

Intensive Reading Remediation in Grade 2 or 3: Are There Effects a Decade Later?

Benita A. Blachman
Syracuse University

Christopher Schatschneider
Florida State University

Jack M. Fletcher
University of Houston

Maria S. Murray
State University of New York at Oswego

Kristen A. Munger
Syracuse University

Michael G. Vaughn
St. Louis University

Despite data supporting the benefits of early reading interventions, there has been little evaluation of the long-term educational impact of these interventions, with most follow-up studies lasting less than 2 years (Suggate, 2010). This study evaluated reading outcomes more than a decade after the completion of an 8-month reading intervention using a randomized design with 2nd and 3rd graders selected on the basis of poor word-level skills (Blachman et al., 2004). Fifty-eight (84%) of the original 69 participants took part in the study. The treatment group demonstrated a moderate to small effect size advantage on reading and spelling measures over the comparison group. There were statistically significant differences with moderate effect sizes between treatment and comparison groups on standardized measures of word recognition (i.e., Woodcock Basic Skills Cluster, $d = 0.53$; Woodcock Word Identification, $d = 0.62$), the primary, but not exclusive, focus of the intervention. Statistical tests on other reading and spelling measures did not reach thresholds for statistical significance. Patterns in the data related to other educational outcomes, such as high school completion, favored the treatment participants, although differences were not significant.

Keywords: reading difficulties, reading intervention, remediation, longitudinal effects

Despite compelling data supporting the benefits of early reading interventions that include explicit instruction in word-level skills and text-based reading (McCardle, Chhabra, & Kapinus, 2008; National Reading Panel (NRP) Report, 2000; Snow, Burns, & Griffin, 1998; Snowling & Hulme, 2005), there has been little evaluation of the long-term effects of these interventions. In a recent review of reading intervention studies that focused on effect sizes, Suggate (2010) found

that in 85 experimental or quasi-experimental studies, only 21 studies had a long-term follow-up, with a mean length of 15.52 months ($SD = 14.34$). Only one of the 21 studies reported a follow-up that lasted longer than 2 years. Suggate concluded that the results of his review were “restricted by the small number of studies reporting long-term follow-up reading performance” (p. 1568).

There are no studies, to our knowledge, that have investigated whether the reading benefits that result from early and intensive reading remediation are evident as students transition from adolescence to young adulthood. The data reported here are the result of a long-term follow-up conducted approximately 11 years after the completion of a reading remediation study using a randomized design with second and third graders selected on the basis of poor word-level reading skills (Blachman et al., 2004). Knowledge of whether there are long-term benefits is especially important given the evidence that childhood reading difficulties are persistent and follow many into adolescence and adulthood with serious public health consequences, including greater likelihood of dropping out of high school, higher unemployment, and negative health and emotional outcomes (Lyon, 2001).

Persistence of Childhood Reading Difficulties

In a series of studies, Bruck (1990, 1993, 1998) investigated the reading of adults with childhood diagnoses of dyslexia. Her main finding was that although reading disabilities are typically diagnosed in childhood, the “condition is not specific to childhood, but

This article was published Online First August 12, 2013.

Benita A. Blachman, Department of Psychology and Reading and Language Arts Center, Syracuse University; Christopher Schatschneider, Department of Psychology and Florida Center for Reading Research, Florida State University; Jack M. Fletcher, Department of Psychology, University of Houston; Maria S. Murray, Department of Curriculum and Instruction, State University of New York at Oswego; Kristen A. Munger, Reading and Language Arts Center, Syracuse University; Michael G. Vaughn, School of Social Work, St. Louis University.

Kristen A. Munger is now in the Department of Counseling and Psychological Services at State University of New York at Oswego.

This research was supported by National Institute of Child Health and Human Development Grant R21 HD060791. We thank all of the young adults who participated in this research. We also express our sincere appreciation to Michelle Marcoe Storie and Tess Dussling for their contributions to this project.

Correspondence concerning this article should be addressed to Benita A. Blachman, Reading and Language Arts Center, Syracuse University, 200 Huntington Hall, Syracuse, NY 13244. E-mail: blachman@syr.edu

persists into adulthood” (Bruck, 1998, p. 179), with reading and cognitive profiles constant across that time span (see also Maughan et al., 2009; Undheim, 2009). Persistent deficits in single word decoding and a variety of phonological processing skills (e.g., phonological recoding in working memory) among this population of adults with childhood diagnoses of reading disabilities have also been found in both international (Breznitz & Misra, 2003; Paulesu et al., 2001) and North American studies (Ransby & Swanson, 2003; Scarborough, 1984; S. E. Shaywitz et al., 2003; Wilson & Lesaux, 2001).

Other researchers who have followed students from various points in elementary school and tracked their reading progress over time have reached similar conclusions. In a longitudinal study, Juel (1988) found that the probability that a child who was a poor reader in Grade 1 would remain a poor reader in Grade 4 was .88. Francis, Shaywitz, Stuebing, Shaywitz, and Fletcher (1996) tracked children from the Connecticut Longitudinal Study from Grade 1 to Grade 9 and found that it did not matter if the poor readers had been diagnosed because of a discrepancy between IQ and achievement or because of low achievement in reading; less than 25% of children in both groups developed adequate reading skills by Grade 9, with these patterns persisting through Grade 12 (S. E. Shaywitz et al., 1999). Even with a more consistent orthography, like German, researchers Landerl and Wimmer (2008) found that 70% of the dysfluent readers in Grade 1 were still dysfluent in Grade 8, demonstrating the persistence of poor word recognition speed. Similar data regarding the stability of individual differences in word decoding and other reading and language skills were reported in a large study of Dutch children followed through the elementary grades (Verhoeven & van Leeuwe, 2008) and a study in Sweden following children through 12th grade (Svensson & Jacobson, 2006).

Snowling, Muter, and Carroll (2007) followed children with high familial risk for dyslexia and again found that the reading difficulties were persistent, with no evidence of catching up between the ages of 8 and 13 (see also Muter & Snowling, 2009). Similar evidence of stability over time has also been found in the Colorado longitudinal twin studies (Astrom, Wadsworth, Olson, Willcutt, & DeFries, 2011; Hulslander, Olson, Willcutt, & Wadsworth, 2010; Wadsworth, DeFries, Olson, & Willcutt, 2007). For example, Wadsworth et al. followed twins with and without reading difficulties (RD) over a 5- to 6-year period when the students were between the ages of 10 and 16 and found that the RD students continued to have difficulties during the follow-up assessments, with a stability correlation across groups of .80 for a composite measure of reading.

Although sobering, these results are consistent with other data documenting the fact that many children who are receiving remedial reading support in schools are reportedly making little or no progress. As Torgesen et al. (2001) put it, the programs are merely “stabilizing their degree of reading failure” (p. 34). Hanushek, Kain, and Rivkin (1998) reported that 4 years of placement in special education in Grades 3 through 6 was associated with growth in reading skills of .04 standard deviations per year. Researchers have long documented the lack of effectiveness of two common sources of remedial reading instruction provided to public school children: Title 1 reading programs and reading instruction provided in the resource room for children with disabilities (Bentum & Aaron, 2003; Foorman, Francis, Beeler, Winikates, & Fletcher, 1997; Kennedy, Birman, & Demaline, 1986; Moody, Vaughn, Hughes, & Fischer, 2000; Puma et al., 1997).

In these studies, minimal gains relative to initial status were reported, and any gains were lost when children left the program (Birman et al., 1987).

Prognosis for Adolescents and Adults With Low Levels of Literacy

The persistence of RD has negative economic and emotional consequences for adolescents and young adults, starting with the fact that reading skill at the end of third grade is a reasonably accurate predictor of whether or not one will graduate from high school (Snow et al., 1998). In addition to increasing the likelihood of dropping out of school, adolescents with reading problems are at higher risk for behavioral and emotional difficulties, such as depression and anxiety (Arnold et al., 2005; Carroll, Maughan, Goodman, & Meltzer, 2005; Svetaz, Ireland, & Blum, 2000). Although grade retentions still have proponents (Snow et al., 1998), there is no evidence of academic benefit (Silbergitt, Appleton, Burns, & Jimerson, 2006), but not meeting grade expectations in reading in the early grades is the main reason for retaining a child (Snow et al., 1998).

More troubling are data from the National Longitudinal Study of Adolescent Health, indicating that being retained in school is related to “higher levels of emotional distress,” violence, and substance abuse (Resnick et al., 1997, p. 831). As adults, those with poorer literacy skills are much less likely to be employed. The results of the National Assessment of Adult Literacy (National Center for Education Statistics [NCES], 2006), administered to over 19,000 adults, indicated that 14% scored at the “below-basic” level (the lowest of four levels of proficiency) and another 29% scored at the basic level, indicating that they too have limited literacy skills. Of those who scored at the “below-basic” level, 51% were unemployed.

