Article

Hammill Institute on Disabilities

Journal of Learning Disabilities
2015, Vol. 48(5) 546–558

DOI: 10.1177/0022219413515511
journaloflearningdisabilities.sagepub.com
SAGE

High School Students With Reading Comprehension Difficulties: Results of a Randomized Control Trial of a Two-Year Reading Intervention

Sharon Vaughn, PhD[1], Greg Roberts, PhD[1], Jade Wexler, PhD[2], Michael G. Vaughn, PhD[3], Anna-Mária Fall, PhD[1], and Jennifer B. Schnakenberg, PhD[1]

Abstract

A 2-year, randomized control trial with 9th to 10th grade students with significant reading problems was provided for 50 minutes a day in small groups. Comparison students were provided an elective class and treatment students the reading intervention. Students were identified as demonstrating reading difficulties through failure on their state accountability test and were randomly assigned to one of three treatment conditions and a business as usual (BAU) condition: reading without dropout prevention, reading with dropout prevention, dropout prevention without reading, or a BAU condition. Findings from the 2-year reading intervention (reading with and without dropout prevention combined and BAU) are reported in this article. Students in reading treatment compared to students in BAU demonstrated significant gains on reading comprehension (effect size = .43), and improved reading was associated with better grades in social studies. Findings from this study provide a rationale for further implementation and investigation of intensive intervention for high school students with reading difficulties.

Keywords

high school, intensive reading, comprehension, content acquisition, intervention

A number of experimental and quasi-experimental studies have been conducted with struggling readers primarily in the elementary grades, increasingly in middle grades, and considerably fewer with high school students with reading difficulties over the past decade (see the following syntheses: Edmonds et al., 2009; Scammacca et al., 2007; Wanzek & Vaughn, 2007; Wanzek, Wexler, Vaughn, & Ciullo, 2010). Reading interventions are designed primarily to remedy problems in word reading and comprehension for students with reading difficulties or disabilities. Treatment effects on reading outcomes from these interventions at the secondary level are typically smaller across all reading domains (e.g., word reading, comprehension, and fluency) compared with effects with primary-grade struggling students (Wanzek & Vaughn, 2007; Wanzek et al., 2010). For example, Wanzek et al. (2010) reported a mean effect size (MES) of 0.20 for word reading as the largest effect for extensive intervention (22 effects) targeting students in grades 4 through 12 with reading difficulties; for studies with students with reading difficulties in grades K-3, the MES for word reading from extensive interventions was .56 (53 effects). Reading comprehension outcomes for students in grades 4 through 12 yielded a small MES of 0.09 (37 effects), whereas for students in kindergarten through third grade, the MES was three times as large (0.46; 25 effects). Although there are numerous explanations for these differences, nevertheless they suggest that for students with reading difficulties in older grades, interventions that are more intensive and occur for more than one academic year need to be investigated. However, to date, there has been little experimental research with high school students with significant reading difficulties, and we are unaware of large-scale trials of multiyear interventions that target basic

[1]The University of Texas at Austin, USA
[2]University of Maryland, College Park, USA
[3]Saint Louis University, MO, USA

Corresponding Author:
Sharon Vaughn, PhD, H.E. Hartfelder/Southland Corp. Regents Chair, Executive Director, The Meadows Center for Preventing Educational Risk, Professor, Department of Special Education, The University of Texas at Austin, 1912 Speedway, D4900, Austin, TX 78712-1284, USA.
Email: srvaughn@austin.utexas.edu

reading outcomes or the effects of improved reading on student performance in high school content area classrooms.

Reading at the Secondary Level

With the implementation of the Common Core State Standards (CCSS) students will be asked to read increasingly difficult texts to build domain-specific knowledge (i.e., acquire content knowledge) and to develop and defend diverse perspectives as a means of becoming content area literate (www.corestandards.org). High-achieving secondary school students are expected to thrive in response to the more challenging standards, whereas students with reading difficulties will no doubt experience considerable challenges.

The distinction between learning to read and building content area literacy is a hallmark of the CCSS. Of course, the two are highly interdependent; students who struggle with the foundation skills necessary for learning to read (e.g., word reading) will be ill-equipped for acquiring more sophisticated knowledge structures or developing domain-specific reading. We are interested in the group of students who continue to struggle with reading in middle school and high school and their prospects for being successful in content area classes when provided with long-term intensive reading intervention that is provided within the texts and topics of their social studies (e.g., world history) and science (e.g., biology) content areas. As a working definition for *struggling*, we use more than 3 grade levels below the student's actual grade and/or performance below the 25th percentile on a standardized measure of reading normed on probability-based samples.

Research with older struggling readers has primarily focused on students in fourth through eighth grades, with few studies with high school students (Kamil et al, 2008). Older students who struggle to read grade-level text benefit from explicit and systematic intervention organized around their instructional needs (Edmonds et al., 2009), including support for morphological awareness (Nagy, Berninger, & Abbott, 2006) and opportunities to learn to read and understand complex, multisyllabic words (Bhattacharya & Ehri, 2004). They need to understand the meanings of challenging words and be able to derive meaning for unfamiliar words encountered across multiple text types by applying knowledge of wood roots and affixes (Baumann, Font, Edwards, & Boland, 2005). For older struggling readers and for English learners, word-level instruction delivered as part of a comprehensive approach to teaching vocabulary has been associated with improved outcomes (Kieffer & Lesaux, 2008). Oral reading fluency represents a useful indicator of student automaticity at the word level and is necessary, though not sufficient, for reading comprehension (RAND Reading Study Group, 2002). Older readers who are fluent but nonetheless struggle with comprehension may benefit from comprehension practices such as monitoring, summarization, and question generation, although strategy-related effects may be conditional on more substantive cognitive structures (Willingham, 2007) or on students' developmental status (Cantrell, Almasi, Carter, Rintamaa, & Madden, 2010).

Experimental studies that scale these effective practices (or subsets of these practices) have been conducted largely in middle school settings and typically for 100 sessions or less (about 20 instructional weeks). The treatment effects have been small. For example, Chamberlain, Daniels, Madden, and Slavin (2007) found no statistically significant differences on the Gates-MacGinitie comprehension subtest in a sample of 405 sixth graders after a yearlong, randomized implementation of Reading Edge, a comprehensive, school-wide literacy model developed for Title I middle schools and aligned with the Success for All program. Kim, Samson, Fitzgerald, and Hartry (2010) found no differences in the posttest reading comprehension of struggling readers participating in an afterschool program that provided 92 hours (60 minutes per day, 4 days per week for 23 weeks) of evidence-based treatment to a randomized sample of fourth through sixth graders. An evaluation of Enhanced Reading Opportunities found no statistically significant differences between two treatment groups and a business-as-usual (BAU) condition on oral language and vocabulary-related outcomes and very small differences on a measure of reading comprehension (Somers et al., 2010). Lang et al. (2009) provided 90 minutes of daily intensive reading instruction to older struggling readers during a 9-month period and found that low-performing readers made no statistically significant gains in reading comprehension. Finally, small to moderate effects were reported on reading comprehension in a sample of sixth to eighth graders provided with daily intensive instruction during an entire school year (Vaughn & Fletcher, 2012).

