Research Article

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
24(8) 1408–1419

DOI: 10.1177/0956797612472204
pss.sagepub.com
SAGE

A Longitudinal Cluster-Randomized Controlled Study on the Accumulating Effects of Individualized Literacy Instruction on Students' Reading From First Through Third Grade

Carol McDonald Connor[1,2], Frederick J. Morrison[3], Barry Fishman[4], Elizabeth C. Crowe[2], Stephanie Al Otaiba[5], and Christopher Schatschneider[2,6]

[1]Learning Sciences Institute, Arizona State University; [2]Florida Center for Reading Research, Florida State University; [3]Department of Psychology, University of Michigan; [4]School of Education, University of Michigan; [5]School of Education and Human Development, Southern Methodist University; and [6]Department of Psychology, Florida State University

Abstract

Using a longitudinal cluster-randomized controlled design, we examined whether students' reading outcomes differed when they received 1, 2, or 3 years of individualized reading instruction from first through third grade, compared with a treated control group. More than 45% of students came from families living in poverty. Following students, we randomly assigned their teachers each year to deliver individualized reading instruction or a treated control condition intervention focused on mathematics. Students who received individualized reading instruction in all three grades showed the strongest reading skills by the end of third grade compared with those who received fewer years of such instruction. There was inconsistent evidence supporting a sustained first-grade treatment effect: Individualized instruction in first grade was necessary but not sufficient for stronger third-grade reading outcomes. These effects were achieved by regular classroom teachers who received professional development, which indicates that policies that support the use of evidence-based reading instruction and teacher training can yield increased student achievement.

Keywords

childhood development, literacy, academic achievement, dyslexia, reading

Received 4/16/12; Revision accepted 11/27/12

Children's reading skills are closely tied to their ongoing academic and life success (National Institute of Child Health and Human Development, NICHD, 2000). Yet only about 65% of fourth graders perform above basic levels on the National Assessment of Educational Progress (NAEP, 2011), with only about 30% achieving the proficient levels needed to fully participate in our 21st century information-based global economy. This percentage is much lower for children from lower-income families than for children from higher-income families. Understanding what effective teachers are actually doing in the classroom and how to support sustained effective instruction for all students as they progress through school is a national challenge. There have been two schools of thought regarding the importance of early elementary instruction (Coyne, Kame'enui, Simmons, & Harn, 2004)—that early intervention can preclude, or inoculate against, later reading difficulties (Vellutino et al., 1996) or that

Corresponding Author:
Carol McDonald Connor, Arizona State University, Department of Psychology, Learning Sciences Institute, P. O. Box 872111, Tempe, AZ 85287
E-mail: carol.connor@asu.edu

ongoing and sustained high-quality instruction is required (Snow, Burns, & Griffin, 1998). Accumulating research suggests that the issue is complex and that students with different cognitive profiles respond differently to interventions (Connor, Morrison, Fishman, Schatschneider, & Underwood, 2007; Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998). These studies suggest that for some children, a dose of intensive intervention is enough to prevent additional problems, whereas for others it is not (Coyne et al., 2004). However, many of these studies focused specifically on children with or at risk for learning disabilities, with interventions provided by highly trained teachers.

What about average students attending higher poverty schools, many of whom are at serious risk for reading failure? Head Start preschool studies suggest that initial gains in reading skills diminish over time as children experience inadequate instruction (Lee & Loeb, 1995), whereas other preschool programs show sustained benefits over time (Reynolds, Temple, Robertson, & Mann, 2002). The national Reading First initiative, which focused on students in the first through third grade, showed positive effects for word reading but not for comprehension (Gamse, Jacob, Horst, Boulay, & Unlu, 2008). Most recently, large data sets and teacher value-added scores[1] have been used in correlational studies to show that when students have more effective teachers, they show greater achievement and that effects may accumulate, with the impact of effective teaching in early grades fading through elementary school (Chetty, Friedman, & Rockoff, 2012; Konstantopoulos & Chung, 2011). Value-added studies have informed policy and, as a result of federal initiatives such as Race to the Top, many states are beginning to hold individual teachers accountable for their students' academic progress using value-added scores (U.S. Department of Education, 2012).

There are, however, concerns with value-added scores, which may be influenced by school and peer effects rather than teachers' instruction (Justice, Petscher, Schatschneider, & Mashburn, 2011; Raudenbush, 2004; Rothstein, 2008) and leave important questions unanswered. It is still not clear, for example, whether the timing of effective instruction is important and whether and how the impact of effective teaching accumulates as students attending higher poverty schools progress through the elementary years. This concern has implications for teacher deployment and training both theoretically (e.g., testing the self-teaching hypothesis; Share, 1995; Stanovich, 1986) as well as practically. Finally, value-added scores say nothing about what teachers and students are actually doing in the classroom and what can be done to improve teacher practice and, hence, student outcomes. Intervention studies that use randomized controlled designs can elucidate effective instructional strategies. However, because of logistics and resources, they are typically single-grade studies (Elleman, Lindo, Morphy, & Compton, 2009; NICHD, 2000) and so cannot test the timing or accumulation of effects.[2] Additionally, many of these produce relatively small or no effects on standardized reading measures (e.g., James-Burdumy et al., 2010).

In the present investigation, we extended and addressed gaps in the current research in three ways. First, we used a cluster-randomized treated-control design (Raudenbush & Bryk, 2002; Shadish, Cook, & Campbell, 2002), which permits stronger causal claims than archival, correlational, or quasiexperimental study designs do (Shavelson & Towne, 2002). Previous correlational and quasiexperimental findings could be related to selection bias; for example, students who received more effective instruction might have been subsequently assigned to classrooms with higher achieving peers (i.e., tracked), which would explain accumulation and inoculation effects. Second, longitudinal data were collected. By following students longitudinally and, each year, randomly assigning their teachers to either a treatment or control condition, we were able to examine whether effects of individualized instruction, relative to the control group, were cumulative and whether effective first-grade individualized instruction protected students attending higher poverty schools against later less effective instruction (i.e., provided inoculation). Finally, using an intervention study, we explicated the elements of more effective literacy instruction and whether these effects accumulate over time such that relatively small effects add up to large effects by the end of third grade.