Evidence-Based Interventions for Children With RD and Follow-Up Studies

Given the high estimates of RD (almost one in five in some studies) (S. E. Shaywitz & Shaywitz, 1996), the persistence of these difficulties into adulthood, and the negative economic and health consequences for both individuals and society when adults do not read well, the National Academy of Sciences established a committee to examine the prevention of RD. The report that resulted (Snow et al., 1998) and the follow-up report commissioned by Congress (NRP, 2000) reviewed data from a large body of research and concluded that there was strong support for providing early instruction in phonological awareness and in the systematic relations between spellings and sounds to improve accuracy and fluency of word recognition, while also reaffirming the importance of extensive reading and an early focus on meaning.

Despite extensive evidence that early intervention and remediation programs can reduce the incidence of reading failure (see Brady, 2011; Vellutino & Fletcher, 2005, for reviews), follow-up studies have been limited, with most lasting less than 2 years (Suggate, 2010). Follow-up studies have reported relative stability of gains, with treatment children continuing to significantly outperform controls 1–3 years after the completion of evidence-based interventions provided in kindergarten (Lennon & Slesinski, 1999; Simmons, Coyne, Kwok, Harn, & Kame'enui, 2008; Vadasy,

Sanders, & Peyton, 2006; although see van der Kooy-Hofland, van der Kooy, Bus, van IJzendoorn, & Bonsel, 2012, for a discussion of “differential susceptibility” to treatment effects), first and second grade (Blachman, Tangel, Ball, Black, & McGraw, 1999; Fuchs, Compton, Fuchs, Bryant, & Davis, 2008; Ryder, Tunmer, & Greaney, 2008; Vadasy, Sanders, & Abbott, 2008; Vellutino & Fletcher, 2005), and to somewhat older remedial students in Grades 2 through 6 (e.g., Blachman et al., 2004; Rashotte, MacPhee, & Torgesen, 2001). A longer follow-up was conducted by Elbro and Petersen (2004), who reported positive effects of a kindergarten intervention for at-risk students on multiple measures of reading in Grades 2, 3, and 7. These findings regarding the durability of effects are promising, but there are no long-term follow-up studies that have investigated whether the gains are still evident when students are making the transition from adolescence to early adulthood. In addition, there is little data on whether early reading interventions can reduce the poor educational outcomes (e.g., lower graduation rates) associated with poor development of literacy skills.

The Current Study and Theoretical Considerations

To address the issues of long-term effects of early reading interventions on reading skills and literacy-related educational outcomes, we conducted an 11-year follow-up of a Grade 2–3 cohort who received a reading intervention. In the original study, Blachman et al. (2004) randomly assigned 69 second- and third-grade struggling readers chosen on the basis of poor word recognition to 8 months of explicit reading tutoring or to treatment as usual provided by the schools, with a 1-year follow-up after the research treatment ended. Treatment children significantly outperformed comparison children on word recognition, reading rate, spelling, and passage reading, with respective effect sizes (d) at the end of the treatment year of 1.69, .96, 1.13, and .78; effect sizes 1 year later were .97, .81, .81, and .57, respectively.

In this long-term follow-up, we hypothesized that students who received the 8-month explicit reading treatment would achieve higher reading and spelling outcomes than students who received the regular school-based intervention. Two theoretical frameworks (Share, 1995; Stanovich, 1986) influenced our predictions regarding long-term outcomes. In a classic article, Stanovich (1986) hypothesized that children who got off to a better start in reading would enjoy a “Matthew effect” in which the gap between good and poor readers increased over time. Snowling and Hulme (2005) explained how this model accounts for success or failure in reading: “Children who are reading well experience more print, acquire more vocabulary, and develop even stronger reading skills; conversely, children who are poor readers read less, fail to develop knowledge and vocabulary, thus inhibiting further growth in reading” (p. 543). Based on this framework, we hypothesized that early improvements in reading among treatment participants, as demonstrated in the original study and 1-year follow-up, would result in higher reading and spelling outcomes a decade later and increasing divergence relative to the comparison group. Further support for this prediction comes from Share’s (1995) self-teaching model. He hypothesized that proficiency in decoding acts as a self-teaching mechanism to facilitate retaining words and spellings in memory (Cunningham, Perry, Stanovich, & Share, 2002) and contributes to the acquisition of sight words (Ehri, 2005). In support of the

self-teaching model, more print exposure has been shown to contribute to the development of automaticity of word recognition (Cunningham, Nathan, & Raheer, 2011), and automaticity has been shown to facilitate comprehension (Perfetti, 1985; Rasinski, Reutzel, Chard, & Linan-Thompson, 2011). If children who learn to decode early are exposed to more print (Stanovich, 1986), the opportunities to benefit from Share’s self-teaching model should have been greater for those who participated in the intervention. We also predicted that distal outcomes of the self-teaching hypothesis might extend to other literacy-related outcomes. Specifically, we explored whether treatment and comparison students would differ in years of special education, grade retentions, high school completion, and attendance at a postsecondary institution, hypothesizing that differences would favor the treatment children. To test our hypotheses, we gathered information about a broad range of academic-related outcomes (e.g., high school completion) as well as included a large battery of academic and cognitive measures.

Method

Participants

Original sample. Participants in the long-term follow-up were recruited from the 69 children in the original study (Blachman et al., 2004). Participants in the original study attended 11 schools in four school districts in upstate New York. The districts represented a range of socioeconomic status levels from poor, urban schools to middle-class, suburban schools. Children were eligible for the study if they had a standard score below 90 (< 25th percentile) on either the Word Identification or the Word Attack subtest of the Woodcock Reading Mastery Tests-Revised (Woodcock, 1987) and also a standard score below 90 on the Basic Skills Cluster of the Woodcock. Children also had a Verbal IQ on the Wechsler Intelligence Scale for Children—third edition (WISC-III; Wechsler, 1991) of at least 80. Treatment children in this study were also participants in a neuroimaging study to investigate the influence of intensive reading intervention on patterns of brain activation in children with reading disabilities (B. A. Shaywitz et al., 2004). The neuroimaging component added constraints to the selection criteria. Specifically, all participants had to be right-handed and could not have medical appliances (e.g., pacemaker) that would preclude a magnetic resonance image. Children who were learning English as a second language were not included. The overall sample size was driven by the capacity to image children, for which families had to travel from Syracuse to New Haven, Connecticut.

There were two cohorts of children recruited when the children were at the end of either first or second grade. Children in the first cohort were recruited in the spring of 1997, and children in the second cohort were recruited in the spring of 1998. Data from the two cohorts were aggregated in the analyses. The final sample for the original study included 37 treatment (22 males and 15 females) and 32 comparison children (20 males and 12 females). (For a detailed description of the selection and randomization process, see Blachman et al., 2004).

No significant differences were found in the original study between the treatment and comparison groups on age, $t(67) = -1.00, p = .32$; sex, $\chi^2(1, N = 69) = .07, p = .80$; race (Fisher’s exact test, $N = 69, p = .94$); mother’s educational level, $t(67) = 0.34, p = .73$; or any of the initial screening measures used to determine eligibility, including Verbal IQ, Performance IQ, Full

Scale IQ, Woodcock Basic Skills Cluster, Woodcock Word Identification subtest, or the Woodcock Word Attack subtest. Additionally, the groups did not differ on the number of school absences during the treatment year, $t(67) = 0.24, p = .81$, or the follow-up year, $t(67) = 1.01, p = .30$.

Follow-up sample. Fifty-eight of the original 69 subjects (84%) participated in the long-term follow-up. They were all over 18 years of age ($M = 20.1$ years; $SD = .65$), with a range of 19–22 years of age, and all provided written consent for the follow-up. Initially, contact information (i.e., telephone numbers and addresses of the students' parent[s], their place of employment, names and addresses of two other friends or relatives) provided at the end of the original study was used to locate the participants. For those whose original contact information was no longer viable (e.g., family moved, participant was away at college or military, family no longer used land lines), other sources were used. These included web-based information sources, such as google.com and whitepages.com.