Older students who have struggled for many years to read may require more than 9 months of intervention to be successful with grade-level text. To evaluate the potential benefits of extended intervention with low reading middle grade students, Vaughn, Klingner et al. (2011) examined the effect of a 3-year reading intervention conducted within a response to intervention framework. Sixth grade students with reading difficulties to a response-based, tiered reading intervention or BAU were randomly assigned to treatment or comparison conditions and maintained initial treatment status across the 3-year study. Response to intervention was reevaluated prior to each school year, with adequate responders returning to core instructional settings and inadequate or low responders continuing daily intervention in subsequent years. Treatment effect was based on growth among the group originally assigned to treatment compared to growth in the group assigned to BAU. When this group of persistently low readers (i.e., the subgroup of treatment

students who received all 3 years of the intensive reading program) was compared to similar students in the BAU on the Gates-MacGinitie Reading Comprehension measure (MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2000), the effects (Hedges g = 1.20) were large (Vaughn, Wexler et al., 2011). Treated students made expected (or normative) progress over the 3 years compared to the BAU students, who lost ground not only to the treatment students but also to normative reading trends. In other words, the effect was a result not of accelerating students' achievement but of minimizing their decline relative to the achievement of same-aged students. Thus, intensive reading treatment provided over extended periods may help the lowest performing students by "keeping their heads above water" at least in terms of reading comprehension.

The findings (Vaughn, Wanzek, Murray, & Roberts, 2012) suggest that intensive intervention may be an effective means of treating secondary students with persistent reading difficulties. However, the two primary goals of our study have been previously unaddressed. First, the effects of more "extensive" interventions (more than one school year) on high school students with reading difficulties have been unexamined. Second, effects of these reading interventions aligned with the social studies and science instruction in their content area classes have not been addressed.

Content Area Literacy

Experimental research on content area reading with high school struggling readers has focused primarily on features of text rather than the impact of instructional practices. Bergerud, Lovitt, and Horton (1988) found that adapting content area textbooks using a graphics-based approach improved content acquisition in life science classes among high school students with learning disabilities. A generic study guide condition and a self-study condition were significantly less efficacious. Horton, Lovitt, and Christensen (1991) extended this work with struggling older students by manipulating the amount of referential support provided in study guides that often accompany science and social studies textbooks. They adapted three study guides from a ninth grade science textbook and a ninth grade social studies textbook, differing on levels of cuing; the highest level presented the study questions, the medium level presented the questions along with the page and paragraph numbers necessary to locate relevant details, and the lowest level presented questions, page numbers, as well as several salient details for composing a response. The guides were used in three instructional settings; teacher-directed study, dyadic study, and independent study. The multilevel guides were more effective than single-level guides for both social studies and science textbooks when dealing with factual questions and when implemented in the teacher-directed instructional setting. On interpretive (or inferential) question types, the multilevel guides were more efficacious than the single-level guides when used with social studies textbook, and they were comparably effective across the three different instructional settings. In science, the type of study guide and the instructional setting were not related to performance on interpretive questions. Computerized approaches to implementing study guides, more so than traditional note-taking methods, also appear to significantly benefit struggling readers' comprehension of a ninth grade world geography textbook (Horton, Lovitt, Givens, & Nelson, 1989), although differences in today's technology compared to that available in 1989 may compromise the utility of this finding.

Design and Research Questions

The CCSS provide high goals for comprehension within content area texts for all students, including those with reading difficulties. Ideally reading interventions would be unnecessary after elementary grades, but increasing numbers of students require reading interventions into adolescents and adulthood (Miller, McCardle, & Hernandez, 2010). Interventions at the secondary level will succeed to the extent that they improve struggling readers' understanding of increasingly difficult text and support their development of content area literacy. At present, there is an absence of research on student-level instructional interventions that target improved reading comprehension within content area texts. This 2-year randomized control trial was designed to address this need. Although the overall study will follow all students through 12th grade, we report findings at the end of 10th grade, when the 2-year reading treatment was terminated. The research questions included the following: (1) To what extent does reading intervention aligned with the content instruction in social studies and science improve students' foundation reading skills, including reading comprehension? (2) To what extent does reading intervention aligned with the content instruction improve students' performance in content area coursework? And (3) to what extent does improved reading influence improved performance in social studies and science content area coursework?

Methods

Findings from the 2-year reading intervention randomized control trial are reported in this article comparing students treated for reading with and without dropout prevention supports to a BAU condition. The dropout prevention supports were a modified version of Check & Connect (Anderson, Christenson, Sinclair, & Lehr, 2004) and are described by Roberts et al. (in review). Students were randomly assigned to a condition within schools. We provided intervention to students during their 9th and 10th grade year

(2010-2012). Follow-up measures of cognitive and behavioral outcomes will be collected during students' 11th and 12th grade years and reported in a subsequent article.

Participants

School sites. Three diverse high schools in a large urban southwestern U.S. district participated in the study, with approximately a third of the sample from each site. In the sampled schools, approximately 43.11% of students are Hispanic, 25.51% are White, 19.44% are African American, 7.85% are Asian, and 4.06% are Native American or biracial. Additionally, 42.6% of students in participating schools are economically disadvantaged. Based on the state's evaluation system, the schools were rated as Academically Acceptable for the 2011-2012 school year.

Participant selection. All eighth grade students from the five middle schools that feed into the three participating high schools were screened using seventh grade scores from the state accountability reading test (Texas Assessment of Knowledge and Skills [TAKS]; Texas Education Agency [TEA], 2004). The district requested that randomization and class scheduling occur during the spring semester (2009) prior to beginning the intervention in the fall semester 2010. The seventh grade TAKS scores were the most current data available at that point in time. The TAKS test, administered by school personnel, is a reliable and valid measure to screen for reading difficulties, as indicated by previous studies, and has been used in previous studies to identify students with reading comprehension difficulties (Vaughn & Fletcher, 2012). The failing score is 2100 or below, and the standard error of measurement is 100. Students qualified for the study if they scored 2200 or below on the first TAKS administration. We included students in the 2100-2200 score range because their observed scores were within the lower bound of the 95% confidence interval, meaning that their true score could be below 2100. Additionally, those students with missing TAKS scores on the reading portion of the test qualified for the study if they made a failing grade in a core class (English language arts, social studies, science, and math) during seventh grade or during the first semester of eighth grade and if they met at least one of the risk indicators for dropping out of school (behavior, attendance, retention, tardies). Students who qualified were rank-ordered, with more struggling readers (i.e., students with the highest failing TAKS scores) placed higher on the list. Based on power analysis, the level of resources necessary for successful implementation, and expected attrition (.10 per year), we needed 456 students, with 152 from each participating school.

