (1991). Curriculum-based measurement: Current applications and future directions. *Exceptional Students, 57,* 466–501.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5,* 239–256.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9,* 20–28.

Gersten, R., & Geva, E. (2003). Teaching reading to early language learners. *Educational Leadership, 60*(7), 44–49.

Good, R. H., III, Kaminski, R., Laimon, D., & Johnson, D. S. (1992). Advances in assessment for the primary prevention of early academic problems. *The Oregon Conference Monograph, 1992* (pp. 100–109). Eugene: University of Oregon College of Education.

Good, R. H., III, Kaminski, R. A., Simmons, D., & Kame'enui, E. J. (2001). *Using dynamic indicators of basic early literacy skills (DIBELS) in an outcomes-driven model: Steps to reading outcomes.* Unpublished manuscript.

Gunn, B., Biglan, A., Smolkowski, K., & Ary, D. (2000). The efficacy of supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. *The Journal of Special Education, 34*(2), 90–103.

Gunn, B., Smolkowski, K., Biglan, A., & Black, C. (2002). Supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school: A follow-up. *The Journal of Special Education, 36*(2), 69–79.

Harrell, F. E. (2001). *Regression modeling strategies with applications to linear models, logistic regression, and survival analysis*. New York: Springer-Verlag.

Hinshaw, S. P. (1992). Academic underachievement, attention deficits, and aggression: Comorbidity and implications for intervention. *Journal of Consulting and Clinical Psychology, 60*(6), 893–903.

Hops, H., & Walker, H. M. (1988). *CLASS: Contingencies for Learning Academic and Social Skills: A classroom behavior management program for children with acting-out behaviors*. Delray Beach, FL: Educational Achievement Systems.

Jenkins, J. R., Fuchs, L. S., Espin, C., van den Broek, P., & Deno, S. L. (February, 2000). *Effects of task format and performance dimension on word reading measures: Criterion validity, sensitivity to impairment, and context facilitation*. Paper presented at Pacific Coast Research Conference, San Diego, CA.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 students from first through fourth grades. *Journal of Educational Psychology, 80*(4), 437–447.

Kellam, S. G., Mayer, L. S., Rebok, G. W., & Hawkins, W. E. (1998). Effects of improving achievement on aggressive behavior and of improving aggressive behavior on achievement through two preventive interventions: An investigation of causal paths. In B. P. Dohrenwend (Ed.), *Adversity, stress, and psychopathology* (pp. 486–505). London: Oxford University Press.

Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling*. London: Sage.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293–323.

Laird, N. M. (1988). Missing data in longitudinal studies. *Statistics in Medicine, 7,* 305–315.

Linan-Thompson, S., & Hickman-Davis, P. (2002). Supplemental reading instruction for students at risk for reading disabilities: Improve reading 30 minutes at a time. *Learning Disabilities Research & Practice, 17*(4), 242–251.

Lindsey, K. A., Manis, F. R., & Bailey, V. E. (2003). Prediction of first grade reading in Spanish-speaking English-language learners. *Journal of Educational Psychology, 95*(3), 482–494.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS system for mixed models*. Cary, NC: SAS Institute.

Little, R. J. A. (1995). Modeling the dropout mechanism in repeated-measures studies. *Journal of the American Statistical Association, 90,* 1112– 1121.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Markell, M. A., & Deno, S. L. (1997). Effects of increasing oral reading: Generalization across reading tasks. *The Journal of Special Education, 31,* 233–250.

McGee, R., Prior, M., Williams, S., Smart, D., & Sanson, A. (2002). The long-term significance of teacher-rated hyperactivity and reading ability in childhood: Findings from two longitudinal studies. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 43*(8), 1004–1016.

Murray, D. M. (1998). *Design and analysis of group-randomized trials*. New York: Oxford University Press.

Nich, C., & Carroll, K. (1997). Now you see it, now you don't: A comparison of traditional versus random-effects regression models in the analysis of longitudinal follow-up data from a clinical trial. *Journal of Consulting and Clinical Psychology, 65,* 252–261.

O'Connor, R. (2000). Increasing the intensity of intervention in kindergarten and first grade. *Learning Disabilities Research & Practice, 15*(1), 43–54.