The search for participants was a protracted process that took place between the summer of 2009 and January 2011. Initially, a preliminary phone contact using an Institutional Review Board (IRB)-approved set of "talking points" was made to explain to potential participants that they would be receiving a consent letter for a follow-up study and to provide basic information about the study. Beginning in September 2009, consent letters were mailed to potential participants as they were located. Approximately 1 week after sending the letter, each potential participant was called to answer any questions. If they indicated an interest in participating, they either mailed back a copy of the signed letter or brought it with them to the testing session(s). No testing took place until signed IRB-approved consent letters were returned. Of the original 69 participants, 65 were located and 58 took part in the long-term follow-up study (33 treatment and 25 comparison). Of the 11 who did not participate, four were not located (two treatment, two comparison), two declined (comparison), four did not respond after they were located (two treatment, two comparison), and one died between the time he was located and scheduled for testing (treatment). Number and percentage of participants by gender are shown in Table 1.

Prior to analyzing the results of the long-term follow-up, we looked for differences between the 58 who participated and the 11 participants from the original study on whom we were not able to gather follow-up data. No significant differences were found between these subgroups on age, $t(12.49) = -0.56, p = .582$; race, $\chi^2(3, N = 69) = 2.31, p = .510$; treatment-comparison attrition, $\chi^2(1, N = 69) = 1.57, p = .211$; or any of the initial screening measures used to determine eligibility, including Verbal IQ, Performance IQ, Full Scale IQ, Woodcock Basic Skills Cluster, Woodcock Word Identification, or the Woodcock Word Attack Subtest ($p > .05$). There were significant differences between

these two groups on sex, $\chi^2(1, N = 69) = 4.96, p = .026$; 10 of the 11 missing participants were male. There were also significant differences on mother's educational level, $t(19.97) = -3.44, p = .003$, with a significantly lower educational level for the mothers of the 11 participants on whom we have no data.

There were no differences between the 33 treatment and 25 comparison participants on age, $t(47.93) = -0.77, p = .446$; sex, $\chi^2(1, N = 58) = .18, p = .672$; race, $\chi^2(3, N = 58) = 1.74, p = .628$; mother's educational level, $t(50.64) = 0.62, p = .536$; or any of the initial screening measures used to determine eligibility or additional measures administered at pretest (see Table 2).

Procedure

Original study. During the school year, the treatment children were given 50 min of one-to-one tutoring, 5 days per week, for 8 months. The reading intervention emphasized the phonologic and orthographic connections in words, focusing on accurate decoding and word recognition, fluency, and spelling, as well as text-based reading. Each lesson was built around a five-step plan. Lessons began with a daily review of sound-symbol associations learned in previous lessons and the introduction of new sound-symbol correspondences. This was followed by practice in phoneme analysis and blending during which children manipulated letter cards on a sound board to make phonetically regular words reflecting particular syllable patterns (e.g., to practice closed syllable words, children might be asked to first make the word *fan* and then change it to *fin* and then *shin*). Next, children focused on fluency by quickly reading words made up of previously learned syllable patterns, as well as words that are high frequency and irregular (e.g., *said*). Oral reading practice provided an opportunity for children to read both phonetically controlled text as well as trade books that were not phonetically controlled (e.g., *Amelia Bedelia* series by Peggy Parish). Finally, tutors dictated words used in earlier steps of the lesson or new words with the same phonetic pattern. Children progressed from reading simple closed syllable words (e.g., *fan*) to more complex syllable types (e.g., *hoist*, *storm*) and finally to multisyllable words made of the syllable types they had learned (e.g., *reptile*, *tarnish*).

Treatment fidelity was high. Each child was observed approximately nine times during the treatment year, and two independent raters reviewed each of the observation protocols. Ninety-six percent of the protocols indicated that the lesson included all five steps, and there was 100% interrater agreement. Comparison children continued to receive whatever services the school district provided (e.g., Title 1 reading, reading instruction in the resource room). Seventy-two percent of the comparison group received special reading intervention from their school during the treatment

Table 1
Number and (Percentage) of Participants by Gender

Variable	Treatment			Comparison		
	Male	Female	Total	Male	Female	Total
Original study ($N = 69$)	22	15	37	20	12	32
Long-term follow-up ($n = 58$) and percent of original study participants in follow-up (84%)	19 (86%)	14 (93%)	33 (89%)	13 (65%)	12 (100%)	25 (78%)

Table 2
Pretest, Posttest, and 1-Year Follow-Up Means for Treatment and Comparison Participants in the Long-Term Follow-Up

Measure	Treatment ^a		Comparison ^a	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Pretest only				
Age at entry	95.33	5.86	94.04	6.68
PPVT-R	97.12	12.30	93.52	12.86
WISC-III Verbal IQ	94.70	10.20	92.64	9.44
WISC-III Performance IQ	99.36	14.01	99.48	13.90
WISC-III Full Scale IQ	96.39	10.93	95.32	11.39
Pretest				
WRMT Basic Skills Cluster	81.55	7.25	82.64	6.39
WRMT Word Identification	82.24	6.92	84.36	6.68
WRMT Word Attack	83.94	8.85	82.16	7.48
WRAT Spelling	82.27	6.84	81.12	7.39
GORT Quotient	73.55	8.23	73.96	6.97
GORT Accuracy	5.52	1.28	5.80	1.35
GORT Rate	5.36	1.11	5.60	1.19
GORT Comprehension	5.76	2.37	5.72	1.84
WJ-R Calculations	89.48	12.05	89.44	14.91
WJ-R Applied Problems	99.73	15.28	96.32	12.58
Posttest				
WRMT Basic Skills Cluster	87.15	11.19	79.68	9.56
WRMT Word Identification	87.67	10.48	81.56	9.39
WRMT Word Attack	88.73	13.14	80.44	9.97
WRAT Spelling	91.61	7.57	83.24	10.50
GORT Quotient	84.64	9.42	78.88	10.04
GORT Accuracy	7.21	2.16	6.32	2.06
GORT Rate	6.64	1.58	5.72	1.72
GORT Comprehension	8.06	2.14	6.92	2.34
WJ-R Calculations	94.61	16.82	99.00	20.04
WJ-R Applied Problems	102.79	12.20	103.80	12.84
1-Year Follow-Up				
WRMT Basic Skills Cluster	86.67	10.88	80.24	10.38
WRMT Word Identification	86.36	9.95	80.76	10.67
WRMT Word Attack	88.76	11.49	82.52	10.25
WRAT Spelling	89.94	8.48	83.24	10.35
GORT Quotient	85.09	10.35	81.40	11.91
GORT Accuracy	6.09	2.60	5.84	2.39
GORT Rate	6.36	2.21	5.36	1.98
GORT Comprehension	8.82	2.13	8.40	2.45
WJ-R Calculations	98.33	14.87	95.12	11.80
WJ-R Applied Problems	105.00	9.72	104.72	11.97

Note. PPVT-R = Peabody Picture Vocabulary Test-Revised; WISC-III = Wechsler Intelligence Scale for Children—third edition; WRMT = Woodcock Reading Mastery Tests; WRAT = Wide Range Achievement Test; GORT = Gray Oral Reading Tests; WJ-R = Woodcock-Johnson Psycho-Educational Battery—Revised.

^a treatment group $n = 33$; comparison group $n = 25$.

year. Both groups received their regular classroom reading instruction. All children were assessed before tutoring began, at the end of tutoring, and again 1 year after the posttest by testers blind to the treatment assignment of the children. There was no attrition during the original study, and there were no missing data on any of the measures (see Blachman et al., 2004, for detailed descriptions of measures and intervention).

Long-term follow-up. Most participants were assessed within 2 years ($M = 21.38$ months, $SD = 7.39$) of their high school graduation

date (or the date on which their graduation was projected to have taken place if they had completed high school). In most instances ($n = 44$), participants were tested in a comfortable room designated for this purpose in our university research lab. In instances where participants were unable to travel to the lab (e.g., they did not have transportation, but lived within 1 hr of the lab), arrangements were made to provide transportation. For other participants ($n = 8$), the tester met them in a quiet public location, such as a public or university library, closer to their homes in upstate New York. In other instances, participants resided out of state, and testers traveled to these out-of-state locations such as Texas, Florida, and Tennessee ($n = 6$). Regardless of where testing took place, all testing sites were quiet, public locations. All participants were offered the option of completing the 4- to 5-hr assessment battery in either one or two sessions. Ninety-three percent of the participants chose to complete it in one session.

All assessments were administered in a fixed order. The assessment battery included tests in the following domains: academic (word identification, decoding, fluency, comprehension, spelling, math), cognitive, vocabulary, health literacy, attention, and adaptive functioning (e.g., resilience, peer affiliations, family support). In addition, participants completed two questionnaires (i.e., technology and reading habits and demographic and medication questionnaires). Measures of adaptive functioning, the demographic and medication questionnaire, and the technology and reading habits survey were read aloud to the participants. All other measures were read by the participants, as dictated by the standardized administration procedures. All participants completed all measures in the battery. Immediately upon completion of the testing battery, participants were paid \$200.