The classroom instructional reading intervention used in this study, Individualizing Student Instruction (ISI), was designed to investigate whether Child Characteristic × Treatment interactions (also called Aptitude × Treatment interactions; Connor et al., 2007; Cronbach & Snow, 1977) were causally related to variability observed in students' reading achievement, even when they shared the same classroom. In our past research, single-grade efficacy trials in kindergarten (Al Otaiba et al., 2011), first grade (Connor et al., 2007; Connor, Morrison, Schatschneider, et al., 2011; Connor, Piasta, et al., 2009), and third grade (Connor, Morrison, Fishman, et al., 2011) produced three separate results. First, optimal amounts and types of reading instruction to achieve proficient skill differed for students and could be computed fairly precisely using algorithms that took into account students' current reading levels, end-of-school-year target, and vocabulary skills. Second, students who received ISI in reading achieved stronger reading outcomes compared with the control group, and, third, the more closely the amount of reading instruction they received matched ISI-recommended amounts, the stronger their reading outcomes were. Using student-outcome and classroom-observation data from these studies, we refined operating algorithms and added new algorithms (see Part A in the Supplemental Material available online), which were used in the present study.

This study had three principal aims: (a) to replicate our findings from the single-grade studies, (b) to investigate whether ISI-reading effects accumulate from first through third grade, and (c) to examine whether receiving ISI in reading in first grade, rather than in later grades, resulted in stronger end-of-third-grade reading ability (i.e., produced an inoculation effect).

Method

Study design and recruitment

Figure 1 provides an overview of the cluster-randomized treated-control design of our study. The advantage of a treated-control design was that all of the teachers in the study received training in an important area of instruction, Hawthorne effects were mitigated, and the motivation of the teachers was more likely to be similar across groups (Shadish et al., 2002). An important objective in this study was to control the amount of professional development provided so that it was the same for both groups, which was not the case in previous studies. In this design, the business-as-usual control condition for the ISI-reading group was typical reading instruction provided to students in the math-intervention condition. Similarly, the business-as-usual control group for the mathematics condition was typical mathematics instruction provided to students in the ISI-reading condition.

Schools, first-grade teachers, and their students (N = 468) were recruited in the 2008 to 2009 school year, and each teacher (N = 28) within schools was randomly assigned to either the ISI-reading or math-intervention condition. We then followed these students into second grade and recruited their teachers (N = 49) and classmates (N = 568). Each classroom within schools was again randomly assigned to condition (see Fig. 1). Of the second-grade students, 398 were also in the first-grade sample (15% attrition). We then followed these students into third grade and recruited their teachers and classmates (N = 40 teachers and 541 students) and again randomly assigned each classroom to condition. Of these students, 523 also participated in the second-grade study (3% attrition), and 357 participated in the first-grade study (24% attrition over all 3 years of the study). A few teachers declined to participate, and in other cases, students moved to new schools, and we lacked the resources to include their schools and teachers in the interventions. However, each of these students (n = 68; 88 data points) was still assessed and assigned to the control group for the year that their teacher did not participate. Their scores were included in the longitudinal part of this study but not the within-grade part of the study. Analyses excluding these children yielded similar results.

Schools were located in north Florida in a district that had not participated in previous ISI studies and served economically and ethnically diverse student bodies in urban, suburban, and rural communities. The percentage of students at each school receiving free or reduced-price lunch, a commonly used indicator of poverty, ranged from 39% to 59%, with an overall mean of 47%. Six schools were recruited in the 2008 to 2009 school year. One school was closed the following year, so five schools participated in the second-grade study. The district opened a new school in 2010 to 2011 serving approximately 90 of the students in the study, which we recruited. Eighty-one percent of participating students were White, 6% were Black, and 13% belonged to other ethnic/racial groups. Classroom teachers were all certified following Florida requirements, including holding a bachelor's degree; 23% held at least a master's degree.

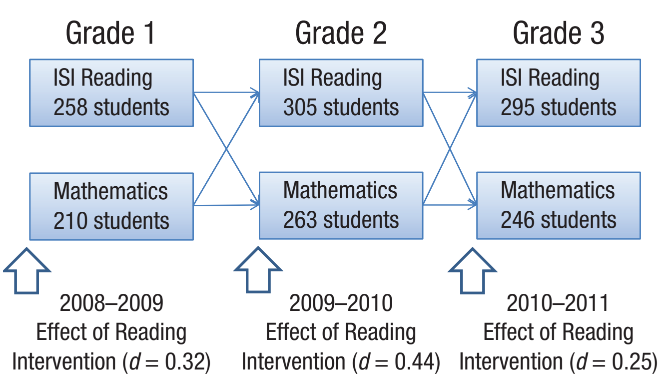

Fig. 1. Longitudinal research design of the present study. Children were followed from first through third grade, and their teachers were randomly assigned to condition—Individualizing Student Instruction (ISI) reading intervention or mathematics control instruction—at each grade. Hence, students could receive 1, 2, or 3 years of the ISI reading intervention in various grade configurations.

Student assessments

Students' literacy and mathematics skills were assessed. For this reading study, we used the Letter-Word Identification and Passage Comprehension tests from the Woodcock-Johnson III Tests of Achievement (Woodcock, McGrew, & Mather, 2001), which were individually administered in the fall, winter, and spring each school year (see Table 1). Raw scores were converted to W scores, which are a variation of Rasch scores and provide equal intervals. A W score of 500 (SD = 15) is the expected score for a 10-year-old child. Students' Letter-Word Identification scores were highly correlated with their Passage Comprehension scores (r = .869, p < .001) and with the scores on the Gates-MacGinitie Reading Test (GMRT;

Table 1. Mean *W* Scores for Word Reading and Passage Comprehension for Students in the First Through Third Grades