Table 1. Student Demographics.

	Overall (*N* = 375)		Treatment (*N* = 170)		Control (*N* = 205)	
	N	%	*n*	%	*n*	%
Gender						
Male	228	60.8	105	61.8	123	60
Female	147	39.2	65	38.2	82	40
Race						
White	66	17.6	32	18.8	34	16.6
Hispanic/Latino, any race	166	44.3	74	43.5	92	44.9
African American	124	33.1	58	34.1	66	32.2
Asian	19	5.1	6	3.5	13	6.3
English as a second language						
Yes	70	18.7	30	17.6	40	19.6
No	304	81.3	140	82.4	164	80.4
Student with special needs						
Yes	66	17.6	34	20	32	15.7
No	308	82.4	136	80	172	84.3

The sample at the beginning of the intervention included 375 students (Table 1). The majority were male (60.8%, n = 228) and Hispanic (44.3%, n = 166). An additional 17.6% (n =66) were White, 33.1% (n =124) African American, and 5.1% (n = 19) Asian. The number of students classified as English Language Learners (ELLs) was 18.7% (n = 70), and 17.6% (n = 66) of the students qualified for special education services.

Attrition analyses. The 457 eighth graders randomized in the spring of 2010 represented the initial sample. As a result of student mobility over the summer, requested schedule changes, and student/parent refusal to participate in the study, 375 (82.1%) students remained in the study at the beginning of the intervention, with additional attrition across the subsequent three assessment waves. At the end of ninth grade, 344 (75.3%) students remained; at the beginning of 10th grade, 325 (71.1%) remained; and at the end of tenth grade, 306 (67%) remained with complete data on all measures.

To establish whether differential attrition is evident across the groups, a two-way analysis of variance was conducted on baseline measures (Cook, Campbell, & Day, 1979) of our primary outcome variables (Gates-MacGinitie reading comprehension, course grades in social studies and science). The factors in the analysis were treatment condition, completer status at the end of Year 2, and the interaction of condition and completer status. A significant main effect for the group of completers indicates significant overall attrition and is a potential threat to the result's external validity. A significant interaction signifies systematic group differences in the characteristics of students who remained in the study and is a potential threat to the internal validity. Data revealed no significant main effects for completer status (p values from .26 to .91) and no significant Condition ×

Completer Status interaction effect (p values from .22 to .44). These findings indicate that attrition among groups was unlikely to influence the observed effects of the intervention.

Reading Intervention

Students in the reading intervention participated in classes of no more than 10 students during their elective period. Intervention was provided daily by a teacher hired and trained by the research personnel. Students participated in 50-minute classes daily over approximately 160 sessions per school year, about 320 sessions over the 2-year treatment. The treatment protocol focused on four areas: word study, vocabulary in content text, comprehension in content texts, and engagement (Boardman et al., 2008; Edmonds et al., 2009; Kamil et al., 2008; Scammacca et al., 2007; Torgesen et al., 2007). The instruction was divided into two phases. In Phase I, during Semester 1 of Year 1, students were explicitly taught an advanced word study strategy through the REWARDS Plus program (Archer, Gleason, & Vachon, 2005). They learned to identify affixes, vowel–vowel combinations, and their corresponding sounds, for the purpose of segmenting multisyllabic words into decodable chunks. Six to eight vocabulary words were identified weekly from the expository textbooks used in students' science and social studies classes and introduced using an explicit, evidence-based, six-step process (Vaughn Gross Center for Reading and Language Arts at The University of Texas at Austin, 2010). Following explicit introduction, word knowledge was reinforced through daily deep processing activities and exposure in text. During Phase I, students were also introduced to a multistep comprehension strategy based on *Collaborative Strategic Reading* (Vaughn, Klingner et al., 2011; Vaughn, Wexler et al., 2011), providing explicit modeling and scaffolded support for each strategy, time for individual practice of each strategy, and support in applying and synthesizing the collective strategies while reading texts related to the content area instruction in social studies and science.

Phase II began in the second semester of project Year 1 (ongoing support was provided to students requiring additional word-level work or practice with the comprehension strategies). Instruction was organized into 6- to 8-day instructional units, focusing on texts from students' content area classes. Within each unit, word study strategies were applied to unfamiliar and multisyllabic words and students were introduced to additional affixes according to the target passage and the impact of each affix on parts of speech and on word meaning. Vocabulary was explicitly taught and reinforced with daily deep processing activities. Application of the multistep comprehension strategy was managed within each unit using the standardized six-step summary writing strategy taught to students in Phase I (Brown & Day, 1983; Cordero-Ponce, 2000; Klingner, Morrison, & Eppolito, 2011), and each instructional unit contained an explicit discussion task and protocol aimed at deepening students understanding of the text's meaning through critical analysis and problem solving (Kamil et al., 2008).

Student engagement was addressed in several ways. The alignment of selected passages with topics in students' social studies and science classes proved motivating for a number of students. Students had specific content learning goals for each unit (Boardman et al., 2008). Also, for each unit students were given a set of three specific relevance instructions identifying key content of the reading passage (McCrudden, Magliano, & Schraw, 2010) used by teachers during passage reading to check for student understanding. Student-developed questions were also included. Finally, 10 minutes of student free choice silent reading was incorporated into each instructional day. During this time, the reading interventionist counseled different students regarding their free choice text, and, within each unit, students were given an opportunity to share a summary of their most recently read free choice text with a selected peer.

Reading Intervention Teachers

Reading interventionists were hired, trained, and supported by research staff. All interventionists were certified teachers with experience teaching struggling readers. They participated in a week-long professional development sequence prior to program implementation that focused on the study design, general features of effective instruction, the intervention program model and related theory of change, expectations related to human subjects research, plans for ongoing coaching, and details on the process for measuring fidelity of implementation. Training accorded with principles of adult learning. Effective intervention was modeled, and participants had iterative opportunities to practice and receive constructive feedback on the match between their practice and the normative program model. Peer groups were established and pairs of participants were encouraged to review and comment on samples of their partner's instructional practice. A research staff member provided regular coaching to interventionists. Students' progress monitoring data were followed as a means of assessing teacher effectiveness. Interventionists in classes with poorly performing students were the priority for coaching. Quarterly half-day "booster" sessions were provided to interventionists, and an abbreviated training (2.5 days) was provided prior to beginning Year 2 of the project.