Also, upon completion of testing, participants signed forms granting permission to obtain their high school transcripts. The original plans were to obtain from the schools information on years of special education and/or remedial help, but this information was not available on school transcripts and would have required school personnel to retrieve the information from archived data. In most instances, school personnel were unwilling to perform this task. For some participants, these data were even more difficult to acquire because the participant had attended multiple schools. Therefore, the information on special education was obtained through self-report. Transcripts were used to determine retention and graduation rates, and self-report was the source for information regarding completion of a General Educational Development (GED) certificate or attendance at a postsecondary institution.

The analyses in this article focus on the primary outcomes, which involve the academic and cognitive measures, as well as information regarding special education services, grade retention, high school completion, and postsecondary education. Our high school completion data included students who received a high school diploma or a GED certificate. As noted by the NCES (2011),

The General Educational Development (GED) credential is often considered to be the equivalent of a high school diploma for students who do not graduate from a high school Nearly all postsecondary institutions (98 percent) that require high school diplomas for application purposes also recognize the GED credential. (p. 1)

The Census also counts those with GEDs as having graduated (Mishel & Roy, 2006), and the “status completion rate” reported by the NCES (2012) “represents the percentage of 18- through 24-year-olds who

are not enrolled in high school and who have earned a high school diploma or an alternative credential, including a GED certificate” (p. 10). Given that the age of participants in the current study ranged from 19 to 22, it seemed appropriate to follow the standard government convention of including those with a GED when reporting the number of study participants who completed high school and to use the data reported by the NCES (2012) for comparison with the percentages reported in this follow-up study.

Tester qualifications, training, and reliability. Assessments were administered by three testers, all of whom were blind to the treatment assignment of the participants. The testers included the assessment coordinator for the project, who has a doctorate degree in school psychology, and two additional testers – one with a doctorate degree in school psychology and the other an advanced doctoral student in school psychology. Tester training took place over four 3-hr sessions and additional one-on-one practice sessions with the assessment coordinator, for a total of 16 hr of training. Practice time for administration of the full battery was provided during the final session. A one-on-one session with the assessment coordinator was held during which each tester had to administer the entire battery. During this session, testers were evaluated on proficiency of administration according to a 5-point Likert-type scale. Testers had to achieve an overall rating of “1” to be “cleared” for testing. This meant they could have no errors that would impact a student’s score on any measure.

Observations of the two primary testers for the purpose of establishing tester reliability were conducted by the assessment coordinator. Although the assessment coordinator also tested seven of the participants (12%), she was not observed by another tester because her testing typically involved traveling to assess out-of-state participants for whom reliability observations were not practical. Reliability observations of the two testers were conducted during 10 test administrations, including the first test battery administered by each tester (observations were conducted for 17% of the sample). Correlations between testers’ and the reliability observer’s raw scores were calculated for the cognitive and achievement sections for which participants provided verbal responses to test items, comprising 39% of the total assessment battery. For the cognitive and achievement sections of the battery, reliability coefficients ranged from .93 to 1.00, with an overall reliability estimate of .9997. These coefficients suggest that there were minimal discrepancies in the administration and scoring procedures followed by individual testers, thus minimizing tester differences as sources of error in these portions of the assessment battery.

Measures

Long-term follow-up. The subset of academic and cognitive measures administered as part of the 5-hr battery are described below.

Woodcock Reading Mastery Tests—Revised (WRMT–R), Word Identification subtest and Word Attack subtest (Form G; Woodcock, 1987). The Word Identification subtest measures word recognition by having participants read words from a graded word list. The median split-half reliability for the Word Identification subtest (Form G) is .97 (Woodcock, 1987). The Word Attack subtest measures word attack skills by having participants read decodable nonwords that increase in difficulty. The median split-half reliability for the Word Attack subtest (Form G) is .87 (Woodcock, 1987). The median split-half reliability for the Basic Skills Cluster (Form

G), the composite of Word Identification and Word Attack, is .96 (Woodcock, 1987).

Test of Word Reading Efficiency (TOWRE): Sight Word Efficiency subtest and Phonemic Decoding Efficiency subtest (Torgesen, Wagner, & Rashotte, 1999). On the Sight Word Efficiency subtest, participants read from a graded word list as quickly as possible, skipping over any unknown words. On the Phonemic Decoding Efficiency subtest, participants read from a pronounceable list of nonwords as quickly as possible. Both subtests have a time limit of 45 s. The average alternate forms reliability coefficient is .93 for Sight Word Efficiency and .94 for Phonemic Decoding Efficiency. Total average Word Reading Efficiency reliability is reported to be .96 (Torgesen et al., 1999).

Group Reading Assessment and Diagnostic Evaluation (GRADE), Passage Comprehension (Level A; Williams, 2001). GRADE Passage Comprehension measures the ability to comprehend text and apply comprehension strategies (e.g., questioning, summarizing, clarifying, and predicting). Level A is recommended for advanced high school and postsecondary individuals. Participants read extended passages silently and answer comprehension questions. This subtest is untimed. The internal consistency coefficient for Level A is .95 for 12th grade and .85 for postsecondary.

WRMT–R, Passage Comprehension subtest (Form G; Woodcock, 1987). The Passage Comprehension subtest measures the ability to read a short passage silently (generally two to three sentences in length) and provide a key word missing from the passage. The subtest is not timed. The median split-half reliability for the Passage Comprehension subtest (Form G) is .92 (Woodcock, 1987).

Wide Range Achievement Test 4 (WRAT4), Spelling (Green Form; Wilkinson & Robertson, 2006). The WRAT4 spelling subtest is individually administered to assess ability to write single words from dictation. Median internal consistency across all ages for this subtest is .90 (Wilkinson & Robertson, 2006).

Woodcock-Johnson-III—Tests of Achievement, Applied Problems subtest (Woodcock, McGrew, & Mather, 2001). The Applied Problems subtest measures the ability to solve word problems read by the examiner. The median internal consistency coefficient for Applied Problems is .93 for all ages (McGrew & Woodcock, 2001).

Peabody Picture Vocabulary Test—fourth edition (PPVT-IV) (Form A; Dunn & Dunn, 2007). This test measures receptive vocabulary and can be used as a measure of verbal ability. Participants point to one of four pictures that match a word spoken by the examiner. The median internal consistency for this subtest across all ages is .94 (Dunn & Dunn, 2007).

Wechsler Adult Intelligence Scale—fourth edition (WAIS-IV), Matrix Reasoning (Wechsler, 2008). The Matrix Reasoning subtest measures nonverbal reasoning ability. For each item, the participant identifies a missing item from a pattern or series (e.g., a sequence of alternating large and small circles are shown, and the participant selects the figure that occurs next from five choices). Split-half reliability for Matrix Reasoning for young adults age 18–24 years ranged from .87 to .88. Split-half reliability estimates for Matrix Reasoning for individuals with reading disabilities ($N = 34$) is estimated at .93.

Original pretest measures used as covariates. As explained in more detail in the Results section, for the academic achievement measures assessed in the long-term follow-up, we used comparable pretest variables measured in the original study as covariates. If the

same measure was not used in the earlier study, we covaried on a pretest measure that was aligned with the same construct.

WRMT-R, Word Identification subtest and Word Attack subtest (Form G; Woodcock, 1987). Both the WRMT-R word identification and Word Attack subtests were administered as pretests in the original study and again during the long-term follow-up. The pretest measures were used as covariates when analyzing the follow-up. The pretest WRMT-R Basic Skills Cluster was used as a covariate when differences between groups on that composite score were analyzed.

Word Reading Efficiency (prepublication version of the TOWRE; Torgesen et al., 1999). This prepublication version of the TOWRE was administered as a pretest and was used as a covariate when analyzing group differences on both subtests of the published TOWRE (the Sight Word Efficiency subtest and Phonemic Decoding Efficiency subtest; Torgesen et al., 1999) administered during the follow-up. For the prepublication version, children read as quickly as possible from a graded word list (Form A), skipping over any unknown words. After 45 s, the examiner instructed the child to stop, and a second timed trial with a new, but comparable, list of words (Form B) was administered. Speed (number of correctly read words per second) was computed for each version of the test, and the two speed measures were averaged. Test-retest reliability of this average based on data collected for the original study was .95.