Measurement occasion and years in ISI reading intervention	*n*	Word Reading score		Passage Comprehension score	
		M	*SD*	*M*	*SD*
First grade: fall					
0	100	421.71	31.413	443.86	24.852
1	201	420.93	27.444	444.16	22.951
2	111	419.86	24.903	442.60	21.173
3	37	415.73	18.458	439.05	19.452
Total	449	420.41	27.135	443.28	22.652
First grade: winter					
0	98	448.78	27.796	465.91	16.172
1	208	450.12	24.144	467.62	15.363
2	113	453.17	22.471	468.81	12.848
3	38	446.97	18.877	463.89	11.714
Total	457	450.32	24.194	467.23	14.704
First grade: spring					
0	105	459.77	25.974	472.55	15.606
1	211	463.35	21.966	475.26	13.529
2	114	465.21	18.325	476.95	11.171
3	38	462.00	16.288	477.43	9.547
Total	468	462.89	21.752	475.28	13.248
Second grade: fall					
0	171	462.63	26.713	443.86	24.852
1	280	464.27	25.420	444.16	22.951
2	147	470.73	20.343	442.60	21.173
3	38	467.50	16.825	439.05	19.452
Total	636	465.51	24.414	443.28	22.652
Second grade: winter					
0	135	477.10	22.788	465.91	16.172
1	256	476.19	22.698	467.62	15.363
2	142	482.78	17.829	468.81	12.848
3	38	479.68	17.037	463.89	11.714
Total	571	478.28	21.390	467.23	14.704
Second grade: spring					
0	140	483.80	21.562	472.55	15.606
1	251	486.40	19.975	475.26	13.529
2	139	490.46	16.663	476.95	11.171
3	38	487.32	14.226	477.43	9.547
Total	568	486.81	19.398	475.28	13.248
Third grade: fall					
0	163	488.36	17.143	485.34	11.312
1	282	490.08	19.817	486.82	11.580
2	134	492.93	17.730	489.31	9.680
3	37	490.89	16.389	488.24	10.056
Total	616	490.29	18.523	487.12	11.058
Third grade: winter					
0	136	494.90	17.777	490.65	9.566
1	228	496.84	19.105	490.83	10.375
2	114	500.87	15.716	493.54	9.655
3	30	499.90	17.050	493.00	8.980
Total	508	497.41	18.004	491.56	9.969
Third grade: spring					
0	136	500.17	17.967	493.42	10.848
1	257	502.82	19.387	494.22	10.284
2	119	507.39	16.012	496.05	9.105
3	29	505.21	17.230	493.97	8.529
Total	541	503.29	18.351	494.43	10.082

Note: ISI = Individualizing Student Instruction.

MacGinitie & MacGinitie, 2006; $r = .746$, $p < .001$), which was administered in the fall and spring. We created a latent variable (reading) using the Passage Comprehension and Letter-Word Identification W scores (see Part B in the Supplemental Material available online).

The ISI reading intervention

Components. The ISI reading classroom intervention has three components (see Part A in the Supplemental Material): teacher professional development, Assessment to Instruction (A2i) software (Connor et al., 2013), and implementation in the classroom. Teacher training for both reading and mathematics interventions was conducted using a coaching model and included half-day workshops at the beginning of the school year; monthly meetings with other teachers, commonly referred to as "communities of practice" (Bos, Mather, Narr, & Babur, 1999); individual meetings as needed; and classroom-based support provided every 2 weeks by research assistants who were also certified teachers. The amount and format of training for both interventions was the same; only the content differed. Using classroom observations (Connor, Morrison, et al., 2009; Pianta, Belsky, Houts, & Morrison, 2007), teachers successfully implemented the instructional intervention to which they were assigned, and there was no appreciable drift of the mathematics intervention to the ISI-reading classrooms and vice versa.

The A2i software uses three test scores for each student—word reading, comprehension, and vocabulary—in dynamic forecasting intervention models (i.e., algorithms) to compute recommended amounts for each of four types of literacy instruction (see Table 2). Using A2i, teachers access the recommended amounts of instruction, students' test scores, and recommended student groupings. There are also planning Web pages and access to training materials and discussion boards. Teachers could view their students' scores and track progress over time throughout the school year using A2i. Paper reports of the scores were provided to teachers in the mathematics control group in November, February, and May. Assessments were administered in a room or quiet hallway close to the students' classrooms by trained research assistants.

The ISI reading intervention characterizes reading instruction across two dimensions: content and management (see Part A in the Supplemental Material and Table 2). The content of reading instruction may be code focused—activities designed to teach the alphabetic principle—or meaning focused—activities that support the extraction and construction of meaning from text. The management dimension identifies who is focusing the students' attention on the learning activity: the teacher and student together—teacher-child managed (TCM)—or the student alone or with peers—child managed (CM). These dimensions operate simultaneously to describe four types of literacy instruction: TCM-code-focused, TCM-meaning-focused, CM-code-focused, and CM-meaning-focused instruction. The recommended amounts of instruction were provided to students with similar reading skills in small groups during the dedicated block of time devoted to literacy instruction. Using their core curriculum and other resources (e.g., Florida Center for Reading Research activities; www.fcrr.org), most teachers used a center or station approach in which the teacher met with small groups of students while other students worked independently or in small peer groups. Schools provided one of several core reading curriculums, including Houghton Mifflin, that were indexed to the four types of instruction. Thus, teachers used instructional materials with which they were familiar and learned how to individualize reading instruction for students when using these materials.

A2i algorithms—dynamic forecasting intervention models. A2i dynamic forecasting intervention models compute recommended daily minutes of small-group

Table 2. Examples of the Four Dimensions of Literacy Instruction in the Individualizing Student Instruction (ISI) Reading Intervention

	Management dimension	
Content dimension	Teacher-child managed	Child managed
Code focused	The teacher is working with a small group of students at the teacher table. The children are learning how to use prefixes and suffixes to decode multisyllabic words.	Working individually and in pairs, students are completing spelling activities from their core reading curriculum, including phonics rules.
Meaning focused	The teacher is discussing a book with a small group of students. Each student takes a turn discussing the main ideas and why characters acted the way they did.	Students are reading books in the library corner.

Note: Examples are from the second-grade intervention.

code- and meaning-focused activities for both the TCM and CM groups, with substantially different algorithms for each grade. These models were developed using correlational research findings that the effect of specific instructional types on children's reading-skill growth depended on students' language and literacy skills. Child Characteristic × Instruction effects have been observed from kindergarten through third grade for different samples (Al Otaiba et al., 2011; Foorman et al., 2006).