Implementation Fidelity

Fidelity data were collected throughout both years of implementation. Each interventionist was observed six times a year by a member of the research team trained on its

Table 2. Mean Fidelity Score for Observed Variables by Year.

Observed Component	Year 1	Year 2
Vocabulary[a]	3.4	3.6
Comprehension[a]	3.4	3.5
Discussion[a]	3.2	3.3
Motivation[a]	2.8	3.6
Student engagement[a]	3.4	3.8
Global classroom management[b]	5.7	6.7
Global classroom instruction[b]	6	5.5
Average total fidelity[b]	84%	90%

[a]Scores rated on 5-point scale, with 5 as highest score. [b]Scores rated on 7-point scale, with 7 as highest score.

elements and on features of its effective implementation. Ratings were calibrated twice annually during a joint observation with the project director; interrater reliability on these occasions was greater than 90%. The fidelity observation tool was based on the program's theory of change and was designed to capture instructional elements thought to be necessary to its effect. We included motivation as part of the fidelity routine, given its importance for engaging adolescent students in meaningful discussion and interpretation of text (Kamil et al., 2008). A measure of student engagement was administered at 15-minute intervals during observed intervention sessions. An overall quality of interventionists' classroom management and instruction was assigned at each observation's conclusion. The Likert-type items on the fidelity protocol were scored from 1 to 4, with 4 representing a component that was implemented with high fidelity and 1 indicating a component that was not implemented. Student engagement was measured on a Likert-type scale ranging from 4 (90% or more engaged) to 1 (<12% engaged). The global quality of classroom management and instruction were measured using 7-point scales, with 6 and 7 indicating high quality and 1 and 2 indicating low quality. Fidelity was calculated as the proportion of assigned points to total possible points. Interventionists' average fidelity ranged from 78% to 87% in Year 1 and 84% to 92% in Year 2. Mean scores for each measured component are presented in Table 2.

Reading intervention was provided during students' elective period (e.g., fine arts, speech, computer skills, and technical applications), so students not enrolled in the reading intervention (i.e., the comparison group) did not participate in a school-provided reading intervention.

Measures

Gates-MacGinitie Reading Tests (MacGinitie et al., 2000). The Gates-MacGinitie is a group-administered, norm-referenced reading test for Grade K through adult. We administered the reading comprehension subtest. Students are provided with expository and narrative reading passages followed by multiple-choice questions. Questions address facts, inferencing, and drawing conclusions. Internal consistency reliability ranges from .91 to .93, and alternate form reliability is reported as .80 to .87. The Gates-MacGinitie Reading Test is a well-normed test that is frequently used in research studies with secondary students. We selected the measure because of its high normative structure, the fact that it uses both information and narrative texts, and the fact that it has been used so frequently in other studies allowing for comparability of findings from this study to previous studies.

Course grades in social studies (history) and science. Grades in content area classrooms were assigned by classroom teachers who were not familiar with which students participated in the reading treatment or in the comparison condition. Grades were provided to the research team by the schools. Ninth-grade content areas include U.S. History and Introduction to Biology. In 10th grade, content area requirements are World History and Introduction to Chemistry. We used end of course grades for the fall and spring semesters of 9th and 10th grades to model student performance over time.

Plan for Analysis

We used latent variable growth models (LVGMs) to address Research Questions 1 and 2. LVGMs maximize the advantages of structural equation modeling in a longitudinal context by explicitly modeling measurement error in observed variables and by estimating measurement variance across time points. They yield more precise trend estimates (Wu, Li, & Zumbo, 2007), provide a more rigorous basis for validity claims about the structure of a given data set (Valentine & McHugh, 2007), and offer a framework for evaluating model characteristics across time. We estimated reading trends over the 2-year project using Gates-MacGinnitie extended scale scores, which use a continuous, equal-interval metric to estimate status across the test's entire range (K-12), allowing within-student and between-group comparisons across multiple years. Performance in content area classrooms was modeled using class grades from the fall and spring semesters of 9th and 10th grades. The randomized design in this study minimizes bias introduced by the measurement error typical of class grades, at least for purposes of contrasting across conditions (i.e., course grades are comparably imprecise in treatment and comparison). Also, measurement considerations aside, performance on class assignments has high social validity. We fit trends from students' history and science classes. Our expectation was that students in the treatment condition would demonstrate more promising trends in reading

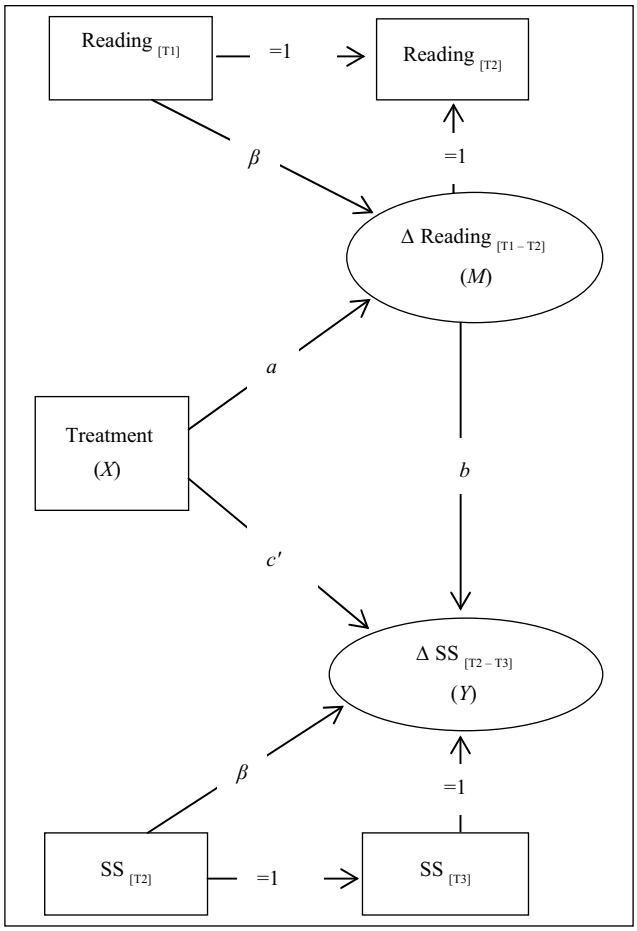

Figure 1. Diagram of the LCS mediation model.

comprehension and in their content area coursework across the 2 years compared to untreated students.