Gray Oral Reading Tests—third edition (GORT-3, Form A: Wiederholt & Bryant, 1992). The GORT-3 measures reading accuracy, rate, and comprehension through timed reading of up to 13 graded passages. Testing yields individual subtest scores as well as an overall oral reading quotient. The overall pretest oral reading quotient was used as a covariate in analyses of group differences on both comprehension measures administered in the follow-up study (i.e., GRADE and WRMT-R Passage Comprehension). Median internal consistency across all ages for the GORT-3 oral reading quotient is .97 (Wiederholt & Bryant, 1992).

Wide Range Achievement Test 3 (WRAT3), Spelling (Tan Form; Wilkinson, 1993). The WRAT3 spelling subtest administered as a pretest was used as the covariate when analyzing group differences on the WRAT4 spelling subtest administered in the follow-up. Median internal consistency across all ages for the WRAT3 spelling subtest is .89 (Wilkinson, 1993).

Woodcock-Johnson Psycho-Educational Battery—Revised (WJ-R), Tests of Achievement, Applied Problems subtest (Woodcock & Johnson, 1989). The WJ-R Applied Problems subtest administered as a pretest was used as a covariate when analyzing group differences on the WJ-III Applied Problems subtest administered in the follow-up. Internal consistency for the WJ-R Applied Problems subtest is .91 (Woodcock & Mather, 1989).

PPVT-R (Dunn & Dunn, 1981). The PPVT-R administered as a pretest was used as a covariate when analyzing group differences on the PPVT-IV administered in the follow-up. The median split-half reliability across all ages for the PPVT-R is .82 (Dunn & Dunn, 1981).

WISC-III (Wechsler, 1991). The WISC-III Performance IQ administered during initial eligibility screening for the original study was used as a covariate when analyzing group differences on the WAIS-IV Matrix Reasoning subtest (Wechsler, 2008) administered in the follow-up. The average split-half reliability across all ages for the WISC-III Performance IQ is .91 (Wechsler, 1991).

Results

To investigate differences between the treatment and comparison groups in the long-term follow-up, we conducted a series of analyses that used original group assignment (treatment vs. comparison) as the main independent variable. For the academic achievement measures assessed in the long-term follow-up, we performed analyses of covariance that included the grouping variable as well as the pretest variable measured before the intervention began. For achievement assessments that did not have a direct pretest measure, we chose a pretest measure that was aligned with the same construct being assessed. For the academically related indices (e.g., grade retention), we did not use a covariate and instead performed either an ANOVA or chi-square tests, depending on the nature of the variable. All statistical tests were conducted at an alpha threshold of $p < .05$, with no correction for the number of dependent tests to avoid Type II errors. Effect sizes were computed for each comparison. Although the original study was adequately powered to detect the anticipated effects of the intervention at posttest, it was not designed with detection of long-term effects in mind.

Treatment Group Effects

Means and standard deviations for the groups at pretest, posttest, and 1-year follow-up for the 58 participants are shown in Table 2. The bar graph in Figure 1 depicts the standard score means on the WRMT-R Basic Skills Cluster (a composite of the Word Identification and Word Attack subtests) for the treatment and comparison groups at pretest, posttest, 1-year follow-up, and long-term follow-up.

Table 3 reports means and standard deviations for the long-term follow-up data. There were statistically significant differences

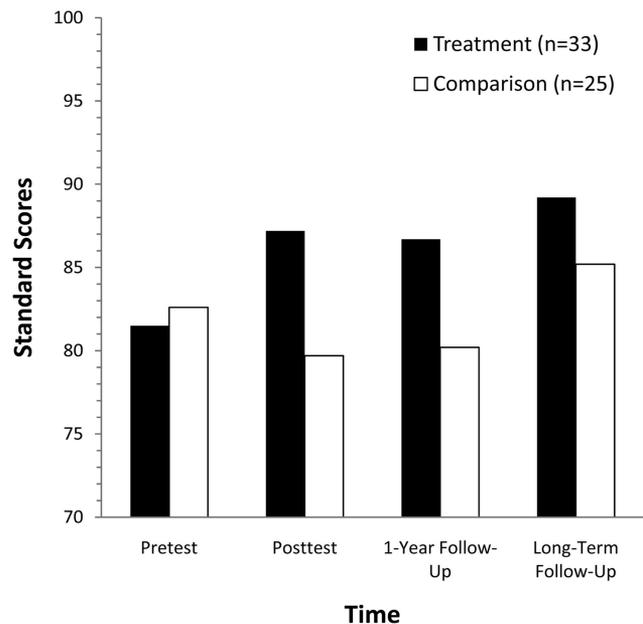


Figure 1. Treatment and comparison group standard score means on the Basic Skills Cluster of the Woodcock Reading Mastery Tests-Revised (Woodcock, 1987) at pretest, posttest, 1-year follow-up, and long-term follow-up.

Table 3
 Long-Term Follow-Up Means and Adjusted Means for Treatment and Comparison Participants in the Long-Term Follow-Up

Measure	Treatment ^a			Comparison ^a			<i>F</i> ^c	<i>p</i>	<i>d</i> ^d
	<i>M</i>	<i>SD</i>	<i>M</i> ^b	<i>M</i>	<i>SD</i>	<i>M</i> ^b			
PPVT-IV	94.76	13.75	93.52	89.28	10.91	90.92	1.58	.213	0.21
WRMT Basic Skills Cluster	89.24	9.50	89.56	85.24	8.44	84.82	5.15	.027	0.53
WRMT Word Identification	81.21	13.69	81.77	74.44	12.34	73.70	5.73	.020	0.62
WRMT Word Attack	90.33	11.87	89.80	87.44	11.08	88.15	0.38	.542	0.14
WRMT Passage Comprehension	90.39	11.38	90.47	89.76	13.70	89.66	0.06	.801	0.06
GRADE Comprehension	84.58	16.56	84.71	82.68	16.60	82.51	0.28	.601	0.13
TOWRE Sight Word Efficiency	82.67	9.44	82.43	80.36	9.57	80.68	0.57	.453	0.18
TOWRE Phonemic Decoding	82.15	11.04	81.85	78.68	10.87	79.08	1.12	.295	0.25
WRAT Spelling	87.39	8.26	87.07	84.08	10.03	84.50	1.50	.226	0.28
WJ-R Applied Problems	90.55	8.13	90.03	89.68	11.47	90.37	0.02	.879	−0.04
WAIS Matrix Reasoning	9.45	2.28	9.46	9.92	2.16	9.92	0.68	.413	−0.21
Grade retentions	0.28	0.46		0.48	0.59		2.07	.156	−0.39

Note. PPVT-IV = Peabody Picture Vocabulary Test—fourth edition; WRMT = Woodcock Reading Mastery Tests; GRADE = Group Reading Assessment and Diagnostic Evaluation; TOWRE = Test of Word Reading Efficiency; WRAT = Wide Range Achievement Test; WJ-R = Woodcock-Johnson Psycho-Educational Battery—Revised; WAIS = Wechsler Adult Intelligence Scale.

^a treatment group *n* = 33; comparison group *n* = 25. ^b Means adjusted by pretest. ^c *F* ratios with one degree of freedom in the numerator and 55 degrees of freedom in the denominator, with the exception of grade retention, which has one degree of freedom in the numerator and 56 degrees of freedom in the denominator. ^d Cohen’s (1988) measure of effect size using the adjusted means in the numerator and the observed pooled standard deviation in the denominator.

between the treatment and comparison groups on the Woodcock Basic Skills Cluster, $F(1, 55) = 5.15, p = .027$, and Woodcock Word Identification, $F(1, 55) = 5.73, p = .020$. Differences between other variables shown in Table 3 did not meet the critical level of alpha ($p < .05$). With the exception of moderate effect sizes for Woodcock Basic Skills Cluster ($d = 0.53$) and Word Identification ($d = 0.62$), all of the other reading and spelling effect sizes were small to negligible. The range was from .28 for WRAT Spelling to .06 for Woodcock Passage Comprehension. The mean effect size for the seven reading and spelling measures in Table 3 (excluding Woodcock Basic Skills Cluster, which is a composite of Word Identification and Word Attack) was .24, with an effect size of .21 for the receptive language measure. The effect size for grade retentions ($d = -.39$) also favored the treatment participants.

Chi-square analyses revealed no significant differences between treatment and comparison participants on the variables in Table 4, although patterns in the data again favor the treatment group. For example, fewer treatment participants reported receiving special education services, more treatment students completed high school, and more treatment participants reported participating in postsecondary education (attending either a trade school or 2-year or 4-year college). With regard to high school completion, 85% of treatment students and 68% of comparison students had a high school diploma or a GED by the time they were assessed for the follow-up study. Treatment student data compares favorably with data reported by the NCES (2012) regarding students with and without disabilities. Specifically, “In 2009, some 89.8 percent of 18- through 24-year-olds not enrolled in high school had received a high school diploma or alternative credential” (NCES, 2012, p. 10). The percentage was somewhat lower (80%) for those with disabilities (NCES, 2012, p. 11).