The algorithms are analogous to those used by meteorologists to predict, for example, the paths of hurricanes (DeMaria, Rhome, Pasch, & Clark, 2009). For each student, the algorithms use a target outcome (end-of-grade-level equivalent or a school-year's gains, whichever is greater), month of the school year, and students' Letter-Word Identification, Vocabulary, and Passage Comprehension grade-equivalent scores to compute recommended amounts of each type of instruction. These recommended minutes per day are displayed in the A2i classroom view (see Figs. S1 and S2 in the Supplemental Material). Research shows that the more precisely children received the recommended instruction, the greater were their reading gains, with large effects (Connor, Morrison, Schatschneider, et al., 2011; Connor, Piasta, et al., 2009). We refined dynamic forecasting intervention algorithms for first and second grade for this study (see Figs. 2 and 3 and Part A in the Supplemental Material). For third grade, we used the already developed algorithms (Connor, Morrison, Fishman, et al., 2011).

The treated control interventions

In first grade, teachers in the control condition learned to use Math Pals (Fuchs et al., 1997) as a supplement to the

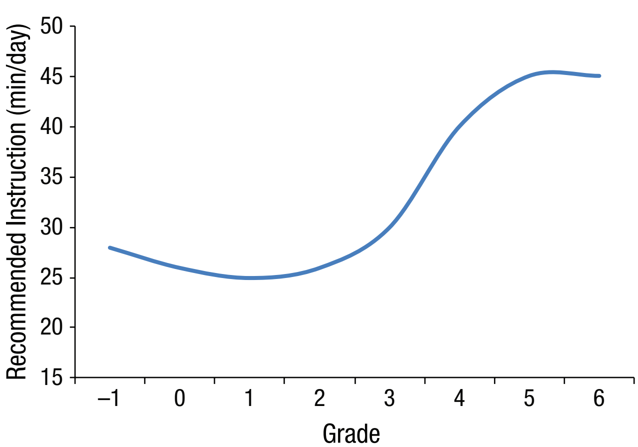

Fig. 2. Recommended amount of teacher-child-managed, meaning-focused instruction in the first grade as a function of students' grade-equivalent reading level (–1 = preschool, 0 = kindergarten). Recommended amounts were computed using Assessment to Instruction software (Connor et al., 2013).

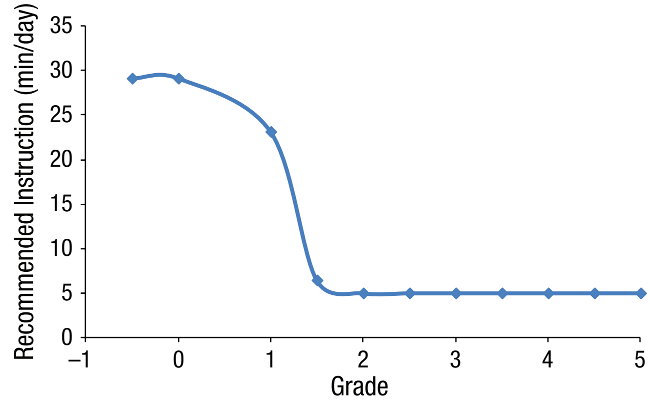

Fig. 3. Recommended amount of child-managed, code-focused instruction in the first grade as a function of students' grade-equivalent reading level (–1 = preschool, 0 = kindergarten). Recommended amounts were computed using Assessment to Instruction software (Connor et al., 2013).

core mathematics curriculum. In second and third grade, teachers in the control condition learned to provide specific types of researcher-developed mathematics activities for each student based on each students' mathematics skills (see Part A in the Supplemental Material). Again, the professional development regimen was the same for both groups.

Results

Research Aim 1: Replicating single-grade studies and testing algorithms

Hierarchical linear models (Raudenbush & Bryk, 2002), with students nested within classrooms (see Part B in the Supplemental Material), revealed that students in the ISI-reading condition achieved significantly stronger Letter-Word Identification reading outcomes than the treated control condition did in first grade (Cohen's $d = 0.32$), in second grade (Cohen's $d = 0.44$), and in third grade (Cohen's $d = 0.25$; see Tables 3 and 4). These are small to moderate effects (Hill, Bloome, Black, & Lipsey, 2008) and represent about a 2-month advantage in skills for each grade by the end of the school year; findings for Passage Comprehension were similar (Cohen's $d = 0.36$, 0.43, and 0.06 for first, second, and third grade,[3] respectively). These findings replicated previous single-year results obtained in a different school district (Connor, Morrison, Fishman, et al., 2011; Connor et al., 2007) and suggest that the algorithms worked as intended.

Research Aim 2

We next tested whether the effect of the ISI reading intervention, compared with the effect of the treated control

Table 3. Fixed Effects From the Within-Grade Hierarchical Linear Model for the Spring Letter-Word Identification Assessment (Reading Score)

Predictor	*b*	*b SE*	*t*	*p*
First grade				
Fitted mean reading, γ_{00}	460.025	1.220	$t(26) = 376.976$	< .001
Effect of ISI in Grade 1, γ_{01}	4.344	1.689	$t(26) = 2.571$	.016
Effect of fall reading, γ_{10}	0.601	0.023	$t(444) = 25.192$	< .001
Second grade				
Fitted mean reading, γ_{00}	479.401	2.441	$t(46) = 196.390$	< .001
Effect of ISI in Grade 2, γ_{01}	6.722	2.83	$t(46) = 2.375$	.022
Effect of fall reading, γ_{10}	0.671	0.022	$t(885) = 29.873$	< .001
Third grade				
Fitted mean reading, γ_{00}	502.134	0.711	$t(37) = 706.189$	< .001
Effect of ISI in Grade 3, γ_{01}	2.161	0.967	$t(37) = 2.233$	.032
Effect of fall reading, γ_{10}	0.826	0.026	$t(38) = 31.769$	< .001