To address questions about participation in reading intervention and its role in improving content area performance (i.e., Question 3), we used latent change score (LCS) mediation analysis (MacKinnon, 2008; McArdle, 2001; Selig & Preacher, 2009). LCS models change across two points in time (rather than the 3 or more data points required by LVGM). We modeled the impact of changes in reading performance in 9th grade (i.e., Gates scores at Time 1 and Time 2) on improved grades in content area coursework (i.e., Time 2 and Time 3 teacher-assigned course grades). The LCS model tested in this article is presented in Figure 1. The diagram represents reading performance measured at two occasions—the fall and the spring of ninth grade. Grades for social studies are indicated for the spring of ninth grade and the spring of 10th grade (science grades are not indicated because there was no direct effect for reading intervention on performance in science, as described in the Results section). For ease of presentation, we label occasions as T1 (Time 1 in fall of ninth grade), T2 (Time 2 in spring of 10th grade), and T3 (Time 3 in the spring of 10th grade). Treatment condition was the independent variable (X), and change (ΔY) was the dependent variable. Improved reading performance was the hypothesized mediator (ΔM), with *mediation* defined as a statistically significant indirect effect ($a*b$) from treatment (X) to changed course grades (Y) via improved reading (M).

We used the M*plus* version 7.0 (Muthén & Muthén, 2007) software package to estimate model parameters. M*plus* provides bias-corrected bootstrap confidence intervals for mediation analysis. Confidence intervals are based on 10,000 bootstrap resamples (Preacher & Hayes, 2008). Support for a mediating role is indicated if the bootstrap (bias-corrected, or BC) confidence interval does not include zero, indicating with 95% probability that the indirect or mediating effect is significant.

Fit of individual models was evaluated by using the chi-square (χ^2) statistic, the comparative fit index (CFI), and the root mean square error of approximation (RMSEA). We followed the Hu and Bentler (1999) guidelines for evaluating the fit between the target model and the observed data: (1) RMSEA values less than 0.05 indicate excellent fit, and values in the vicinity of 0.08 indicate acceptable fit; (2) CFI and Tucker Lewis Index (TLI) values of .95 or greater indicate an excellent fit, and coefficients of 0.90 indicate a good fit. Missing data were treated as missing at random, and a full information maximum likelihood estimator was used to fit models. Parameter differences were evaluated by constraining growth-related values (means and variance estimates for intercept and slope) as equal across groups and comparing the relative fit (in terms of $\Delta\chi^2$) of nested models (Bovaird, 2007). Treatment effects were estimated in two ways: (1) as differences between achievement of students in the comparison and treatment conditions (Feingold, 2009), with achievement represented by the slope of the average trend line over the 2-year trial, and (2) as differences in intercepts, with intercept estimated as the terminal endpoint of the trend. Option 2 was preferred when there were no statistical differences prior to treatment. Option 1 was used to model and compare trends that differed in status at Time 1.

Results

Means and standard deviations for the Gates-MacGinnitie Reading Comprehension Test (standard scores and extended scale scores) and for course grades in history and science are summarized in Table 3 for the sample available at each time point. Differing sample sizes across measurement occasions (as indicated in Figure 1) reflect attrition. We estimated intraclass correlations for observed variables. Clustering was minimal (values ranged from 0% on the Time 2 Gates to .014% for Time 4 Gates), and we elected to fit LVGMs as single-level rather than multilevel models. We also examined the distribution of course grades prior to

Table 3. Extended Scale Scores, Standard Scores, and Standard Deviations for Reading Outcome (Gates-MacGinnitie).

	Fall, 9th Grade			Spring, 9th Grade			Fall, 10th Grade			Spring, 10th Grade		
	M	*SD*	*N*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
Standard score												
Reading comprehension												
Comparison	91.62	11.13	204	88.62	11.75	190	88.20	9.60	181	92.47	9.45	166
Treatment	88.41	12.46	171	89.49	11.78	154	87.58	9.56	144	93.53	8.43	139
Extended scale score												
Reading comprehension												
Comparison	522.78	29.69	204	520.59	29.82	190	528.68	20.93	181	538.82	21.69	166
Treatment	514.87	32.02	171	521.44	33.62	154	527.19	21.75	144	541.14	20.00	139

analysis to ensure univariate normality. Skewness and kurtosis of the distributions were within acceptable bounds. We also confirmed that the effect of dropout prevention did not differ in the sample of students assigned to reading treatment (R + DO) and the sample not assigned to reading treatment (DO only). This supported our use of the R + DO and R only groups to estimate treatment effects of reading intervention and the use of DO only and BAU as the comparison.

Growth Trends in Overall Reading

An unconditional, single-group linear growth model was fit to clarify the basic structure of the data and to establish a baseline for comparing conditional models. With several modifications, this model fit the data reasonably well (χ^2 = 7.27, *df* = 2, *p* = .03; CFI = .99, TLI = .98, RMSEA = .083, RMSEA 90% CI = [.024 to .152]). For purposes of model identification, we allowed Gates scores at Time 1 to correlate with Time 3 scores, Gates scores at Time 2 to correlate with Time 4 scores, and the intercept for the observed Time 4 Gates score to freely estimate (it is 0 by default). The model-estimated intercept (estimated as Time 1) was 519.18, expressed as an extended scale score. This is comparable to a 89 standard score, with a mean of 100 and standard deviation of 15. The average slope per time interval was 4.01 (total change from Time 1 to Time 4 of 12.03), and the average variation around slope was 41.60. A multigroup model with the above-listed modifications also fit the data reasonably well (χ^2 = 14.30, *df* = 5, *p* = .01; CFI = .98, TLI = .96, RMSEA = .089, RMSEA 90% CI = [.041 to .161]), though the RMSEA estimate was somewhat greater than recommended. Mean intercept differed significantly across the two groups ($\Delta\chi^2$ = 5.82, Δdf = 1, *p* = .016), with the comparison group scoring higher, on average, at Time 1 than students in the treatment. Mean slope also differed significantly ($\Delta\chi^2$ = 8.20, Δdf = 1, *p* = .004), with the treatment group outperforming the comparison over time. There were no group differences in the variance estimates for intercept (*p* = .136) or for slope (*p* = .144). An effect size of .43 was associated with the difference in unadjusted slopes.

Growth Trends in Content Area Coursework

We used a similar model building process to establish trends in students' social studies and science class grades. As before, school-level clustering was minimal, and because we were unable to link individual students with specific content area teachers and because teachers differed by student across semesters, we decided to fit single-level models, rather than attempt to model the effect of teacher. The single-group model for history fit the data well (χ^2 = 12.63, *df* = 4, *p* = .01; CFI = .99, TLI = .98, RMSEA = .071, RMSEA 90% CI = [.029 to .116]). The Time 1 estimate was 77.88, which represents the average grade in U.S. history in the Fall of ninth grade in the total sample. The average slope was -.871 ($p < .001$; σ^2 = 23.17) per measurement interval, meaning a total decline of about 2.61 percentage points, on average, over the four measurement occasions.