Discussion

The major finding was that the treatment group demonstrated a moderate to small effect size advantage on reading and spelling measures over the comparison group more than a decade after a reading intervention that took place when the students were in Grades 2 or 3. The results provided some support for our hypothesis that students who received the 8-month explicit reading treatment would achieve higher reading (although not spelling) outcomes than students who received the regular school-based intervention when the effect size data are considered. However, the results were not consistent with the hypothesis that these postintervention differences would increase because of the Mathew effect. Instead, the findings are more consistent with studies of the Mathew effect in *untreated* longitudinal samples that do not find evidence of increased divergence over time among good and poor readers for decoding skills (B. A. Shaywitz et al., 1995). A year of successful reading intervention seems to shift the poor readers upward, but there is stability at that point, at least after 1 year of intervention. In addition, the evidence that these effects would extend to other domains and outcomes because of self-teaching was not strongly supported. There were statistically significant differences with moderate effect sizes between treatment and comparison groups on standardized measures of word recognition, the primary, but not exclusive, focus of the intervention, but statistical tests on other reading and spelling measures, as well as measures of other educational outcomes, did not meet the critical level of alpha.

As discussed earlier, a recent review of reading intervention studies (Suggate, 2010) that focused on effect sizes revealed that in a database of 85 experimental or quasi-experimental studies, only 21 studies had a long-term follow-up, nine of which were characterized as “phonics” studies (including the original study on which this follow-up is based; Blachman et al., 2004). The mean length of the follow-up across all 21 studies was 15 months ($SD = 14.34$).

Table 4
Percent of Participants Who Received Special Education Services, Completed High School, and Are Attending a Postsecondary Institution

Variable	Treatment (<i>n</i> = 33)	Comparison (<i>n</i> = 25)
Special education services	42	60
High school completion	85	68
Postsecondary education	70	52

The mean weighted effect size calculated from the nine phonics studies was .32. The mean effect size for reading and spelling measures in our long-term follow-up ($d = 0.24$) compares favorably with the mean effect size of follow-up studies that were much shorter in duration. Our mean effect size also compares favorably with the mean effect size ($d = 0.25$) based on data collected from three reading subtests of the Woodcock at the end of Grade 2 after 3 years of participation in the Success For All school reform effort (Borman et al., 2007). The take away message, however, is that the field is clearly lacking the long-term follow-up studies needed to understand how and whether early reading interventions influence long-term school performance and outcomes after high school.

What might be required to enhance the long-term outcomes of an early reading intervention like the one in the original study, especially given the school factors that work against maintaining gains (e.g., evidence that public school remedial and special education programs do little more than maintain the students' degree of reading failure; Torgesen, 2005). Ideally, one would want to build on the initial large effects seen immediately posttreatment on word recognition, reading rate, spelling, and passage reading (with respective effect sizes of 1.69, .96, 1.13, and .78) by providing the kind of extended instruction that would facilitate an accelerated growth rate over time, especially in fluency (automaticity) and comprehension. To close the achievement gap between struggling readers and typical readers, more extensive efforts are clearly required.

In the original study and 1-year follow-up, Blachman et al. (2004) reported that, although students largely maintained their significant posttreatment gains at the 1-year follow-up, the significant differences in *rates of growth* that favored the treatment group during the treatment year were no longer evident during the follow-up year. As we explained at the time:

During the follow-up year, however, when children in both groups received only the standard services traditionally available in schools, rates of growth between the two groups no longer differed. Because school-based treatments (either from Chapter 1 or resource teachers) have been shown to be relatively ineffective (Kennedy et al., 1986; Moody et al., 2000; Puma et al., 1997; Snow et al., 1998), it is not completely surprising that the rate of growth of our children slowed when they returned to standard instruction. The challenge in translating research to practice is to alter standard instruction so that an accelerated growth trajectory is the norm, rather than the exception, for low-achieving children. (Blachman et al., 2004, p. 458)

Thus, 1 year of reading intervention in second or third grade did not appear to be adequate to strongly accelerate growth in subsequent years. In a recent series of adolescent reading interventions summarized in Vaughn and Fletcher (2012), 1 year of intervention produced small effects that largely were not statistically significant

in a much larger sample. However, continuing the intervention with adolescents who had not responded adequately for 2–3 years led to moderate effect size advantages after the second year and a large effect on reading comprehension after 3 years, the latter partly because the standard scores for the comparison group declined from Grades 6 to 8.

A number of factors might have influenced the drop off in acceleration in the Blachman et al. (2004) study. It is possible that some of the treatment children did not end the treatment at a level of proficiency in decoding skills that would fully engage the self-teaching mechanism described by Share's (1995) theoretical model in which correct decoding of new words facilitates retaining those words in memory and contributes to the acquisition of sight words (Ehri, 2005) and memory for spellings (Cunningham et al., 2002). This might have been especially true for treatment children who began the study with lower initial skills at screening. Although in the original study there were no reliable differences in gain scores during the treatment year between treatment children who began the study with lower initial skills and those who began with higher initial skills, lower skilled children were less likely to complete the program. Initial reading skills and level of program reached by the children at the end of the treatment year were both significantly related to end-of-year reading and spelling scores, prompting us to speculate that a "longer or more intense program may be especially important for the lowest achieving children" (Blachman et al., 2004, p. 458). Also, as noted in Blachman et al., students differed significantly on the one measure of comprehension administered at the end of the treatment year ($d = 0.55$), but this difference was not statistically significant 1 year later ($d = 0.24$). Evidence suggests that explicit instruction can improve not only word reading and spelling but also comprehension (Rand Reading Study Group, 2002), and, as we reported, this was the area in which our program was the least explicit and systematic. It remains to be seen whether remediation research similar to the Blachman et al. study, but with attention to some of the variables noted above (e.g., a longer, multiyear intervention), would produce larger effect sizes in an extended follow-up study like the one described in this article.

An important limitation to this study is the small sample size. The original sample ($N = 69$) was limited partly because of financial constraints imposed by the cost of the neuroimaging component of the research (see B. A. Shaywitz et al., 2004, for details). Although 84% of the original subjects participated in the long-term follow-up, the sample size was far from ideal. We also had to rely on self-report data regarding special education services. As noted earlier, our original intent was to get this information from the schools. However, because this information was not typically included on transcripts, school personnel would have had to access archived data, and most were unwilling to do so. Finally, there were significant differences in those who participated in the long-term follow-up and those who did not participate in terms of mother's educational level and gender. It is unclear how or whether the outcomes would have been affected if these differences did not exist. Given the lack of long-term follow-up studies in the reading intervention literature, we believe that this study, despite its limitations, provides an important contribution in evaluating the long-term benefits of reading remediation in the primary grades.

The results from this long-term follow-up provide further support for the hypothesis that reading intervention (especially when provided to remedial students, as opposed to younger at-risk students in kindergarten and first grade) is more appropriately viewed as analogous to insulin therapy, rather than as an inoculation against further reading failure (see Coyne, Kame'enui, Simmons, & Harn, 2004, for a discussion of this debate). That is, students in need of explicit and systematic instruction in the early stages of reading acquisition are likely to require ongoing evidence-based support to acquire more complex skills. The challenge posed by Blachman et al. (2004) almost a decade ago to alter standard instruction so that an accelerated growth trajectory is the norm remains a challenge for the field today.