Note: First-grade deviance = 3,690.28; second-grade deviance = 5,978.767976; third-grade deviance = 3,689.469255. The fitted mean spring reading score for the control group is represented by γ_{00}; the effect of the Individualizing Student Instruction (ISI) reading intervention in Grade 1 is represented by γ_{01} (total fitted score = $\gamma_{00} + \gamma_{01}$); and the effect of the fall reading *W* score is represented by γ_{10}.

condition, might accumulate as students experience more years of the intervention. To do this, we used crossclassified random effects growth-curve modeling (Raudenbush & Bryk, 2002) because students share a classroom and teacher in first grade but have different teachers and classmates in second and third grade (see Part B in the Supplemental Material). Additionally, to utilize the two reading measures administered, we created factor scores using Letter-Word Identification and Passage Comprehension *W* scores (Part B in the Supplemental Material). Students' fall, winter, and spring reading scores in first, second, and third grade ($N = 882$ students and 95 teachers) were used to model the fitted growth curve, centered in the spring of second grade. The end of second grade was selected because it provides a meaningful intercept while remaining close to the midpoint of the data. The fitted growth curve from first through third grade was nonlinear (see Fig. 4), including both quadratic and cubic trends with most rapid growth during first grade, slowing growth during second grade, and more linear growth during third grade. We then tested the effect of having 1, 2, or 3 years of ISI reading on student outcomes as compared with the control condition. Results revealed that students who spent more years in ISI reading classrooms than in mathematics intervention classrooms made significantly greater gains in reading (Cohen's $d = 0.20$ per year or 0.60 for 3 years compared with no years, which is a large effect; see Tables 5 and 6). Similar results for the

Table 4. Random Effects From the Within-Grade Hierarchical Linear Model for the Spring Letter-Word Identification Assessment (Reading Score)

Predictor	*SD*	Variance component	χ^2	*p*
First grade				
Classroom Level 2, u_0	3.116	9.713	$\chi^2(26, N = 28) = 47.64582$	.006
Student Level 1, *r*	13.636	185.962	—	—
Second grade				
Classroom Level 2, u_0	12.844	164.971	$\chi^2(46, N = 49) = 491.57049$	< .001
Student Level 1, *r*	15.172	230.190	—	—
Third grade				
Classroom Level 2, u_0	2.057	4.231	$\chi^2(36, N = 40) = 62.10752$	.005
Fall reading score, u_1	0.086	0.007	$\chi^2(37, N = 40) = 46.59341$	.134
Student Level 1, *r*	8.614	74.203	—	—

Note: Deviance = 37,511.511728. Time is in months and is centered at the end of second grade.

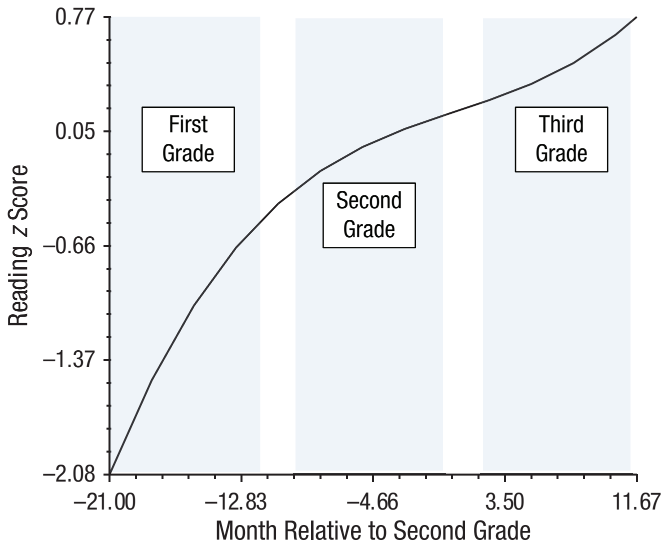

Fig. 4. Fitted mean growth curve for students' reading scores from the fall of first grade through the spring of third grade for all students (end of second grade = 0 months).

GMRT are provided in Table S1 in the Supplemental Material.

Research Aim 3

We then examined whether individualized reading instruction in first grade might have a larger effect on third-grade outcomes than in second and/or third grade and protect students from later less effective literacy instruction. Results are shown in Fig. 5 (see also Tables 7 and 8) and reveal inconsistent evidence for a first-grade "inoculation effect." Students who were in ISI-reading classrooms in first grade and were in control classrooms in second and third grade (TCC) achieved significantly higher fitted scores than did students who received 1 year of ISI reading instruction either in second or in third grade (CTC, CCT, respectively). However, when we considered students who received 2 years of ISI, the results were less consistent. For those students who were in ISI-reading classrooms for 2 years and control classrooms for 1 year, there was an advantage in receiving individualized reading instruction in first and third grade (TCT) compared with receiving effective reading instruction in first and second (TTC) or second and third (CTT) grade. At the same time, students who were in ISI reading classrooms for all three grades (TTT) achieved significantly higher fitted reading scores by the end of third grade compared with students who were in control classrooms all 3 years (CCC; Cohen's d = 0.90), 2 years, or 1 year. Students in the TTT groups achieved fitted scores that were, on average, well above grade-level expectations (grade equivalent = 5.0).

Discussion

The results of this 3-year, cluster-randomized controlled, longitudinal efficacy study have implications for practice, policy, and research. First, they provide evidence that individualized reading instruction is more effective in improving students' reading skills than instruction of similar quality that is not individualized; this was true at a single grade, and the findings were more consistent for first and second grade. This finding replicates other single-grade studies but with a different population of

Table 5. Fixed Effects From the Cross-Classified, Growth-Curve Model for Total Number of Years Students Were in ISI Intervention Classrooms for Reading Score (z score)

Effect	b	b SE	t	p
Fitted mean, θ_0	0.204	0.033667	t(2911) = 6.055	< .001
Total ISI, γ_{01}	0.060	0.026434	t(880) = 2.279	.023
Time, θ_1	0.037	0.002315	t(2911) = 16.159	< .001
Total ISI, γ_{11}	−0.004	0.001568	t(2911) = −2.656	.008
Time^2, θ_2	−0.0005	0.000156	t(2911) = −3.178	.001
Total ISI, γ_{21}	0.0001	0.000106	t(2911) = 1.092	.275
Time^3, θ_3	0.0001	0.000010	t(2911) = 13.368	< .001
Total ISI, γ_{31}	0.00002	0.000007	t(2911) = 3.429	< .001

Note: The fitted mean, θ_0, represents the fitted mean end-of-second-grade score for students who were in control classrooms all 3 years. Time is in months and is centered at the end of second grade. Time^2 represents the quadratic trend, and Time^3 represents the cubic trend. Total ISI, γ_{01}, represents the number of years (1, 2, or 3) that students were in Individualizing Student Instruction (ISI) classrooms. Total ISI, γ_{11}, is the effect of ISI on the slope; Total ISI, γ_{21}, is the effect of ISI on the quadratic trend; and Total ISI, γ_{31}, is the effect of ISI on the cubic trend.