The multigroup model fit the data reasonably well (χ^2 = 29.72, *df* = 9, *p* = .01; CFI = .97, TLI = .96, RMSEA = .103, RMSEA 90% CI = [.063 to .145]), though RMSEA was larger than expected. As before, the decision to ignore the minimal clustering may have been a factor. Several modifications were required for purposes of model identification. In the group-specific model for comparison students, the residual variance for course grades in Fall of ninth grade was constrained as 0. The model estimate was a negative value, which prevents model identification, but because its absolute value did not differ from 0, constraints were appropriate. In the treatment group model, we allowed the Fall ninth and Spring ninth grade scores to correlate and the Fall ninth grade and Fall 10th grade scores to correlate. The Time 1 estimate in the comparison group was 77.90 and variance was 86.45, and in the treatment group, the estimated Time 1 mean was 77.98 with variance of 63.45. Slopes in the two groups were -1.32 in the comparison and -.52 in the treatment group. Intercept values (estimated as Time 4, as described earlier) were 73.98 and 76.42 in the comparison and treatment groups, respectively, a difference that was statistically significant ($\Delta\chi^2$ = 4.09, Δdf = 1, *p* = .04). The effect size was .23.

The single-group model for science grades fits the data well (χ^2 = 7.90, *df* = 3, *p* = .05; CFI = .99, TLI = .99,

RMSEA = .061, RMSEA 90% CI = [.005 to .115]), with the following constraints for purposes of identification: (1) Fall of ninth and Fall of 10th allowed to freely correlate; (2) Spring of ninth and Fall of 10th freely correlated; (3) intercept for Spring of ninth allowed to freely estimate; and (4) residual variance for Fall of ninth constrained to 0 (the estimated value was negative and did not differ statistically from 0). This same model fit the multigroup case comparably well (χ^2 = 16.05, *df* = 7, *p* = .03; CFI = .99, TLI = .98, RMSEA = .077, RMSEA 90% CI = [.026 to .128]). Time 1 mean estimates (74.49 and 74.86 in comparison and treatment, respectively) did not differ statistically (*p* = .67). Average slope in the comparison was -.622 (*p* = .04), with a variance of 11.79 (*p* < .001). In the treatment group, the estimated average slope was -.433 (*p* = .16), and the slope variance was 10.14 (*p* < .001). There were no significant differences in the estimated Time 4 scores for the two groups ($\Delta\chi^2$ = .58, Δdf = 1, *p* = .45).

Mediating Effects of Reading Comprehension on Content Area Coursework Performance

As outlined earlier, mediation effects were evaluated using LCS modeling, where change in reading performance was conceptualized as:

$$\Delta\ reading_{[T1-T2]} = \beta_{[T1]} reading_{[T1]} \quad (1)$$

where $\Delta\ reading_{[T1\text{-}T2]}$ is a latent factor that represents the score difference at T1 and T2 and $\beta_{[T1]}$ is the effect of reading at T1 on the change in reading performance. Reading performance at T2 is expressed as a function of reading at T1 plus the latently expressed change in reading performance (Equation 1):

$$reading_{[T2]} = reading_{[T1]} + \Delta\ reading_{[T1-T2]} \quad (2)$$

In a similar manner, the change in grades for social studies from T2 to T3 ($\Delta\ social\ studies_{[T1\text{-}T2]}$) and average grades for social studies at T3 were represented as follows:

$$\Delta\ social\ studies_{[T2-T3]} = \beta_{[T2]} social\ studies_{[T2]} \quad (3)$$

$$social\ studies_{[T3]} = social\ studies_{[T2]} + \Delta\ social\ studies_{[T2-T3]} \quad (4)$$

Science estimates were not calculated because there were no treatment main effects on performance in science content areas. The model in Figure 1 fit the data well (χ^2 = 2.486, *df* = 3, *p* = .48; CFI = 1.00, TLI = 1.00, RMSEA = .00, RMSEA 90% CI = [.00 to .07]). The direct effect of intervention on changes in reading performance (*a* = 7.02; BC 95% CI = 2.34, 11.88) was statistically significant. The coefficient for the path from changes in reading performance to changes in grades for social studies was also significant (*b* = .06; BC 95% CI = .01 to .12), and the indirect effect of treatment on social studies grades via improved reading differed significantly from 0 (*c'* = .44; BC 95% CI = .08, 1.13), supporting the idea that improved grades in social studies among struggling high school students may be due to improved basic reading skills.

Discussion

A 2-year reading intervention with at-risk 9th to 10th grade students was effective in improving students' reading comprehension. The treatment effect was statistically significant and represents one of the few successful interventions with high school struggling readers and the only positive experimental finding of an extensive reading intervention (more than 1 year) in a high school setting. Treatment in Year 1 was also associated with improved grades in social studies in Year 2, and the indirect effect of treatment through improved reading differed statistically from 0. There were no direct effects on students' performance in science classes.

The findings are important for several reasons. Although the 2-year effects were moderate in size, they have considerable clinical significance when compared to effect sizes typical of similarly rigorous studies of less extensive interventions. Hill, Bloom, Black, and Lipsey (2008) argue that the best guidance for the practical interpretation of effect sizes is to consider the normative expectations of growth over time, policy relevance of the findings, and how the effect sizes of the treatment compare with effect sizes from previous interventions. We interpret our findings in all three recommended areas. Using data provided by Hill et al. (2008), the average annual gain in effect sizes from nationally normed reading tests for students in primary grades (K-2) ranges from 0.97 to 1.52, whereas effects for students in Grades 9 through 11 in reading are considerably lower, 0.19. These same patterns of considerably lower effects for students in older grades were reported in two syntheses of extensive reading interventions for students in primary grades (Wanzek & Vaughn, 2007) and older grades (4 through 12) (Wanzek et al., 2010), with effects across reading interventions for students in fourth grade and older averaging 0.14. Thus, the MES of 0.43 in our study would be considered quite large relative to previous work. Another view that Hill et al. (2008) argue should be considered is the policy implications of the findings. Using their reported data about achievement gaps in reading for Black-White and Hispanic-White in eighth grade of 0.80 and 0.76, respectively, the gains in ES of our primarily minority sample of 0.43 indicates that the intervention was associated with almost a 50% reduction in the gap for minority-White students, again suggesting high practical significance. Finally, Hill et al. (2008) ask that researchers consider effects in terms of other interventions provided with the

same population. Although there are few comparisons with a similar sample, generally effects for older students are in the small range, providing a very favorable comparison for the effects from this study.

The results also indicate that improved reading due to intervention is associated with better content area outcomes (grades) in social studies but not in science. Although students across the entire sample made progressively lower grades over time, on average the downward trend of the slope for the treatment group students was less steep than that of students in the comparison condition. The downward slope in content area grades as students proceed through the grades is a common finding in samples of struggling students and in some samples of typically achieving high school students, depending on content area. Coursework becomes more challenging even within the content area classes required of all high school students for graduation, and the average downward trend in grades reflects this increased difficulty.