References

- Arnold, E. M., Goldston, D., Walsh, A. K., Reboussin, B. A., Daniel, S. S., Hickman, E., & Wood, F. B. (2005). Severity of emotional and behavioral problems among poor and typical readers. *Journal of Abnormal Child Psychology*, *33*, 205–217. doi:10.1007/s10802-005-1828-9
- Astrom, R. L., Wadsworth, S. J., Olson, R. K., Willcutt, E. G., & DeFries, J. C. (2011). DeFries-Fulker analysis of longitudinal reading performance data from twin pairs ascertained for reading difficulties from their nontwin siblings. *Behavior Genetics*, *41*, 660–667. doi:10.1007/s10519-011-9445-6
- Bentum, K. E., & Aaron, P. G. (2003). Does reading instruction in learning disability resource rooms really work?: A longitudinal study. *Reading Psychology*, *24*, 361–369. doi:10.1080/02702710390227387
- Birman, B. F., Orland, M. E., Jung, R. K., Anson, R. J., Garcia, G. N., Moore, M. T., . . . Reisner, E. R. (1987). *The current operation of the Chapter 1 program*. Washington, DC: U. S. Government Printing Office.
- Blachman, B. A., Schatschneider, C., Fletcher, J. M., Francis, D. J., Clonan, S. M., Shaywitz, B. A., & Shaywitz, S. E. (2004). Effects of intensive reading remediation for second and third graders and a 1-year follow-up. *Journal of Educational Psychology*, *96*, 444–461. doi:10.1037/0022-0663.96.3.444
- Blachman, B. A., Tangel, D. M., Ball, E. W., Black, R., & McGraw, C. K. (1999). Developing phonological awareness and word recognition skills: A two-year intervention with low-income, inner-city children. *Reading and Writing: An Interdisciplinary Journal*, *11*, 239–273. doi:10.1023/A:1008050403932
- Borman, G. D., Slavin, R. E., Cheung, A. C. K., Chamberlain, A. M., Madden, N. A., & Chambers, B. (2007). Final reading outcomes of the national randomized field trial of Success for All. *American Educational Research Journal*, *44*, 701–731. doi:10.3102/0002831207306743
- Brady, S. (2011). Efficacy of phonics teaching for reading outcomes: Indicators from post-NRP research. In S. A. Brady, D. Braze, & C. A. Fowler (Eds.), *Explaining individual differences in reading: Theory and evidence* (pp. 69–96). New York, NY: Psychology Press.
- Breznitz, Z., & Misra, A. (2003). Speed of processing of the visual-orthographic and auditory-phonological systems in adult dyslexics: The contribution of “asynchrony” to word recognition deficits. *Brain and Language*, *85*, 486–502. doi:10.1016/S0093-934X(03)00071-3
- Bruck, M. (1990). Word-recognition skills of adults with childhood diagnoses of dyslexia. *Developmental Psychology*, *26*, 439–454. doi:10.1037/0012-1649.26.3.439
- Bruck, M. (1993). Component spelling skills of college students with childhood diagnoses of dyslexia. *Learning Disability Quarterly*, *16*, 171–184. doi:10.2307/1511325
- Bruck, M. (1998). Outcomes of adults with childhood histories of dyslexia. In C. Hulme & R. M. Joshi (Eds.), *Reading and spelling: Development and disorders* (pp. 179–200). Mahwah, NJ: Erlbaum.
- Carroll, J. M., Maughan, B., Goodman, R., & Meltzer, H. (2005). Literacy difficulties and psychiatric disorders: Evidence for comorbidity. *Journal of Child Psychology and Psychiatry*, *46*, 524–532. doi:10.1111/j.1469-7610.2004.00366.x
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Coyne, M. D., Kame'enui, E. J., Simmons, D. C., & Harn, B. A. (2004). Beginning reading intervention as inoculation or insulin: First-grade reading performance of strong responders to kindergarten intervention. *Journal of Learning Disabilities*, *37*, 90–104. doi:10.1177/00222194040370020101
- Cunningham, A. E., Nathan, R. G., & Raheer, K. S. (2011). Orthographic processing in models of word recognition. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. P. Afflerbach (Eds.), *Handbook of reading research*: Vol. 4 (pp. 259–285). New York, NY: Routledge.
- Cunningham, A. E., Perry, K. E., Stanovich, K., & Share, D. (2002). Orthographic learning during reading: Examining the role of self-teaching. *Journal of Experimental Child Psychology*, *82*, 185–199. doi:10.1016/S0022-0965(02)00008-5
- Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test-Revised (PPVT-R)*. Circle Pines, MN: American Guidance Service.
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test-Fourth Edition (PPVT-IV)*. Minneapolis, MN: NCS Pearson.
- Ehri, L. C. (2005). Development of sight word reading: Phases and findings. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 135–154). Oxford, England: Blackwell Publishing. doi:10.1002/9780470757642.ch8
- Elbro, C., & Petersen, D. K. (2004). Long-term effects of phoneme awareness and letter sound training: An intervention study with children at risk for dyslexia. *Journal of Educational Psychology*, *96*, 660–670. doi:10.1037/0022-0663.96.4.660
- Foorman, B. R., Francis, D. J., Beeler, T., Winikates, D., & Fletcher, J. M. (1997). Early interventions for children with reading problems: Study designs and preliminary findings. *Learning Disabilities*, *8*, 63–71.
- 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. doi:10.1037/0022-0663.88.1.3
- Fuchs, D., Compton, D. L., Fuchs, L. S., Bryant, J., & Davis, G. N. (2008). Making “secondary intervention” work in a three-tier responsiveness-to-intervention model: Findings from the first-grade longitudinal reading study of the National Research Center on Learning Disabilities. *Reading and Writing: An Interdisciplinary Journal*, *21*, 413–436. doi:10.1007/s11145-007-9083-9
- Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (1998). *Does special education raise academic achievement for students with disabilities?* (Working Paper No. 6690). Cambridge, MA: National Bureau of Economic Research.
- Hulslander, J., Olson, R. K., Willcutt, E. G., & Wadsworth, S. J. (2010). Longitudinal stability of reading-related skills and their prediction of reading development. *Scientific Studies of Reading*, *14*, 111–136. doi:10.1080/10888431003604058
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology*, *80*, 437–447. doi:10.1037/0022-0663.80.4.437
- Kennedy, M. M., Birman, B. F., & Demaline, R. E. (1986). *The effectiveness of Chapter 1 services: An interim report from the national assessment of Chapter 1*. Washington, DC: U. S. Department of Education.
- Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology*, *100*, 150–161. doi:10.1037/0022-0663.100.1.150