Ortiz, A., & Graves. A. W. (2001). English language learners with literacy-related learning disabilities. *The International Dyslexia Association Commemorative Booklet, 52nd Annual Conference* (pp. 31–35). Albuquerque, NM.

Patterson, G. R., DeBaryshe, B. D., & Ramsey, E. (1989). A developmental perspective on antisocial behavior. Special Issue: Students and their development—knowledge base, research agenda, and social policy application. *The American Psychologist, 44,* 329–335.

Quiroga, T., Lemos-Britton, Z., Mostafapour, E., Abbott, R. D., & Berninger, V. W. (2002). Phonological awareness and beginning reading in Spanish-speaking ESL first graders: Research into practice. *Journal of School Psychology, 40*(1), 85–111.

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,* 31–74.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research, methods, and data analysis*. San Francisco: McGraw-Hill.

SAS Institute. (1999). *SAS/STAT® user's guide, version 8*. Cary, NC: SAS.

Share, D. L., & Stanovich, K. E. (1995). Cognitive processes in early reading development: Accommodating individual differences into a model of acquisition. *Issues in Education: Contributions From Educational Psychology, 1*, 1–57.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 24*(4), 323–355.

Singer, J. D. (2002). Fitting individual growth models using SAS PROC MIXED. In D. S. Moskowitz & S. L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: Methods and applications*. Mahwah, NJ: Erlbaum.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Slavin, R. E., Madden, N. A., Dolan, L. J., Wasik, B. A., Ross, S., Smith, L., et al. (1996). Success for all: A summary of research. *Journal of Education for Students Placed at Risk, 1*(1), 41–76.

Smolkowski, K., Biglan, A., Barrera, M., Taylor, T. K., Black, C., & Blair, J. (2005). Schools and Homes in Partnership (SHIP): Long-term effects of a preventive intervention focused on social behavior and reading skill early elementary school. *Prevention Science*.

Snow, C. E., Burns, S. M., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy of Sciences, National Research Council, Commission on Behavioral and Social Sciences and Education.

Stahl, S., & Miller, P. (1989). Whole language and language experience approaches for beginning reading: A quantitative research synthesis. *Review of Educational Research, 59*(87), 116.

Stanford Achievement Test, Ninth Edition. (1996). San Antonio, TX: Harcourt Brace.

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16,* 32–71.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360–406.

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice, 15*(1), 55–64.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1997). Approaches to the prevention and remediation of phonologically based reading disabilities. In B. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (pp. 287–304). Mahwah, NJ: Erlbaum.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., & Herron, J. (1999). *A comparison of two computer assisted approaches to the prevention of reading disabilities in young children*. Unpublished manuscript.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A. C., Chen, R., et al. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88,* 601–638.

Walker, H., & McConnell, S. (1988). *The Walker-McConnell scale of social competence and school adjustments*. Austin, TX: PRO-ED, Inc.

Walker, H. M., Ramsey, E., & Gresham, F. M. (Eds.). (2004). *Antisocial behavior in school: Evidence-based practices* (2nd ed.). Belmont, CA: Wadsworth/Thomson Learning.

Wallace, D., & Green, S. B. (2002). Analysis of repeated measures designs with linear mixed models. In D. S. Moskowitz & S. L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: Methods and applications*. Mahwah, NJ: Erlbaum.

Webster-Stratton, C. (1992a). *The incredible years: A trouble-shooting guide for parents of children aged 3–8*. Toronto, Canada: Umbrella Press.

Webster-Stratton, C. (1992b). *The Teachers and Children's Videotape Series: Dina Dinosaur's Social Skills and Problem-Solving Curriculum*. Seattle, WA.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development, 69,* 848–872.

Woodcock, R. W., & Mather, N. (1989). WJ-R tests of achievement: Examiner's manual. In R. W. Woodcock & M. B. Johnson (Eds.), *Woodcock-Johnson psycho-educational battery–Revised*. Allen, TX: DLM Teaching Resources.

Woodcock, R. W., & Mather, N. (1990). WJ-R tests of achievement: Examiner's manual. In R. W. Woodcock & M. B. Johnson (Eds.), *Woodcock-Johnson psycho-educational battery–Revised*. Allen, TX: DLM Teaching Resources.