Implications

Although the observed improvements in reading comprehension are encouraging, treated students continued to read at levels well below average, suggesting ongoing challenges with the complex text they are likely to encounter in high school. Likewise, students' improved performance (compared to what might have otherwise been) in content area social studies classes does not represent the development of content area literacy as outlined by the CCSS. Participants in the treatment continued to score in the average to low average grade range on their social studies coursework. At the same time, the results document the effectiveness of intensive reading intervention provided over extended time periods, as a means of improving struggling students' basic reading skills and as a mechanism for improving their access to content in subject areas that are necessary for success in and graduation from high school. We do not interpret our findings as evidence that struggling readers be excluded from the raised expectations that will accompany the CCSS. However, we do think it is important to acknowledge the extent of intervention that will be necessary for a sizable minority of older students to have reasonable access to print-based content and to graduate from high school with a fairly complete complement of foundation reading skills. Our findings should also be considered when thinking about the notion of content area literacy as outlined in the CCSS. Two years of intensive reading intervention "moved the needle" for basic reading comprehension and improved content area performance when using text that is typical of existing high school curricula. CCSS recommends introduction of more difficult text and that students' use text to build and support increasingly sophisticated arguments, generally but particularly as it relates to students who struggle to read.

Limitations

Several caveats should be considered when interpreting the results. First, although many would argue that a 2-year reading intervention is extensive, considering the significance and pervasiveness of the reading difficulties, the length of the intervention may be more in line with what is necessary. Second, class grades are influenced by a variety of factors, including the completeness (or incompleteness) of classwork and homework and the correctness of completed work. We did not measure number of completed assignments or changes in patterns of completeness and correctness, but the reading intervention met daily and we have strong evidence that participants reported that they found classes motivating, both in terms of the intervention itself and in relation to school work more generally. Participants also had opportunities for less formal types of assistance, tutoring, and general support (from project interventionists) that may not have been available to students in the BAU. It is fair, as well, to acknowledge the limitations of class grades as measures of learning. We argued earlier that class grades provide unbiased estimates of treatment effects to the extent that sources of imprecision are comparably distributed across the two groups and as long as the comparisons of interest are based on group means. However, covariance-based estimates, like those in the reported mediation analysis, are more sensitive to a measure's (un)reliability, due to differences in standard error calculation, and this should be a part of any interpretation.

The findings also suggest questions about the lack of a direct treatment effect on science grades. We interpret the findings in favor of improved history grades for treated over comparison students and not in science as expected for the following reasons. First, science in high school rarely involves text reading. Thus, students in the treatment condition had little opportunity to advantage themselves of their acquiring reading comprehension. Second, because science and history content differ in their levels of accessibility (i.e., science content is more vocabulary-driven and construct-oriented while history has a more "narrative" structure), project interventionists may have been less able to support students' acquisition of biology- and chemistry-related knowledge, a difference that would, conceivably, be evident in the completeness and correctness of student work over time (i.e., treatment students did not lose ground to students in the BAU; they also did not improve relative to these same students).

Future Research

Several areas of future research are suggested. More extensive interventions should be considered. Promising work in middle school settings has been implemented over a 3-year treatment frame (Vaughn et al., 2012), yielding effect sizes on a standardized reading comprehension measure greater

than 1.00. A similar approach would be of interest in high schools. Measures of content area performance other than (or in addition to) course grades would be valuable. Many states are moving towards formal measures of content acquisition in core subject areas. Data collected from these measures may provide a reasonable estimate of posttreatment knowledge to the extent that they are made available and to the extent that the tests are "motivated." Also, the research on discipline-related differences in effective reading strategies (beyond prior knowledge effects) may have applications for struggling students (Shanahan & Shanahan, 2008). Although this work is generally organized around "expert" learning among readers with strong foundation skills, there has been no attempt to adapt and apply related evidence-based practices to a different population of learners. We are not proposing the potential for a "simple fix." Indeed, our findings suggest the contrary. However, to the extent that more challenging text will characterize high school content area classrooms, considerations for how students' access to reading for meaning can be enhanced are needed.

Conclusion

With the CCSS in 2014, expectations for secondary students, including those who struggle with reading and comprehending grade-level text, have risen. Consequently, even more emphasis has been placed on confirming effective practices to help secondary struggling learners to read, comprehend, and acquire domain-specific knowledge from complex expository content area text. Evidence from a handful of previously conducted experimental and quasi-experimental reading intervention studies with secondary struggling readers have produced only small effects. Findings from this 2-year reading treatment with 9th to 10th graders are from the first multiyear high school–level experimental study that shows evidence of effectiveness for providing intensive reading intervention aligned with content from secondary social studies and science expectations. Although we recognize the magnitude of the resources to implement interventions at the high school level, we interpret the findings from the current study to be encouraging in light of the difficulty of making an impact on reading comprehension and content acquisition with secondary struggling learners.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A100022 to The University of Texas at Austin as part of the "Preventing School Dropout With Secondary Students: The Implementation of an Individualized Reading Intervention and Dropout Prevention Intervention" grant. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

Anderson, A. R., Christenson, S. L., Sinclair, M. F., & Lehr, C. A. (2004). Check & Connect: The importance of relationships for promoting engagement with school. *Journal of School Psychology*, *42*, 95-113.

Archer, A. L., Gleason, M. M., & Vachon, V. (2005). *REWARDS Plus: Reading strategies applied to social studies passages*. Longmont, CO: Sopris West Educational Services.

Baumann, J. F., Font, G., Edwards, E. C., & Boland, E. (2005). Strategies for teaching middle-grade students to use word-part and context clues to expand reading vocabulary. *Teaching and Learning Vocabulary: Bringing Research to Practice*, *3*, 179-205.

Bergerud, D., Lovitt, T. C., & Horton, S. (1988). The effectiveness of textbook adaptations in life science for high school students with learning disabilities. *Journal of Learning Disabilities*, *21*(2), 70-76.

Bhattacharya, A., & Ehri, L. C. (2004). Graphosyllabic analysis helps adolescent struggling readers read and spell words. *Journal of Learning Disabilities*, *37*(4), 331-348.

Boardman, A. G., Roberts, G., Vaughn, S., Wexler, J., Murray, C. S., & Kosanovich, M. (2008). *Effective instruction for adolescent struggling readers: A practice brief*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Bovaird, T. (2007). Beyond engagement and participation: User and community coproduction of public services. *Public Administration Review*, *67*(5), 846-860.

Brown, A. L., & Day, J. D. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior*, *22*(1), 1-14.

Cantrell, S. C., Almasi, J. F., Carter, J. C., Rintamaa, M., & Madden, A. (2010). The impact of a strategy-based intervention on the comprehension and strategy use of struggling adolescent readers. *Journal of Educational Psychology*, *102*(2), 257.

Chamberlain, A., Daniels, C. N., Madden, N., & Slavin, R. (2007). A randomized evaluation of the Success for All Middle School reading program. *Middle Grades Research Journal*, *2*(1), 1-21.

Cook, T. D., Campbell, D. T., & Day, A. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghton Mifflin.

Cordero-Ponce, W. L. (2000). Summarization instruction: Effects on foreign language comprehension and summarization of expository texts. *Literacy Research and Instruction*, *39*(4), 329-350.

Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Tackett, K. K., & Schnakenberg, J. W. (2009). A synthesis of reading interventions and effects on reading comprehension

outcomes for older struggling readers. *Review of Educational Research*, *79*(1), 262-300.

Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods*, *14*(1), 43-53.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*(3), 172-177.

Horton, S. V., Lovitt, T. C., & Christensen, C. C. (1991). Notetaking from textbooks: Effects of a columnar format on three categories of secondary students. *Exceptionality: A Special Education Journal*, *2*(1), 19-40.

Horton, S. V., Lovitt, T. C., Givens, A., & Nelson, R. (1989). Teaching social studies to high school students with academic handicaps in a mainstreamed setting effects of a computerized study guide. *Journal of Learning Disabilities*, *22*(2), 102-107.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., & Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices: A practice guide* (NCEE #2008-4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc

Kieffer, M. J., & Lesaux, N. K. (2008). The role of derivational morphology in the reading comprehension of Spanish-speaking English language learners. *Reading and Writing*, *21*(8), 783-804.

Kim, J. S., Samson, J. F., Fitzgerald, R., & Hartry, A. (2010). A randomized experiment of a mixed-methods literacy intervention for struggling readers in grades 4–6: Effects on word reading efficiency, reading comprehension and vocabulary, and oral reading fluency. *Reading and Writing*, *23*(9), 1109-1129.

Klingner, J., Morrison, A., & Eppolito, A. (2011). Metacognition to improve reading comprehension. In R. E. O'Connor & P. F. Vadasy (Eds.), *Handbook of reading interventions* (pp. 220-253). New York: The Guilford Press.

Lang, L., Torgesen, J., Vogel, W., Chanter, C., Lefsky, E., & Petscher, Y. (2009). Exploring the relative effectiveness of reading interventions for high school students. *Journal of Research on Educational Effectiveness*, *2*(2), 149-175.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., & Hughes, K. E. (2000). *Gates-MacGinitie reading tests* (4th ed.). Itasca, IL: Riverside Publishing Company.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York: Erlbaum.

McArdle, J. J. (2001). A latent difference score approach to longitudinal dynamic structural analysis. In R. Cudeck, M. du Toit & D. Sorbom (Eds.), *Structural equation modeling: Present and future* (pp. 341-380). Lincolnwood, IL: Scientific Software International.

McCrudden, M. T., Magliano, J. P., & Schraw, G. (2010). Exploring how relevance instructions affect personal reading intentions, reading goals and text processing: A mixed methods study. *Contemporary Educational Psychology*, *35*(4), 229-241.

Miller, B., McCardle, P., & Hernandez, R. (2010). Advances and remaining challenges in adult literacy research. *Journal of Learning Disabilities*, *43*(2), 101-107.

Muthén, L. K., & Muthén, B. O. (2007). *Mplus: Statistical analysis with latent variables; User's guide* (Version 5). Los Angeles, CA: Muthén & Muthén.

Nagy, W., Berninger, V. W., & Abbott, R. D. (2006). Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle-school students. *Journal of Educational Psychology*, *98*(1), 134.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879-891.

RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.

Roberts, G., Vaughn, M., Vaughn, S., Fall, A., Schnakenberg, J., Coleman, M., & Wexler, J. (2013). School engagement: Malleability and treatment effects in a randomized control trial. Manuscript submitted for publication.

Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, C. K., & Torgesen, J. (2007). *Intervention for adolescent struggling readers: A meta-analysis with implication for practice*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Selig, J. P., & Preacher, K. J. (2009). Mediation models for longitudinal data in developmental research. *Research in Human Development*, *6*(2-3), 144-164.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, *78*(1), 40-59.

Somers, M. A., Corrin, W., Sepanik, S., Salinger, T., Levin, J., & Zmach, C. (2010). *The Enhanced Reading Opportunities Study final report: The impact of supplemental literacy courses for struggling ninth-grade readers* (NCEE 2010–4022). Retrieved from http://ies.ed.gov/ncee/pubs/20104021/pdf/20104021.pdf

Texas Education Agency. (2004). *Appendix—Technical Digest 2004-2005*. Retrieved from http://ritter.tea.state.tx.us/student.assessment/resources/techdig05/appendices.html

Torgesen, J. K., Houston, D., Rissman, L., Decker, S., Roberts, G., Vaughn, S., & Lesaux, N. K. (2007). *Academic literacy instruction for adolescents: A guiding document from the Center on Instruction*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Valentine, J. C., & McHugh, C. M. (2007). The effects of attrition on baseline comparability in randomized experiments in education: A meta-analysis. *Psychological Methods*, *12*(3), 268.

Vaughn, S., & Fletcher, J. M. (2012). Response to intervention with secondary school students with reading difficulties. *Journal of Learning Disabilities*, *45*(3), 241-253.

Vaughn Gross Center for Reading and Language Arts at The University of Texas at Austin. (2010). *Vocabulary and oral language development*. Austin, TX: Author.

Vaughn, S., Klingner, J. K., Swanson, E. A., Boardman, A. G., Roberts, G., Mohammed, S. S., & Stillman-Spisak, S. J. (2011). Efficacy of collaborative strategic reading with middle school students*. *American Educational Research Journal*, *48*(4), 938-964.

Vaughn, S., Wanzek, J., Murray, C. S., & Roberts, G. (2012). *Intensive interventions for students struggling in reading and mathematics: A practice guide*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Vaughn, S., Wexler, J., Leroux, A. J., Roberts, G., Denton, C. A., Barth, A. E., & Fletcher, J. M. (2011). Effects of intensive reading intervention for eighth grade students with persistently inadequate response to intervention. *Journal of Learning Disabilities*, *45*(6), 515-525.

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, *36*(4), 541-561.

Wanzek, J., Wexler, J., Vaughn, S., & Ciullo, S. (2010). Reading interventions for struggling readers in the upper elementary grades: A synthesis of 20 years of research. *Reading and Writing*, *23*(8), 889-912.

Willingham, D. T. (2007). Critical thinking: Why it is so hard to teach? *American Educator, Summer*, 8-19.

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multigroup confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research and Evaluation*, *12*(3), 1-26.