- Lennon, J. E., & Slesinski, C. (1999). Early intervention in reading: Results of a screening and intervention program for kindergarten students. *School Psychology Review*, 28, 353–364.
- Lyon, G. R. (2001, March 8). *Measuring success: Using assessments and accountability to raise student achievement (statement to the Committee on Education and the Workforce)*. Retrieved from <http://www.hhs.gov/asl/testify/t010308.html>
- Maughan, B., Messer, J., Collishaw, S., Pickles, A., Snowling, M., Yule, W., & Rutter, M. (2009). Persistence of literacy problems: Spelling in adolescence and at mid-life. *Journal of Child Psychology and Psychiatry*, 50, 893–901. doi:10.1111/j.1469-7610.2009.02079.x
- McCardle, P., Chhabra, V., & Kapinus, B. (2008). *Reading research in action: A teacher's guide for student success*. Baltimore, MD: Brookes.
- McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock-Johnson-III Technical Manual*. Itasca, IL: Riverside.
- Mishel, L., & Roy, J. (2006). Accurately assessing high school graduation rates. *Phi Delta Kappan*, 88, 287–292.
- Moody, S. W., Vaughn, S., Hughes, M. T., & Fischer, M. (2000). Reading instruction in the resource room: Set up for failure. *Exceptional Children*, 66, 305–316.
- Muter, V., & Snowling, M. J. (2009). Children at familial risk of dyslexia: Practical implications from an at-risk study. *Child and Adolescent Mental Health*, 14, 37–41. doi:10.1111/j.1475-3588.2007.00480.x
- National Center for Education Statistics. (2006). *The health literacy of America's adults: Results from the 2003 National Assessment of Adult Literacy*. Retrieved from <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006483>
- National Center for Education Statistics. (2011). *Issue brief: Characteristics of GED recipients in high school: 2002–06 (NCES 2012–025)*. Retrieved from <http://nces.ed.gov/pubs2012/2012025.pdf>
- National Center for Education Statistics. (2012). *Trends in high school dropout and completion rates in the United States: 1972–2009 (IES 2012–006)*. Retrieved from <http://nces.ed.gov/pubs2012/2012006.pdf>
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implication for reading instruction: Reports of the subgroups*. Washington, DC: National Institute of Child Health and Human Development.
- Paulesu, E., Démonet, J-F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., . . . Frith, U. (2001). Dyslexia: Cultural diversity and biological unity. *Science*, 291, 2165–2167. doi:10.1126/science.1057179
- Perfetti, C. A. (1985). *Reading ability*. New York, NY: Oxford University Press.
- Puma, M. J., Karweit, N., Price, C., Ricciuti, A., Thompson, W., & Vaden-Kiernan, M. (1997). *Prospects: Final report on student outcomes*. Cambridge, MA: ABT Associates. (ERIC Document Reproduction Service No. ED413411).
- Rand Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.
- Ransby, M. J., & Swanson, H. L. (2003). Reading comprehension skills of young adults with childhood diagnoses of dyslexia. *Journal of Learning Disabilities*, 36, 538–555. doi:10.1177/00222194030360060501
- Rashotte, C. A., MacPhee, K., & Torgesen, J. K. (2001). The effectiveness of a group reading instruction program with poor readers in multiple grades. *Learning Disability Quarterly*, 24, 119–134. doi:10.2307/1511068
- Rasinski, T. V., Reutzel, D. R., Chard, D., & Linan-Thompson, S. (2011). Reading fluency. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. P. Afflerbach (Eds.), *Handbook of reading research: Vol. 4* (pp. 320–338). New York, NY: Routledge.
- Resnick, M. D., Bearman, P. S., Blum, R. W., Bauman, K. E., Harris, K. M., Jones, J., . . . Sieving, R. E. (1997). Protecting adolescents from harm: Findings from the National Longitudinal Study on Adolescent Health. *Journal of the American Medical Association*, 278, 823–832. doi:10.1001/jama.1997.03550100049038
- Ryder, J. F., Tunmer, W. E., & Greaney, K. T. (2008). Explicit instruction in phonemic awareness and phonemically based decoding skills as an intervention strategy for struggling readers in whole language classrooms. *Reading and Writing: An Interdisciplinary Journal*, 21, 349–369. doi:10.1007/s11145-007-9080-z
- Scarborough, H. S. (1984). Continuity between childhood dyslexia and adult reading. *British Journal of Psychology*, 75, 329–348. doi:10.1111/j.2044-8295.1984.tb01904.x
- Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, 55, 151–218. doi:10.1016/0010-0277(94)00645-2
- Shaywitz, B. A., Holford, T. R., Holahan, J. M., Fletcher, J. M., Francis, D. J., Stuebing, K. K., & Shaywitz, S. E. (1995). A Matthew effect for IQ but not for reading: Results from a longitudinal study of reading. *Reading Research Quarterly*, 30, 894–906. doi:10.2307/748203
- Shaywitz, B. A., Shaywitz, S. E., Blachman, B. A., Pugh, K. R., Fulbright, R. K., Skudlarski, P., . . . Gore, J. C. (2004). Development of left occipitotemporal systems for skilled reading in children after a phonologically-based intervention. *Biological Psychiatry*, 55, 926–933. doi:10.1016/j.biopsych.2003.12.019
- Shaywitz, S. E., Fletcher, J. M., Holahan, J. M., Shneider, A. E., Marchione, K. E., Steubing, K. K., . . . Shaywitz, B. A. (1999). Persistence of dyslexia: The Connecticut Longitudinal Study at adolescence. *Pediatrics*, 104, 1351–1359. doi:10.1542/peds.104.6.1351
- Shaywitz, S. E., & Shaywitz, B. A. (1996). Unlocking learning disabilities: The neurological basis. In S. C. Cramer & W. Ellis (Eds.), *Learning disabilities: Lifelong issues* (pp. 255–260). Baltimore, MD: Paul H. Brookes.
- Shaywitz, S. E., Shaywitz, B. A., Fulbright, R. K., Skudlarski, P., Mencl, W. E., Constable, R. T., . . . Gore, J. C. (2003). Neural systems for compensation and persistence: Young adult outcomes of childhood reading disability. *Biological Psychiatry*, 54, 25–33. doi:10.1016/S0006-3223(02)01836-X
- Silberglitt, B., Appleton, J. J., Burns, M. K., & Jimerson, S. R. (2006). Examining the effects of grade retention on student reading performance: A longitudinal study. *Journal of School Psychology*, 44, 255–270. doi:10.1016/j.jsp.2006.05.004
- Simmons, D. C., Coyne, M. D., Kwok, O., Harn, B. A., & Kame'enui, E. J. (2008). Indexing response to intervention: A longitudinal study of reading risk from kindergarten through third grade. *Journal of Learning Disabilities*, 41, 158–173. doi:10.1177/0022219407313587
- Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Snowling, M. J., & Hulme, C. (Eds.). (2005). *The science of reading: A handbook*. Oxford, England: Blackwell Publishing.
- Snowling, M. J., Muter, V., & Carroll, J. (2007). Children at family risk of dyslexia: A follow-up in early adolescence. *Journal of Child Psychology and Psychiatry*, 48, 609–618. doi:10.1111/j.1469-7610.2006.01725.x
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360–407. <http://dx.doi.org/> doi:10.1598/RRQ.21.4.1
- Suggate, S. P. (2010). Why what we teach depends on when: Grade and reading intervention modality moderate effect size. *Developmental Psychology*, 46, 1556–1579. doi:10.1037/a0020612
- Svensson, I., & Jacobson, C. (2006). How persistent are phonological difficulties? A longitudinal study of reading retarded children. *Dyslexia*, 12, 3–20. doi:10.1002/dys.296
- Svetaz, M. V., Ireland, M., & Blum, R. (2000). Adolescents with learning disabilities: Risk and protective factors associated with emotional well-being: Findings from the National Longitudinal Study of Adolescent Health. *Journal of Adolescent Health*, 27, 340–348 (Erratum in *Journal*

- of *Adolescent Health*, 2001, 28, 355). doi:10.1016/S1054-139X(00)00170-1
- Torgesen, J. K. (2005). Recent discoveries on remedial interventions for children with dyslexia. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 521–537). Oxford, England: Blackwell Publishing. doi:10.1002/9780470757642.ch27
- Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. K. S., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities*, 34, 33–58. doi:10.1177/002221940103400104
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency (TOWRE)*. Austin, TX: Pro-Ed.
- Undheim, A. M. (2009). A thirteen-year follow-up of young Norwegian adults with dyslexia in childhood: Reading development and educational levels. *Dyslexia*, 15, 291–303. doi:10.1002/dys.384
- Vadasy, P. F., Sanders, E. A., & Abbott, R. D. (2008). Effects of supplemental early reading intervention at 2-year follow up: Reading skill growth patterns and predictors. *Scientific Studies of Reading*, 12, 51–89. doi:10.1080/10888430701746906
- Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trial with paraeducator implementers. *Journal of Educational Psychology*, 98, 508–528. doi:10.1037/0022-0663.98.3.508
- van der Kooy-Hofland, V. A., van der Kooy, J., Bus, A. G., van IJzendoorn, M. H., & Bonsel, G. J. (2012). Differential susceptibility to early literacy intervention in children with mild perinatal adversities: Short- and long-term effects of a randomized control trial. *Journal of Educational Psychology*, 104, 337. doi:10.1037/a0026984
- Vaughn, S., & Fletcher, J. M. (2012). Response to intervention with secondary school students. *Journal of Learning Disabilities*, 45, 244–256. doi:10.1177/0022219412442157
- Vellutino, F. R., & Fletcher, J. M. (2005). Developmental dyslexia. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 362–378). Oxford, England: Blackwell Publishing. doi:10.1002/9780470757642.ch19
- Verhoeven, L., & van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*, 22, 407–423. doi:10.1002/acp.1414
- Wadsworth, S. J., DeFries, J. C., Olson, R. K., & Willcutt, E. G. (2007). Colorado longitudinal twin study of reading disability. *Annals of Dyslexia*, 57, 139–160. doi:10.1007/s11881-007-0009-7
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third Edition (WISC-III)*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV)*. San Antonio, TX: Psychological Corporation.
- Wiederholt, J. L., & Bryant, B. R. (1992). *Gray Oral Reading Tests—Third Edition (GORT-3)*. Austin, TX: Pro-Ed.
- Wilkinson, G. S. (1993). *The Wide Range Achievement Test 3 (WRAT3)*. Wilmington, DE: Wide Range.
- Wilkinson, G. S., & Robertson, G. J. (2006). *Wide Range Achievement Test 4 (WRAT4)*. Lutz, FL: Psychological Assessment Resources.
- Williams, K. T. (2001). *Group reading assessment and diagnostic evaluation: Technical manual*. Circle Pines, MN: American Guidance Service.
- Wilson, A. M., & Lesaux, N. K. (2001). Persistence of phonological processing deficits in college dyslexics with age-appropriate reading skills. *Journal of Learning Disabilities*, 34, 394–400. doi:10.1177/002221940103400501
- Woodcock, R. W. (1987). *Woodcock Reading Mastery Tests—Revised*. Circle Pines, MN: American Guidance Service.
- Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery—Revised (WJ-R)*. Chicago, IL: Riverside.
- 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* (pp. 11–19). Chicago, IL: Riverside.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson-III Tests of Achievement*. Itasca, IL: Riverside.

Received October 7, 2012

Revision received June 3, 2013

Accepted June 10, 2013 ■