Table 6. Random Effects From the Cross-Classified, Growth-Curve Model for Total Number of Years Students Were in ISI Intervention Classrooms for Reading Score (*z* score)

Effect	*SD*	Variance component	χ^2	*p*
Intercept 1/row, b_{00j}	0.547	0.299	$\chi^2(770) = 9737.369$	< .001
Time/row, b_{10j}	0.019	0.0004	$\chi^2(770) = 2429.529$	< .001
Level 1, *e*	0.294	0.086	—	—

Note: *N* = 882 students and 95 teachers. ISI = Individualizing Student Instruction.

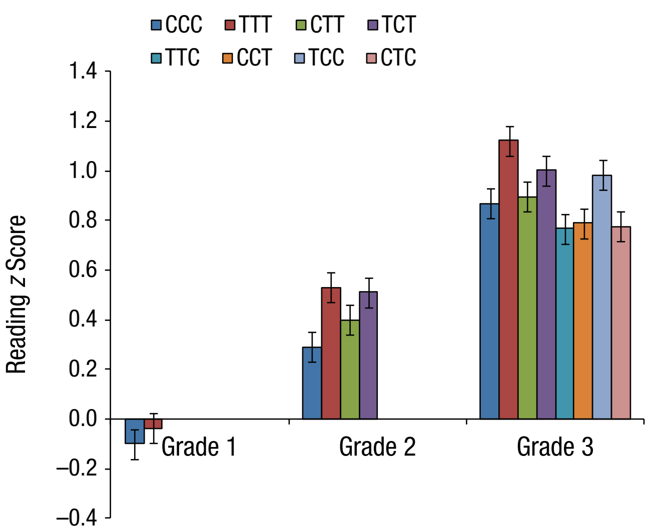

Fig. 5. Fitted end of first-, second-, and third-grade reading outcomes as a function of whether students (*N* = 882) participated in Individualizing Student Instruction (ISI)-reading instruction or the mathematics intervention. In the key, "C" indicates the mathematics control group and "T" indicates the ISI-reading treatment condition, with the order of letters signaling which type of instruction was received in Grades 1, 2, and 3 (e.g., TTC students received ISI reading in first and second grade and were in the control group in third grade). Error bars show standard errors. Values were calculated using data shown in Table 7.

students and teachers and with a control group that received the same amount of professional development, albeit in mathematics.

This study also revealed that once is not enough, even in first grade. The support for a first-grade inoculation effect was inconsistent and inconclusive. Rather, students who received the more efficacious ISI instruction in first, second, and third grade had stronger reading skills at the end of third grade than did children who received fewer years of ISI instruction, regardless of when the ISI instruction was received, and the accumulated impact was large by any standard. Students who attended ISI classrooms all 3 years achieved reading skills that were well above grade-level expectations by the end of third grade, as measured by nationally normed reading-achievement tests. Keeping in mind that over 45% of the students in our sample qualified for the free and reduced lunch under the U.S. National School Lunch Program,

Table 7. Fixed Effects From the Cross-Classified, Growth-Curve Model on the Grade at Which Students Were in ISI Intervention Classrooms for Reading Score (*z* score)

Predictor	*b*	*b SE*	*t*(880)	*p*
For Intercept 1, π_0				
TTT, γ_{01}	0.497	0.093	5.319	< .001
TTC, γ_{02}	0.529	0.075	7.022	< .001
TCC, γ_{03}	0.511	0.058	8.827	< .001
CCC, γ_{04}	0.318	0.041	7.658	< .001
TCT, γ_{05}	0.514	0.103	4.997	< .001
CTC, γ_{06}	0.271	0.054	4.983	< .001
CCT, γ_{07}	0.271	0.053	5.045	< .001
CTT, γ_{08}	0.539	0.077	7.006	< .001
For Time, π_1				
TTT, γ_{11}	0.035	0.005	6.717	< .001
TTC, γ_{12}	0.027	0.004	6.336	< .001
TCC, γ_{13}	0.038	0.003	11.302	< .001
CCC, γ_{14}	0.041	0.003	13.916	< .001
TCT, γ_{15}	0.034	0.005	5.962	< .001
CTC, γ_{16}	0.053	0.004	14.170	< .001
CCT, γ_{17}	0.036	0.004	8.569	< .001
CTT, γ_{18}	0.043	0.005	8.801	< .001
For Time^2, π_2				
TTT, γ_{21}	−0.001	0.0003	−3.593	< .001
TTC, γ_{22}	−0.0006	0.0003	−2.354	.019
TCC, γ_{23}	−0.001	0.0002	−4.883	< .001
CCC, γ_{24}	−0.0009	0.0002	−5.114	< .001
TCT, γ_{25}	−0.001	0.0004	−2.972	.003
CTC, γ_{26}	−0.002	0.0003	−7.590	< .001
CCT, γ_{27}	−0.0007	0.0002	−3.175	.002
CTT, γ_{28}	−0.002	0.0003	−6.182	< .001
For Time^3, π_3				
TTT, γ_{31}	0.0002	0.00002	8.417	< .001
TTC, γ_{32}	0.0002	0.00002	11.726	< .001
TCC, γ_{33}	0.0001	0.00001	10.782	< .001
CCC, γ_{34}	0.0001	0.00001	10.772	< .001
TCT, γ_{35}	0.0002	0.00002	7.944	< .001
CTC, γ_{36}	0.00009	0.00002	5.757	< .001
CCT, γ_{37}	0.0002	0.00002	10.430	< .001
CTT, γ_{38}	0.0001	0.00002	6.478	< .001

Note: Deviance = 4,781.377294. In group names, the letters (T = treatment, C = control) indicate the order in which the treatments were received from first grade through third grade. Time is in months and is centered at the end of second grade. Time^2 represents the quadratic trend, and Time^3 represents the cubic trend. Intercepts were removed from the model so all groups could be modeled. Thus, intercepts represent end-of-second-grade means for each group. ISI = Individualizing Student Instruction.

Table 8. Random Effects From the Cross-Classified, Growth-Curve Model on the Grade at Which Students Were in ISI Intervention Classrooms for Reading Scores (*z* score)

Effect	*SD*	Variance component	χ^2	*p*
Intercept 1/row	0.53211	0.28314	$\chi^2(772) = 2330.00$	< .001
Time/row	0.01954	0.00038	$\chi^2(772) = 1490.00$	< .001
Level 1, *e*	0.28059	0.07873	—	—
Intercept 1/column c_{38k}	0.17050	0.02907	$\chi^2(30) = 182.21787$	< .001
$Time^3$/CTT	0.00002	0.00000	$\chi^2(22) = 21.91850$	> .500

Note: Level 1 units = 4,777 (repeated measures), row-level units = 882 (students), column-level units = 95 (teachers). We computed χ^2 statistics only for units that had sufficient data. The table shows the final estimation of row and Level 1 variance components and column-level variance components. Time is in months and is centered at the end of second grade. $Time^3$ represents the cubic trend only for the CTT group (who received control mathematics instruction in Grade 1 and reading-intervention treatment in Grades 2 and 3); all other groups were fixed. ISI = Individualizing Student Instruction.

a frequently used marker of family poverty, this is encouraging. Indeed, no student who received 3 years of ISI reading instruction had a standard score below 85 and only 2 of the students who received ISI reading instruction all 3 years had a third-grade reading standard score that was less than 90 (national mean = 100, *SD* = 15; < 85 is considered below grade expectations); 75% of students who received 3 years of ISI reading instruction had standard scores above 100. In contrast, 22% of the students who were in control classrooms all 3 years had standard scores below 90, including some very low scores below 85, compared with the 2 students (6%) who received ISI reading instruction all 3 years.

This study is the first longitudinal cluster-randomized controlled trial to be conducted with the specific research aim of examining the accumulation of instructional effects from first through third grade. The results, coupled with other correlational and quasiexperimental results, stress the importance of providing high quality evidenced-based reading instruction throughout the early grades. This finding has policy and practical implications. The essence of the ISI reading intervention is that literacy instruction is precisely matched to the instructional needs of each individual student within a classroom (Connor, Morrison, Fishman, et al., 2011; Connor, Piasta, et al., 2009). This study provides the strongest evidence to date that sustained, carefully planned, differentiated literacy instruction supported by technology and using the best assessments and research-based strategies available is more likely to be effective than what is typically seen in classrooms. It is likely that as the algorithms used to prescribe amounts and types of reading instruction and technology to support teachers' implementation improve, so too will students' reading outcomes.

Given national policy and media attention regarding teacher accountability, the current findings bear direct relevance to the discussion of teacher value-added scores. The statistical models used in this study are highly similar to those used to compute teacher value-added scores; hence, the positive treatment effect means that teachers in the ISI condition generally had higher value-added scores compared with teachers in the control condition. Most teachers in this study, when given training and support, were able to provide more effective literacy or mathematics instruction in line with their treatment assignment. Thus, as policymakers place more emphasis on the effectiveness of individual teachers and the long-term benefits of effective teaching (Chetty et al., 2012; Konstantopoulos & Chung, 2011), it is clear that what teachers do in the classroom, their use of evidence-based instruction, their ability to differentiate this instruction, and the training and professional development they receive all make a difference in student learning and, hence, in teachers' value-added scores. By implication, teachers who are generally less effective can learn to improve their classroom practices in ways that improve their students' test scores. In this study, participating teachers learned to use the A2i software, assessment information, and classroom implementation well enough within 1 school year to make a meaningful difference in their students' reading achievement when they received a professional development regimen that, while intensive, could be replicated at scale. This study demonstrates that the effectiveness of the current teaching force can be improved and contradicts the assumption that teaching is a talent that some possess and others do not and that there is nothing to be done when teachers have low value-added scores. The discussion around value-added scores can be expanded to focus not just on replacement of teachers with low value-added scores, but also on the training and development of teachers.

Finally, continued rigorous research on effective instructional models is vitally needed to provide the cumulative benefits demonstrated in the present study. Too many curriculums and interventions are brought to schools without rigorous evidence of efficacy, with the explanation that it is not possible to conduct randomized controlled trials in schools. This series of studies and others belie that explanation and call for better evidence for the curricula and interventions used with students. As the understanding about how students learn in the classroom improves, along with how to support teachers' efforts to teach each student effectively, research can increase the effectiveness of instruction and ensure that every child has the skills he or she will need in the future.

Author Contributions

All authors designed the study and interpreted the results. Testing and data collection were performed under the direction of C. M. Connor and E. C. Crowe. C. M. Connor and C. Schatschneider performed the data analyses. C. M. Connor drafted the original manuscript, and all authors contributed to the final version of the manuscript. All authors approved the final version of the manuscript for submission.

Acknowledgments

We thank the Individualizing Student Instruction Project team as well as Phyllis Underwood, Stephanie Day, and Jennifer Dombek, and the children, parents, teachers, and schools participants.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

Funding for this project was provided by the U.S. Department of Education Institute of Education Sciences (R305H04013 and R305B070074) and the National Institute of Child Health and Human Development (R01HD48539, R21HD062834, and P50 HD052120).

Supplemental Material

Additional supporting information may be found at http://pss.sagepub.com/content/by/supplemental-data

Notes

1. Teacher value-added scores reflect the mean gain in achievement of the students in the teacher's classroom. These scores are computed relative to the mean gain in achievement of students in other teachers' classrooms using gain scores or controlling for previous achievement (Raudenbush, 2004).
2. We thank an anonymous reviewer for this observation.
3. The effect size for third grade was not significantly different than 0.

References

Al Otaiba, S., Connor, C. M., Folsom, J. S., Greulich, L., Meadows, J., & Li, Z. (2011). Assessment data-informed guidance to individualize kindergarten reading instruction: Findings from a cluster-randomized control field trial. *The Elementary School Journal*, *111*, 535–560.

Bos, C., Mather, N., Narr, R. F., & Babur, N. (1999). Interactive, collaborative professional development in early literacy instruction: Supporting the balancing act. *Learning Disabilities Research & Practice*, *14*, 227–238.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2012). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood* (Working paper 17699). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://obs.rc.fas.harvard.edu/chetty/value_added.pdf

Connor, C. M., Fishman, B., Crowe, E., Underwood, P., Schatschneider, C., & Morrison, F. J. (2013). Teachers' use of Assessment to Instruction (A2i) software and third graders' reading comprehension gains. In A. Shamir & O. Korat (Eds.), *Technology as a support for literacy achievements for children at risk* (pp. 123–140). New York, NY: Springer.

Connor, C. M., Morrison, F. J., Fishman, B., Giuliani, S., Luck, M., Underwood, P. S., . . . Schatschneider, C. (2011). Testing the impact of Child Characteristics × Instruction interactions on third graders' reading comprehension by differentiating literacy instruction. *Reading Research Quarterly*, *46*, 189–221.

Connor, C. M., Morrison, F. J., Fishman, B. J., Ponitz, C. C., Glasney, S., Underwood, P., . . . Schatschneider, C. (2009). The ISI classroom observation system: Examining the literacy instruction provided to individual students. *Educational Researcher*, *38*, 85–99.

Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). The early years: Algorithm-guided individualized reading instruction. *Science*, *315*, 464–465. doi:10.1126/science.1134513

Connor, C. M., Morrison, F. J., Schatschneider, C., Toste, J., Lundblom, E. G., Crowe, E., & Fishman, B. (2011). Effective classroom instruction: Implications of child characteristic by instruction interactions on first graders' word reading achievement. *Journal of Research on Educational Effectiveness*, *4*, 173–207.

Connor, C. M., Piasta, S. B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E., . . . Morrison, F. J. (2009). Individualizing student instruction precisely: Effects of child x instruction interactions on first graders' literacy development. *Child Development*, *80*, 77–100.

Coyne, M. D., Kame'enui, E. J., Simmons, D. C., & Harn, B. A. (2004). Beginning reading intervention as inoculation or insulin. *Journal of Learning Disabilities*, *37*, 90–104. doi:10.1177/00222194040370020101

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York, NY: Irvington.

DeMaria, M., Rhome, J., Pasch, R., & Clark, J. S. (2009). *Technical summary of the National Hurricane Center track and intensity models*. Retrieved from http://www.nhc.noaa.gov/pdf/model_summary_20090724.pdf

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness*, *2*, 1–44.

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 children. *Journal of Educational Psychology*, *90*, 37–55.

Foorman, B. R., Schatschneider, C., Eakin, M. N., Fletcher, J. M., Moats, L. C., & Francis, D. J. (2006). The impact of instructional practices in grades 1 and 2 on reading and spelling achievement in high poverty schools. *Contemporary Educational Psychology*, *31*, 1–29.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., Karns, K., & Dutka, S. (1997). Enhancing students helping behavior during peer-mediated instruction with conceptual mathematical explanations. *The Elementary School Journal*, *97*, 223–249.

Gamse, B. C., Jacob, R. T., Horst, M., Boulay, B., & Unlu, F. (2008). *Reading First Impact Study final report* (NCEE 2009-4038). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Hill, C., Bloome, H., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*, 172–177.

James-Burdumy, S., Deke, J., Lugo-Gil, J., Carey, N., Hershey, A., Gersten, R., . . . Faddis, B. (2010). *Effectiveness of selected supplemental reading comprehension interventions: Findings from two student cohorts*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Justice, L. M., Petscher, Y., Schatschneider, C., & Mashburn, A. (2011). Peer effects in preschool classrooms: Is children's language growth associated with their classmates' skills? *Child Development*, *82*, 1768–1777.

Konstantopoulos, S., & Chung, N. (2011). The persistence of teacher effects in elementary grades. *American Educational Research Journal*, *48*, 361–386.

Lee, V., & Loeb, S. (1995). Where do Head Start attendees end up? One reason why preschool effects fade out. *Educational Evaluation and Policy Analysis*, *17*, 62–82.

MacGinitie, W. H., & MacGinitie, R. K. (2006). *Gates-MacGinitie Reading Tests* (4th ed.). Iowa City, IA: Houghton Mifflin.

National Assessment of Educational Progress. (2011). *National Assessment of Educational Progress: The nation's report card*. Retrieved from http://nces.ed.gov/nationsreportcard/

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

Pianta, R. C., Belsky, J., Houts, R., & Morrison, F. J. (2007). Teaching: Opportunities to learn in America's elementary classrooms. *Science*, *315*, 1795–1796.

Raudenbush, S. W. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics*, *29*, 121–129.

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

Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2002). Age 21 cost-benefit analysis of the Title I Chicago child-parent centers. *Educational Evaluation and Policy Analysis*, *24*, 267–303.

Rothstein, J. (2008). *Teacher quality in educational production: Tracking, decay, and student achievement* (Working paper 14442). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w14442.pdf?new_window=1

Shadish, W. R., Cook, T. D., & Campbell, J. R. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York, NY: Houghton Mifflin.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, *55*, 151–218.

Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education*. Washington, DC: National Academy Press.

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

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360–407.

U.S. Department of Education. (2010). *Nine states and the District of Columbia win second round Race to the Top grants*. Retrieved from http://www.ed.gov/news/press-releases/nine-states-and-district-columbia-win-second-round-race-top-grants

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (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.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside.