Effects of Individualized and Standardized Interventions on Middle School Students With Reading Disabilities

SHARON VAUGHN,
(Texas CEC), H. E. Hartfelder/Southland Corp. Regents Chair; Executive Director, The Meadows Center for Preventing Educational Risk, and Professor, Department of Special Education, The University of Texas at Austin

JADE WEXLER,
(Texas CEC), Research Assistant Professor, Department of Special Education and Dropout Institute Director, The Meadows Center for Preventing Educational Risk, The University of Texas at Austin

GREG ROBERTS,
(Texas CEC), Director, Vaughn Gross Center for Reading and Language Arts, and Associate Director, The Meadows Center for Preventing Educational Risk, The University of Texas at Austin

AMY A. BARTH,
(Texas CEC), Research Assistant Professor, Texas Institute for Measurement, Evaluation, and Statistics, University of Houston, Texas

PAUL T. CIRINO,
Research Associate Professor, Department of Psychology, University of Houston, Texas

MELISSA A. ROMAIN,
(Texas CEC), Research Assistant Professor, Department of Psychology, University of Houston, Texas

DAVID FRANCIS,
Hugh Roy and Lillie Cranz Cullen Distinguished Professor, Texas Institute for Measurement, Evaluation, and Statistics, University of Houston, Texas

JACK FLETCHER, and
(Texas CEC), Distinguished University Professor, Department of Psychology, University of Houston, Texas

CAROLYN A. DENTON
(Texas CEC), Associate Professor, Department of Pediatrics, Center for Academic and Reading Skills, University of Texas Health Science Center at Houston, Texas

Abstract

This study reports the effectiveness of a year-long, small-group, tertiary (Tier 3) intervention that examined 2 empirically derived but conceptually different treatments and a comparison condition. The researchers had randomly assigned all students to treatment or comparison conditions. The participants were seventh- and eighth-grade students from the previous year who received an intervention and did not meet exit criteria. The researchers assigned them to one of two treatments: standardized ($n = 69$) or individualized ($n = 71$) for 50 min a day, in group sizes of 5,

Correspondence concerning this article should be addressed to Sharon Vaughn, Department of Special Education, The University of Texas at Austin, 1 University Station, D4900, Austin, TX 787120365, SRVaughnUM@aol.com.

for the entire school year. Comparison students received no researcher-provided intervention ($n =$ 42). The researchers used multigroup modeling with nested comparisons to evaluate the statistical significance of Time 3 estimates. Students in both treatments outperformed the comparison students on assessments of decoding, fluency, and comprehension. Intervention type did not moderate the pattern of effects, although students in the standardized treatment had a small advantage over individualized students on word attack. This study provides a framework from which to refine further interventions for older students with reading disabilities.

The literature on multitiered, research-based reading interventions provides strong evidence for the critical role of early reading instruction and the benefits of early intervention for children who are struggling to learn to read (Denton, Fletcher, Anthony, & Francis, 2006). Although applying effective prevention programs that address early reading difficulties is fundamental to preventing further reading difficulties in at-risk children, intervention approaches for students who have already exhibited reading failure have less empirical support (Kamil et al., 2008). In particular, few experimental studies document the effects of multitiered intervention approaches in reading for students in the middle grades. To address this need, we designed a series of studies to determine the effects of these interventions on students with reading disabilities in Grades 6 through 8. This article reports findings from a study in which students who were inadequate responders to a previously provided Tier 1 intervention (enhanced instructional practices in vocabulary and comprehension) and Tier 2 intervention (supplemental daily reading instruction for at-risk students; Vaughn, Cirino, et al., 2010; Vaughn, Wanzak, et al., 2010) received one of two conceptually different tertiary interventions (Tier 3): individualized or standardized treatments. At-risk students randomly assigned to the comparison condition in Year 1 remained in that condition.

BACKGROUND ON READING INSTRUCTION FOR OLDER STUDENTS WITH READING DIFFICULTIES

A significant number of students demonstrate reading difficulties that persist into their middle and high school years. In 2007, the National Assessment of Educational Progress reported that 69% of eighth-grade students were unable to successfully derive meaning from grade-level text. With such a high prevalence of reading problems in the middle grades and an increasing focus on improving high school retention and preparing students for postsecondary learning, adolescent reading instruction has become increasingly important (Kamil et al., 2008).

Older students demonstrate a broad and complex range of difficulties related to reading. These include problems in recognizing words, understanding word meanings, and understanding and connecting with text; students often lack background knowledge required for reading comprehension (Biancarosa & Snow, 2004). We examined several syntheses on interventions for secondary students with reading difficulties to identify effective interventions to meet this range of reading difficulties. Edmonds et al. (2009) conducted a meta-analysis examining the effects of adolescent reading interventions (Grades 6 through 12) that included instruction in decoding, fluency, vocabulary, or comprehension on reading comprehension outcomes. Analyses revealed a mean weighted effect size in the moderate range in favor of treatment students over comparison students. Promising approaches were those that provided targeted reading intervention in comprehension, multiple reading components, or word-recognition strategies.

In a related meta-analytic synthesis, Scammacca et al. (2007) examined single- and multi-component interventions to determine the effect of various intervention components on reading-related outcomes, including and in addition to reading comprehension outcomes.

Most of the studies reported outcomes using nonstandardized measures, which inflated the overall effect sizes. Of the 11 studies that used only standardized measures, the mean effect size was 0.42, with lower effects for word-study interventions than for comprehension- or vocabulary-focused studies. Using only standardized outcome measures (not available for vocabulary), the impact of moderator variables shifted, with word study and comprehension strategy instruction demonstrating the highest effect sizes. The researchers found higher effect sizes associated with researcher-implemented interventions and middle school participants rather than high school participants.

There are, however, several significant differences between the studies reported in these syntheses and the current study: (a) none of the studies from the syntheses focused on students with demonstrated low response to previous interventions, whereas the study reported here included the lowest responders from a previous year-long intervention; (b) none of the synthesis studies were large-scale (e.g., multiple schools across sites) as was our study—these large-scale, multiple-site studies are consistently associated with lower effects; and (c) many studies in the syntheses used researcher-developed measures, and we used only standardized measures. Moreover, none of these studies provided a multicomponent, comprehensive approach to remediating reading difficulties for secondary students with significant and persistent reading disabilities.

INDIVIDUALIZED OR STANDARDIZED APPROACHES TO TERTIARY INTERVENTIONS

We were interested in determining the efficacy of two conceptually different but empirically derived treatment approaches to remediating reading disabilities with individuals whose response to a Tier 2 intervention was inadequate. Within a response-to-intervention (RTI) framework, researchers have highlighted two approaches to intervention: a standard protocol and a problem-solving approach (e.g., Fuchs, Mock, Morgan, & Young, 2003). A standard protocol intervention uses research-based instructional programs and is provided in a specified manner to all students with learning difficulties. Typically, a standard protocol includes a well-specified treatment furnished in a step-by-step sequence. Educators consider that standard protocols are easier for school personnel to implement because they typically (a) have teachers' guides and student materials that give instructional support; (b) furnish clear expectations allowing for ease of implementation and fidelity determination; (c) enable schools to document what educators have taught, thereby guiding decision making and placement in special education; and (d) leverage school resources more efficiently by allowing districts to focus interventions on fewer options.

Rather than implement a problem-solving approach that is grounded in school psychology and involves behavioral problem solving (Bergan, 1977) with limited empirical support for reading interventions, we chose an individualized approach with roots in special education. We derived this approach from a clinical teaching perspective that designs instruction to meet students' instructional needs and documents it through daily instructional monitoring and weekly progress monitoring. Minimal data exist about establishing the effectiveness of individualized intervention approaches (Fuchs et al., 2003), especially those related to adolescent populations (Scammacca et al., 2007). Even at the elementary school level, where research on effective reading interventions is prevalent, information about the effectiveness of individualized intervention approaches is scarce (Wanzek & Vaughn, 2007). Historically, special education provides specialized instruction to students with disabilities to accommodate individual differences (DeStefano & Snauwaert, 1989). Within an individualized approach, educators make instructional decisions and adaptations on the basis of student progress and individual variation (Deno, Fuchs, Marston, & Shin, 2001; Fuchs, Fuchs, & Compton, 2004). Although individualized interventions epitomize clinical

approaches to intervention with students with reading disabilities, few empirically based intervention studies use such techniques (Fuchs et al., 2003), particularly for older students.

PREVIOUS STUDIES AND CURRENT RESEARCH QUESTIONS

This article reports findings from the second year of a multiyear middle-school study. During the Year 1 study (2006–2007), in which we randomly assigned students to treatment and comparison conditions, we identified middle school students with reading difficulties and provided a year-long, Tier 2 intervention (more than 100 hours; Vaughn, Cirino, et al., 2010; Vaughn, Wanzek, et al., 2010) to treatment students. All students received the benefits of content-area teachers (e.g., social studies, science) who participated in researcher-provided professional development designed to integrate vocabulary and comprehension practices throughout the school day. Treatment students showed small gains on measures of decoding, fluency, and comprehension over the course of the year (median $d = +0.16$).

After the Year 1 treatment, we identified students who failed to attain benchmarks for RTI. For the subsequent school year (2007–2008), students whom we had assigned in Year 1 to treatment and comparison groups remained in those conditions, and treatment students received either individualized or standardized conditions. We hypothesized that students who participated in the individualized intervention would outperform students who participated in the standardized intervention on reading-related outcomes. We also hypothesized that both treatment conditions would yield outcomes that were statistically significantly higher than those for students in the comparison condition. We expected to see benefits for decoding, fluency, and comprehension outcomes. However, we expected that the overall effect of the intervention would be less than that suggested by the Edmonds et al. (2009) and Scammacca et al. (2007) reviews, given that the students participating in this Tier 3 intervention were significantly more impaired, as indicated by their low response to a previous year-long intervention in reading, than students who participated in most studies reported in previous syntheses.

METHOD

Participants

School Sites—The researchers conducted this study with institutional review board approval in two urban cities (one large district; one medium-sized district) in the southwestern United States, with approximately half the sample from each site. School populations ranged from 498 to 1,145 students. Seventh- and eighth-grade students from six middle schools participated in the study.

Criteria for Participation—In 2006–2007 (Year 1), we selected all struggling readers by using the state accountability test results (Texas Assessment of Knowledge and Skills [TAKS]; Texas Education Agency, 2004) to identify struggling readers with a scale score that approximates the 30th percentile (TAKS below 2150) on other norm-referenced reading comprehension assessments. We also included students exempted from the TAKS because of special education status attributable to very low reading achievement. Students randomly assigned to treatment in Year 1 received an intensive standardized protocol in groups of 10 to 15 students. Students who were adequate responders (i.e., those with TAKS scale scores at or above 2,150) exited the intervention, and we did not include them in this study. We randomly assigned inadequate responders (i.e., students who scored below 2,150) to one of two Tier 3 treatments for a second full year of intervention, either another year of standardized intervention protocol or an individualized intervention. Year 1 comparison students with TAKS scores below benchmark remained in the comparison condition for the current study.

Student Participants—The prospective design of the study required two randomizations: one in Year 1 and another in Year 2. To accommodate the Year 2 randomization, the researchers overrandomized students in Year 1 (i.e., assigned a greater number of qualifying students to the treatment condition) at three treatment students for each one comparison student. The sample for this study included a total of 182 students (86% free and reduced lunch): 42 comparison students (10% White, 31% special education, 29% limited English proficiency [LEP]); 71 individualized students (7% White, 39% special education, 21% LEP); and 69 students in the standardized condition (16% White, 35% special education, 20% LEP).

The researchers randomized students after posttesting in the spring of Year 1 (2007). We included all eligible cases (i.e., $N = 182$) in the analysis sample. Of the 182 sampled cases, 150 returned in fall 2007: 38 comparison students, 55 in the individualized condition, and 57 in the standardized protocol condition. In spring 2008, there were 36 comparison students, 51 students in the individualized condition, and 46 in the standardized protocol condition. The researchers compared pretest scores for students not continuing through spring 2008 across the three groups. There were no statistically significant differences. More than 90% of the coverage estimates (amount of data present in each cell of the measure by occasion matrix) were at or above .75 across all outcomes and measurement occasions.

Teacher Participants—The researchers hired six female intervention teachers with a median of 8.5 years of teaching experience. They received 60 hr of professional development before teaching. They also participated in biweekly staff development meetings with ongoing on-site feedback and coaching (once every 1–2 weeks).

Description of Interventions

Students in both treatment conditions received interventions during their elective periods, in classes of four to five students for 50 min daily for approximately 160 lessons during their elective periods. To ensure that effects from the treatment could not be attributed to teachers, the researchers trained all teachers systematically on both treatments and then randomized each teacher's classes to standardized or individualized conditions.

Standardized Intervention Protocol—The standardized treatment protocol reflected three phases of intervention. Phase I (about 20 to 25 lessons over 6 to 7 weeks) began with an emphasis on word study and fluency, with additional instruction on vocabulary and comprehension.

- The intervention supported *fluency* through daily repeated reading practice in a partner- reading format that paired skilled readers with less skilled readers.

- The teachers tracked student progress through regularly administered assessments of oral reading fluency.

- The researchers addressed *word study* by using REWARDS (Archer, Gleason, & Vachon, 2003), a program designed to teach advanced strategies for decoding multisyllabic words. Students received daily instruction and practice with individual letter sounds, letter combinations, and affixes and learned a segmentation strategy for decoding and spelling multisyllabic words.

- The protocol addressed *vocabulary* daily by teaching the meanings of words from text being read. Vocabulary instruction included providing student-friendly definitions along with examples and nonexamples of the proper use of new words.

- The teachers taught *text comprehension* by asking students to answer questions of varying levels of difficulty (literal and inferential) while reading a passage and after

they had finished reading it to check for understanding and model active thinking during reading. Students learned to use text as a resource for answering questions and justifying their responses.

Phase II (about seventeen to eighteen weeks) focused on vocabulary and comprehension, with additional instruction and practice on the word study and fluency skills and strategies from Phase I. After reading new vocabulary words, students engaged in practice activities, including identifying the appropriate word to match various examples or descriptions. In addition, the teachers introduced students to word relatives and parts of speech (e.g., preserve, preservation, preservable). Finally, students reviewed applying word study to spelling words. The researchers selected vocabulary words for instruction from the text used in the fluency and comprehension component.

Three days a week, teachers addressed fluency and comprehension using REWARDS Plus Science text and lesson materials (Archer et al., 2003). Teachers taught vocabulary related to the text reading and dictated spelling words, and then students previewed the passage. Teachers next guided the students in reading the passage, asking questions to check for understanding and to model active thinking during reading. While students read, they completed a graphic organizer that summarized key information. Students engaged in repeated reading activities to increase fluency and answered questions about the passage content, providing the reasons for their choices. Students also engaged in writing content summaries. Two days a week, teachers used novels with researcher-developed lessons that reflected the use of strategies learned in REWARDS Plus.

Phase III occurred over approximately eight to ten weeks. In Phase III, students continued the instructional emphasis on vocabulary and comprehension, with application of skills and strategies in expository texts. Teachers taught comprehension and critical thinking at the sentence, paragraph, and multiparagraph levels.

Individualized Intervention Protocol—Students in the individualized intervention received instruction tailored to meet their individual needs, with students' test scores dictating the reading components that made up each lesson and the amount of time allocated to the different reading components; advancement and pacing of lessons was based on individual student mastery rather than group mastery. Educators used assessment data to develop students' profiles in phonics, word reading, fluency, vocabulary, and comprehension. For example, we used diagnostic data from the Woodcock-Johnson III subtests (WJ-III; Woodcock, McGrew, & Mather, 2001) to identify students with a standard score of 95 or above on the Word Attack subtest and focused their instruction on upper-level multisyllabic word reading, as well as vocabulary and comprehension strategies. Students who scored below a 95 standard score received a more intensive focus on word-study instruction, as well as vocabulary and comprehension strategy instruction. Teachers documented these relative emphases for each student within their group and then used weekly progress monitoring to adjust the emphasis. Overall, teachers followed a similar scope and sequence of research-based comprehension strategy instruction (e.g., strategies for finding main ideas and summarizing text) for all students in the individualized condition but had access to a variety of instructional materials and could modify pacing and materials in response to students' needs. Teachers used a variety of narrative and expository texts to teach; and they scaffolded use of the strategies before, during, and after reading. Word-study instruction was also flexible; but teachers primarily used an explicit, intensive multisensory word-study program (Wilson Reading System, 1996) that targeted both reading and spelling. Teachers progressed through the program in a flexible manner, varying pacing and lesson implementation according to students' needs. In addition, a motivation component built into the daily individualized lessons included weekly expectations for

purposeful and motivational text selection, student and teacher goal setting, evaluation conferences, and positive telephone calls home.

Individualized instruction included specified guidelines for use of instructional time. Students with higher-level decoding skills received 35 to 45 min of instruction in vocabulary/morphology, 170 to 180 min in comprehension/text reading, and 15 to 25 min of the motivational component during a 5-day week. Students with below-average word reading received 100 to 110 min of word study/text reading instruction, 35 to 45 min of vocabulary/morphology instruction, 70 to 80 min of comprehension/text reading instruction, and 15 to 25 min of the motivational component. Teachers made decisions to modify instruction by relying on biweekly curriculum-based measures (CBMs). Teachers developed CBMs on the basis of instructed objectives to determine skill mastery and guide instruction, whereas they based more formal decisions regarding student progress on standardized monthly progress-monitoring checks.

Fidelity Data

The researchers conceptualized fidelity as the difference between the intended (or normative) program model and the implemented model (Chen, 1990). To document teachers' adherence to program elements and quality of implementation, trained observers collected fidelity data for each condition four to five times a year for each teacher. Teachers following specified program elements/activities (i.e., fluency, vocabulary instruction, oral blending activities) within specified time limits represented the normative model for the standardized condition. The researchers collected fidelity information by using a 3-point Likert-type rating scale ranging from 1 (*low*) to 3 (*high*) to assess the extent to which the teacher completed each required instructional program element/activity and the overall quality of implementation, which included the active engagement of the students during each instructional program element/activity. The researchers conceptualized the individualized program model to respond to the "individualized" intent of the treatment. Teachers taught particular reading components (e.g., word study/text reading; vocabulary/ morphology) for a set time on a weekly basis according to student needs. The researchers collected fidelity information by using the same 3-point Likert-type rating scale used in the standardized condition to assess compliance with each required reading component. The researchers also collected overall quality of implementation data for each specified reading component.

Individual teachers' mean implementation scores for the individualized intervention ranged from 2.13 to 3.0, with a group average of 2.61 on a 4-point scale (0 to 3). Mean quality scores for the individualized intervention ranged from 2.13 to 2.93, with a group average of 2.53. Teachers' mean total fidelity rankings, which included implementation and quality ratings, ranged from 2.13 to 2.98 for the individualized intervention, with an average of 2.58. Teachers' mean implementation scores for the standardized intervention ranged from 2.44 to 3.0, with an average of 2.72 on a 4-point scale (0 to 3). Mean quality scores for the standardized intervention ranged from 2.19 to 2.88, with an overall average of 2.55. The mean total fidelity ranking, including implementation and quality ratings, ranged from 2.31 to 2.90 for the standardized intervention, with a group average of 2.66 (to obtain copies of both the standardized and individualized fidelity protocols, see http:// www.texasldcenter.org/research/project3.asp).

Measures

For further descriptions and more reliability and validity data on these measures, see http:// www.texasldcenter.org. The researchers obtained all measures at pretest and posttest in Years 1 and 2 unless otherwise indicated.

Decoding and Spelling—We assessed word-reading accuracy for real words and pseudowords with the Letter-Word Identification and Word Attack subtests of the WJ-III Tests of Achievement (Woodcock et al., 2001). Educators administered the WJ-III Spelling subtest at posttest. Coefficient alphas based on a sample from the previous year of 327 struggling readers and 249 typical readers who contributed data throughout the year for Letter-Word Identification and Word Attack ranged from .93 to .97; coefficient alpha for Spelling at posttest was .84.

Fluency—The Sight Word Efficiency and Phonemic Decoding Efficiency subtests from the Test of Word Reading Efficiency (TOWRE; Torgesen et al., 1999) assessed word list fluency for real words and pseudowords. Internal consistency for different forms of this well-standardized test exceeds .90.

Comprehension—The Texas Assessment of Knowledge and Skills (TAKS; Texas Education Agency, 2004), a criterion-referenced reading comprehension test, is the Texas accountability test. The TAKS is not timed and uses different assessments for each grade; these criterion-referenced assessments align with grade-based standards from Texas Essential Knowledge and Skills (TEKS). The internal consistency (coefficient alpha) of the Grade 7 test is .89 (Texas Education Agency, 2004). We used it as an initial screening assessment and then as a benchmark assessment because it is reliable, represents an accountability high-stakes assessment implemented in all states, and has good construct validity as a measure of reading comprehension. In preliminary latent-variable analyses of the students in Grades 6 to 8, the TAKS measure loaded strongly on the WJ-III Passage Comprehension subtest and the Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001). The WJ-III Passage Comprehension subtest is a cloze-based assessment in which students read a passage and fill in a missing word. Coefficient alphas in the entire sample of 327 struggling readers and 249 typicals were .94 at pretest and .85 at posttest.

Statistical Analyses

Latent variable growth modeling (LGM) as a type of structural model has advantages over observed-score approaches. It provides more precise score estimates by explicitly estimating measurement error. LGM generates indexes of overall model fit, making possible the evaluation of a model's adequacy and the comparison of competing models. It handles missing data by using a direct maximum likelihood (ML) estimator to compute a likelihood function for each case using all available data; it is more efficient than listwise deletion or imputation of missing values and yields more precise estimates and greater power. Also, because LGM analyzes covariance structures representing different levels of aggregation, it is more appropriate than traditional approaches when data are nested, whether by design (i.e., stratified sampling strategy); circumstance (e.g., students in schools); or in the case of growth models, the nesting of time within students. Finally, LGM provides a flexible framework for analyzing the differential effects of covariates.

Modeling linear growth requires at least three data points; we used spring 2007, fall 2007, and spring 2008 data (the relatively brief timeframe and the results of preliminary analyses suggested a linear model). This timeframe encompassed the summer months, suggesting the possibility of learning loss from Time 1 to Time 2 and casting doubt on the usefulness of fitting the raw data to a linear model. Accordingly, we used standard scores, thereby making the linear model easier to fit and minimizing the confounding effects of summer learning loss. We modeled intercept as the end point of the trend (i.e., Time 3 or spring 2008) to accommodate the comparison of posttest differences. We modeled slope as a fixed effect for purposes of parsimony and based it on the statistically nonsignificant slope variance within

groups. "Expected" growth in this model has slope of 0, given the use of standard scores. In addition, nonzero slopes in Table 1 are in a counterintuitive direction, because the slope estimate represents movement from right to left (i.e., from Time 3 to Time 1). In this way, the researchers can evaluate both final performance and progress since the end of the previous intervention (i.e., Year 1 intervention). Intercept represents performance at the end of Year 2, whereas slope indicates the rate of change over time. We did not specifically model school-level clustering effects because of difficulties with model identification (there were more parameters than clusters). Previous work using data from these students on these measures (Vaughn, Cirino, et al., 2010; Vaughn, Wanzek, et al., 2010) found minimal clustering effects at the teacher/interventionist level, and inclusion of such effects did not substantively alter results.

The researchers used multigroup modeling with nested comparisons to evaluate the statistical significance of slope and intercept estimates (Bovaird, 2007; Mehta & Neale, 2005). The difference test involved constraining the groups as equal on parameters of interest and comparing the fit of the constrained and the fully specified models. If groups were comparable on spring 2008 performance, for instance, the fit for the constrained and full models did not significantly differ. Constraints resulting in less adequate fit suggested significant group differences. We conceptualized the main effects of treatment as differences between the full treatment group and the comparison group on mean intercept and slope. The comparison group is generally lower in spring 2007 (i.e., Time 1) than the two treatment groups, suggesting a possible confounding of spring 2008 differences with spring 2007 differences. We interpreted the absence of statistically significant Time 3 (i.e., intercept) differences as evidence of no effect. For the same reason, in cases of significant Time 3 differences, we also contrasted the groups' slope estimates as a follow- up to determine whether the rate of progress differed regardless of Time 1 status. We interpreted a statistically significant slope difference as evidence of a main treatment effect. We also considered differences between the two treatment conditions and between each treatment condition and the comparison as a test of the moderating effect of intervention type. We incorporated other moderators (special education studies, LEP) into the models, as well.

RESULTS

Primary Outcomes

Table 2 presents descriptive summaries. There were no statistically significant group differences in mean scores at pretest. Across the 18 comparisons (i.e., three comparisons each for six measures), the mean standardized difference (using the absolute value for each effect) was .07 and the average 95% confidence interval for these effects was about ±2.5 standard score points. Results for the multigroup unconditional model, as previously described, are in Table 1. The intercept in Table 1 is the model-derived score for performance in spring 2008. Variance is not indicated for slope because we estimated it as a fixed effect. We calculated effect sizes as the difference in estimated means (i.e., model-derived intercepts) divided by the weighted, pooled standard deviation for each estimate and adjusted for small sample size using Hedges's g formula. Progress on reading comprehension for the combined treatment group was statistically significantly greater than progress in the comparison group.

Decoding and Spelling—The multigroup model (i.e., three groups) for WJ-III Letter Word Identification fit the data well (χ^2 = 10.33 [9], p = .324; comparative fit index [CFI] = .99, Tucker- Lewis index [TLI] = .99; root mean square error of approximation [RMSEA] = .05). The score estimates for spring 2008 in the two treatment groups were 89.77 and 91.09 for the individualized and standardized protocol groups, respectively, and 90.48 for

the treatment groups combined. These values represent standard scores. The comparison group average was 86.34. The difference between the total treatment group (i.e., individualized and standardized protocol combined) and the comparison was not statistically significant ($\Delta\chi^2 = 3.29$, $df = 1$, $p > .05$), where $\Delta\chi^2$ represents the difference in χ^2 between the constrained and fully specified multigroup models. The performance of the individualized group did not differ significantly from the comparison ($\Delta\chi^2 = 1.68$, $df = 1$, $p > .05$), nor did the average performance of participants in the standardized protocol ($\Delta\chi^2 = 3.71$, $df = 1$, $p > .05$). Differences between the individualized and standardized protocol groups were not statistically significant ($\Delta\chi^2 = .40$, $df = 1$, $p > .05$). The effect sizes on WJ Letter Word Identification were .28 and .44 for the individualized protocol and the standardized protocol, respectively. Readers should consider these effect estimates in light of the previously discussed Time 1 differences in comparison and treatment groups. The effect size for the difference between the individualized and standardized groups was −.11, favoring the standardized group (i.e., the standardized protocol group outscored the individualized group by about .10 standard deviations).

The model for WJ-III Word Attack also fit the data well ($\chi^2 = 4.95$ [9], $p = .838$; CFI = .99, TLI = .99; RMSEA = .001). The estimated Time 3 standard score for the comparison group was 90.25, somewhat lower than the average score for the individualized group ($\gamma_{ij} = 91.78$) and for the standardized protocol ($\gamma_{ij} = 94.64$), although not statistically significantly so ($\Delta\chi^2 = .42$, $df = 1$, $p > .05$ for individualized and $\Delta\chi^2 = 3.78$, $df = 1$, $p > .05$ for standardized protocol). The effect sizes were .14 and .45 for the individualized and standardized protocol conditions, respectively. There were no significant differences between the individualized and standardized protocols ($\Delta\chi^2 = 1.93$, $df = 1$, $p > .05$; Hedges's g unbiased = −.27).

Like the models for Word Attack and Letter Word Identification, the fit for WJ-III Spelling was excellent ($\chi^2 = 6.36$ [9], $p = .703$; CFI = .99, TLI = .99; RMSEA = .001). The estimated average scores for the two treatment groups ($\gamma_{ij} = 88.11$ and $\gamma_{ij} = 88.86$ for individualized and standardized protocol) were higher than the comparison group estimate ($\gamma_{ij} = 84.07$), although the differences were not statistically significant ($\Delta\chi^2 = 1.74$, $df = 1$, $p > .05$ and $\Delta\chi^2 = 2.42$, $df = 1$, $p > .05$). The effect size for the individualized condition was .31. For the standardized protocol, the standardized mean difference was .36. The standardized difference size between the standardized protocol and the individualized group was −.06, again favoring the standardized group.

Fluency—Fit was marginal for TOWRE—Sight Word Efficiency ($\chi^2 = 21.62$ [9], p = .010; CFI = .97, TLI = .97; RMSEA = .154, confidence interval = .071 – .238), given the high RMSEA value. Estimated Time 3 scores were 90.74 for the individualized group, 90.57 for the standardized protocol, and 87.91 for the comparison condition. Group differences were not statistically significant ($\Delta\chi^2 = 1.32$, $df = 1$, $p > .05$ and $\Delta\chi^2 = 1.28$, $df = 1$, $p > .05$ for the individualized and standardized protocols, respectively); and effect sizes for individualized (.24) and standardized protocol (.27) were in the small to moderate range. The model for TOWRE—Phonemic Decoding represented a good fit with the data ($\Delta\chi^2 = 11.48$ [9], $p = .245$; CFI = .99, TLI = .99; RMSEA = .068). The treatment conditions scored comparably at posttest ($\gamma_{ij} = 91.91$ and $\gamma_{ij} = 91.70$; Hedges's g unbiased = .01). Both outscored the comparison (87.81), on average, although the difference was not statistically significant ($\Delta\chi^2 = 1.38$, $df = 1$, $p > .05$ and $\Delta\chi^2 = 1.26$, $df = 1$, $p > .05$). The effect sizes were .26 and .27 for the individualized and standardized protocols, respectively.

Comprehension—The model for the WJ-III Passage Comprehension measure fit the data well ($\chi^2 = 13.67$ [9], $p = .135$; CFI = .99, TLI = .99; RMSEA = .094, .90 confidence interval = .00 – .19), although the RMSEA value was higher than is typically desirable. The

treatment groups collectively (i.e., a two-group comparison of treatment and comparison groups) outperformed the comparison in spring 2008 ($\Delta\chi^2 = 6.50$, $df = 1$, $p < .01$). The intercept estimate for the individualized group was 84.84. For the standardized protocol condition, it was 85.10. The comparison group average intercept was 79.20. Given the statistically significant differences in intercept, we also contrasted slope estimates. Slope for the combined treatment group was −.51 (.00 in the standardized group and −.97 for the individualized). For the comparison, slope was .45. This difference was also statistically significant ($\Delta\chi^2_{slope} = 3.68$, $df = 1$, $p < .05$). The contrast of the individualized slope and standardized slope, although comparatively large (.00 versus −.97), was not statistically significant ($\Delta\chi^2 = 5.91$, $df = 3$, $p > .05$). Effect sizes were moderately sized, − .52 for the individualized group and .56 for the standardized protocol condition. The standardized difference between the standardized protocol and individualized conditions was −.03.

Moderating Effects

We estimated variation in treatment effects across special education status and language minority status by regressing the variables of interest (e.g., special education and primary language status) on the group-specific intercepts derived from the earlier fit models (i.e., Table 1). The values in Table 3 are unstandardized regression coefficients for each treatment condition. They represent differences in standard score points for different levels of the variable in question. Table 3 also summarizes differences related to special education and primary language across the three treatment conditions (i.e., moderated effects).

Special Education and Limited English Proficiency—On WJ-III Letter Word Identification, students in special education who were in the comparison condition performed more poorly than their counterparts who were not in special education by almost 14 scale score points ($\gamma_a = -13.57$, $p < .001$, where γ_a represents the coefficient for a dummy-coded variable representing special education status and the associated p-value reflects the extent to which it differs statistically from 0). In the individualized group, the difference between special education students and students who did not participate in special education was approximately 17 points ($\gamma_a = -16.72$, $p < .001$). Special education students fared less well than their peers who did not participate in special education in the individualized and comparison conditions. By contrast, special education-related differences within the standardized protocol were not statistically significant ($\gamma_a = -7.99$, $p = .158$), suggesting that the two groups were comparable in spring 2008. Across conditions (i.e., the moderating effect of special education status), special education students in the standardized protocol were at less of a disadvantage than special education participants in the individualized group ($\Delta\chi^2 = 5.72$, $p < .05$). In other words, the special education-related disadvantage in the individualized condition was greater than the disadvantage for students in the standardized protocol. There were no statistically significant differences between the comparison group and either of the treatment conditions.

On WJ-III Word Attack, special education students lagged significantly behind their peers who were not in special education in all three conditions ($\gamma_a = -11.90$, $p < .001$; $\gamma_a = -8.76$, $p < .001$; $\gamma_a = -10.99$, $p < .001$ for individualized, standardized protocol, and comparison conditions, respectively), although these values did not significantly differ across the groups.

On WJ Passage Comprehension, special education students in the individualized condition were significantly less successful than participants who were not in special education in the same condition ($\gamma_a = -7.97$, $p = .002$). Differences in the standardized protocol and the comparison conditions did not differ statistically from 0. Across conditions, the difference between individualized participants and the comparison group was statistically significant

$(\Delta\chi^2 = 3.95, p < .05)$. Patterns on WJ-III Spelling, TOWRE—Sight Word, and TOWRE—Phonemic Decoding were generally similar. Students in special education performed less well than students who were not in special education in the same condition, and no statistically significant differences occurred across conditions in the effect of special education on the spring 2008 performance.

There were no statistically significant differences across treatment conditions in the effect of LEP status on intercept. On the TOWRE—Phonemic Decoding, LEP students outperformed non-LEP students in the standardized protocol ($\gamma_a = 11.79, p = .004$) and in the individualized ($\gamma_a = 11.46, p = .016$) treatment conditions. Within groups, performance on WJ-III Passage Comprehension was significantly lower for LEP students than for non-LEP students in the comparison ($\gamma_a = -9.57, p = .014$) and in the standardized protocol condition ($\gamma_a = -7.28, p = .016$).

DISCUSSION

We report findings from a study in which we randomly assigned students who did not meet exit criteria from a previously provided Tier 1 (enhanced instructional practices in vocabulary and comprehension for all content-area teachers) and Tier 2 intervention (supplemental daily reading instruction for at-risk students; Vaughn, Cirino, et al., 2010; Vaughn, Wexler, et al., 2010) to one of two conceptually different interventions: standardized or individualized treatments. Students whom the researchers had randomly assigned to the comparison condition (Tier 1 only) and were low responders in Year 1 remained in the comparison condition for Year 2.

Standardized Versus Individualized Treatments

The primary question addressed is the efficacy of two tertiary interventions for students with reading disabilities. Findings did not confirm our research hypothesis that students in the individualized treatment would outperform students in the standardized treatment. These findings aligned with previous research with beginning readers who had reading difficulties in which researchers compared a more standardized treatment with a more responsive or individualized treatment and found no statistically significant differences between the two treatments (Mathes et al., 2005). We do not believe that this single study provides convincing data that standardized approaches might be at least as effective as more individualized approaches for secondary students with intensive reading difficulties, but we do think that it provides compelling data to consider when designing interventions for older students with reading disabilities. Educators should also consider the findings in light of the significant level of training, supervision, and feedback provided to the teachers in this study, because it might not represent the level or quality of training typically provided to teachers.

When considering the findings for the treatments combined (standardized and individualized combined), we found statistically significant differences for reading comprehension but not for tasks involving word reading, word attack, or fluency. We consider the impact on reading comprehension meaningful in light of the challenge to successfully influence the reading comprehension of students with significant reading problems.

Special Education Status and English Language Proficiency

We had hypothesized that students with identified disabilities might perform significantly better in the individualized condition than in the standardized condition; however, the results did not confirm our hypothesis. Overall, students identified with disabilities (i.e., served by special education) were at more of a disadvantage (i.e., had poorer outcomes) in the individualized condition than in the standardized condition. This finding was valid for word

attack and for reading comprehension. Additionally, the word attack and reading comprehension outcomes for students with disabilities were significantly lower than outcomes for their peers who did not have identified disabilities. We found this outcome for all three conditions (standardized, individualized, and comparison).

We also anticipated that students with LEP might perform better in the individualized condition because there was a greater focus on responding to students' individual learning needs and increasing students' academic motivation through building home-school connections, goal setting, and choice. Neither treatment condition was significantly more effective for students who were identified as having LEP. Students with the LEP designation demonstrated overall lower scores on reading comprehension than non-LEP students. However, on phonemic-decoding, LEP students outperformed non-LEP students.

The Impact of Reading Intervention on Older Students' Comprehension

The findings from several similar studies (Corrin, Somers, Kemple, Nelson, & Sepanik, 2008; Kemple et al., 2008; Lang et al., 2009), as well as our previous studies (Vaughn, Cirino, et al., 2010; Vaughn, Wanzek, et al., 2010), indicate that a 1-year-long intervention will adequately meet the needs of relatively few struggling middle school readers and that most students, particularly those with significant reading problems, will require more intensive interventions that last for longer than 1 year. Considering these findings, we believe that the progress of the students in our treatment conditions in reading comprehension is noteworthy. These students made significant gains in reading comprehension, as evidenced by a moderately high impact (ES = .56); and we found an association between these gains and acceleration in their standard score performance of about one third of a standard deviation, suggesting that not only were students improving in their overall outcomes in reading comprehension but that they were also closing the gap between their current reading performance and grade-level expectations. We do not believe that students in middle grades with significant reading problems are likely to make rapid and readily remediated progress in reading. Many of these students with low comprehension also demonstrate low vocabulary and limited background knowledge, which is associated with low reading comprehension (Cromley & Azevedo, 2007); and compensating for these challenges is not going to happen in a 50-min-long daily intervention.

Limitations and Future Directions

The context for this study is within relatively large urban settings in which most of the students are from low-income homes. Findings for students in other settings in which more home resources mitigate the daily challenges of schooling may be different. The findings might also have differed if the researchers had specifically selected students with defined reading disabilities for the study or if the study had used a benchmark other than performance on a state reading comprehension test. However, our sample included many identified for special education, and the benchmark is valid.

The cost of this intervention is an issue to consider. In this intervention, well-trained, experienced teachers taught students in groups of five students. This group size is costly when compared with typical class instruction, in which there may be one teacher for every 20 to 25 students. We believe that in addition to improved student outcomes in reading (realized by students in the treatment conditions), determining whether treated students are also more likely to remain in school and graduate would be valuable. The impact from this study on reading comprehension for treatment students yields an effect size that is larger than that in many studies of struggling adolescent readers who received a year of intervention, even though we selected the group because of evidence of intractability to a previous year of intervention, representing a group with severe reading problems. The study

points to the need to seriously consider the intensity needed to remediate reading difficulties in middle school. Even with 2 years of intervention, most students do not evidence grade-level reading for understanding and will require further intervention.

Acknowledgments

References

Archer AL, Gleason MM, Vachon VL. Decoding and fluency: Foundation skills for struggling older readers. Learning Disabilities Quarterly. 2003; 26:89–101.

Bergan, JR. Behavioral consultation. Columbus, OH: Merrill; 1977.

Biancarosa, G.; Snow, CE. Reading next—A vision for action and research in middle and high school literacy: A report to Carnegie Corporation of New York. Washington, DC: Alliance for Excellent Education; 2004.

Bovaird, JA. Multilevel structural equation models for contextual factors. In: Little, TD.; Bovaird, JA.; Card, NA., editors. Modeling contextual effects in longitudinal studies. Mahwah, NJ: Lawrence Erlbaum Associates; 2007. p. 149-182.

Chen, H-T. Theory-driven evaluations. Newbury Park, CA: Sage; 1990.

Corrin, W.; Somers, M-A.; Kemple, J.; Nelson, E.; Sepanik, S. The enhanced reading opportunities study: Findings from the second year of implementation (NCEE 2009–4036). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education; 2008.

Cromley JG, Azevedo R. Testing and refining the direct and inferential mediation model of reading comprehension. Journal of Educational Psychology. 2007; 99:311–325.10.1037/0022-0663.99.2.311

Deno SL, Fuchs LS, Marston D, Shin J. Using curriculum-based measurement to establish growth standards for children with learning disabilities. School Psychology Review. 2001; 30:507–524.

Denton CA, Fletcher JM, Anthony JL, Francis DJ. An evaluation of intensive intervention for students with persistent reading difficulties. Journal of Learning Disabilities. 2006; 35:447–466. [PubMed: 17004676]

DeStefano, L.; Snauwaert, D. A value-critical approach to transition policy analysis (No. 300-85-0160). Washington, DC: U.S. Department of Education; 1989.

Edmonds MS, Vaughn S, Wexler J, Reutebuch CK, Cable A, Tackett KK, Schnakenberg JW. A synthesis of reading interventions and effects on reading comprehension outcomes for older struggling readers. Review of Educational Research. 2009; 79:262–300. [PubMed: 20072704]

Fuchs D, Fuchs LS, Compton DL. Identifying reading disabilities by responsiveness-to-instruction: Specifying measures and criteria. Learning Disabilities Quarterly. 2004; 27:216–227.

Fuchs D, Mock D, Morgan PL, Young CL. Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. Learning Disabilities Research & Practice. 2003; 18:157–171.

Kamil, ML.; Borman, GD.; Dole, J.; Kral, CC.; Salinger, T.; Torgesen, J. 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; 2008. Retrieved from http://ies.ed.gov/ncee/wwc

Kemple, J.; Corrin, W.; Nelson, E.; Salinger, T.; Herrmann, S.; Drummond, K. The enhanced reading opportunities study: Early impact and implementation findings (NCEE 2008–4015). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education; 2008.

Lang L, Torgesen J, Vogel W, Carol C, Lefsky E, Petscher Y. Exploring the relative effectiveness of reading interventions for high school students. Journal of Research on Educational Effectiveness. 2009; 2:149–175.10.1080/19345740802641535

Mathes PG, Denton CA, Fletcher JM, Anthony JL, Francis DJ, Schatschneider C. The effects of theoretically different instruction and student characteristics on the skills of struggling readers. Reading Research Quarterly. 2005; 40:148–182.

Mehta PD, Neale MC. People are variables too: Multilevel structural equations modeling. Psychological Methods. 2005; 10:259–284. [PubMed: 16221028]

National Assessment of Educational Progress. The nations report card: Trial urban district assessment reading 2007 (NCES 2008–455). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education; 2007.

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

Texas Education Agency. Appendix 10—Technical digest 2004–2005. 2004. Retrieved from http://www.tea.state.tx.us/index3.aspx?id=4397&menu_id=793

Torgesen JK, Wagner RK, Rashotte CA, Lindamood P, Rose E, Conway T, Garvan C. Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. Journal of Educational Psychology. 1999; 91:579–593.

Vaughn S, Cirino PT, Wanzek J, Wexler J, Fletcher JM, Denton CA, Francis DJ. Response to intervention for middle school students with reading difficulties: Effects of a primary and secondary intervention. School Psychology Review. 2010; 39:3–21. [PubMed: 21479079]

Vaughn S, Wanzek J, Wexler J, Barth A, Cirino PT, Fletcher J, Francis DJ. The relative effects of group size on reading progress of older students with reading difficulties. Reading and Writing: An Interdisciplinary Journal. 2010; 23:931–956.10.1007/s11145-009-9183-9

Wanzek J, Vaughn S. Research-based implications from extensive early reading interventions. School Psychology Review. 2007; 36:541–561.

Williams JP. Commentary: Four meta-analyses and some general observations. Elementary School Journal. 2001; 101:349–354.

Wilson Language Training Corporation. Wilson Reading System. Willbury, MA: Author; 1996.

Woodcock, RW.; McGrew, K.; Mather, N. Woodcock-Johnson III tests of achievement. Itasca, IL: Riverside; 2001.

TABLE 1

Effect Estimates for Treatment Condition

Outcome	χ^2/df(p)[a]	CFI/TLI	RMSEA	I (p)	Intercept σ^2	S (p)	Δ_{ltotal}[b]	Δ_{ltreat}[c]
WJ Letter Word Identification	10.33/9 (.324)	.99/.99	.05					
Treatment—individual				89.77 (<.001)	163.7 (<.001)	−.549 (.110)	.28	−.11
Treatment—standard				91.09 (<.001)	108.7 (<.001)	−.612 (.22)	.44	
Comparison				86.34 (<.001)	131.2 (<.001)	−.290 (.630)	—	
WJ Word Attack	4.952/9 (.838)	.99/.99	.001					
Treatment—individual				91.78 (<.001)	129.1 (<.001)	.501 (.331)	.14	−.27
Treatment—standard				94.64 (<.001)	87.66 (<.001)	−.743 (.139)	.45	
Comparison				90.25 (<.001)	104.2 (<.001)	.551 (.221)	—	
WJ Spelling	6.36/9 (.703)	.99/.99	.001					
Treatment—individual				88.11 (<.001)	167.7 (<.001)	−.097 (.880)	.31	−.06
Treatment—standard				88.86 (<.001)	173.8 (<.001)	−.004 (.995)	.36	
Comparison				84.07 (<.001)	194.6 (<.001)	−.136 (.874)	—	
TOWRE Sight Word	21.62/9 (.010)	.97/.97	.154					
Treatment—individual				90.74 (<.001)	141.96 (<.001)	.063 (.856)	.24	.02
Treatment—standard				90.57 (<.001)	87.55 (<.001)	.363 (.452)	.27	
Comparison				87.91 (<.001)	125.71 (<.001)	.188 (.739)	—	
TOWRE Phonemic Decoding	11.48/9 (.245)	.99/.99	.068					
Treatment—individual				91.91 (<.001)	263.9 (<.001)	.497 (.332)	.26	.01
Treatment—standard				91.70 (<.001)	212.1 (<.001)	.513 (.318)	.27	
Comparison				87.81 (<.001)	219.53 (<.001)	1.60 (.100)	—	
WJ Passage Comprehension	13.67/9 (.135)	.99/.99	.094					
Treatment—individual				**84.84 (.001)**	109.4 (.007)	**−.967 (.062)**	.52	−.03
Treatment—standard				**85.10 (<.001)**	100.2 (<.001)	**0.00 (.000)**	.56	
Comparison				79.20 (<.001)	135.4 (<.001)	.455 (.345)	—	

Note. Values in bold italics differ from nonitalicized, nonbolded values at *p* < .05. CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation. TOWRE = Test of Word Reading Efficiency. WJ = Woodcock-Johnson test.

[a]Represents the multigroup (*N* = 3) model with slope fixed.

[b] ΔI_{total} represents the effect for the two treatment groups in relation to the comparison. Effect sizes were calculated as the difference in estimated means (i.e., model-derived intercepts) divided by the weighted, pooled standard deviation for each estimate, adjusted for small sample size using Hedges's g formula.

[c] ΔI_{treat} represents the difference between the two treatment groups (i.e., standardized and individualized). Effect sizes were calculated as the difference in estimated means (i.e., model-derived intercepts) divided by the weighted, pooled standard deviation for each, adjusted for small sample size using Hedges's g formula. Effects are reported in relation to the individualized treatment group.

TABLE 2

Descriptive Statistics for Outcome Measures by Group and Measurement Occasion

Group	Measurement Occasion	WJ Word Identification	WJ Word Attack	WJ Spelling	TOWRE Sight Word	TOWRE Phonemic Decoding	WJ Passage Comprehension
Individualized	Spring 2007	88.97 (13.25)	93.22 (13.59)	88.33 (13.35)	91.25 (12.65)	93.18 (17.52)	83.22 (11.70)
	Fall 2007	89.23 (13.28)	91.77 (12.69)	88.56 (13.68)	91.40 (11.19)	91.67 (15.89)	84.06 (10.24)
	Spring 2008	91.52 (14.25)	92.08 (12.14)	89.32 (15.38)	91.50 (12.68)	91.67 (17.44)	86.77 (12.57)
Standardized	Spring 2007	90.18 (11.18)	93.42 (10.91)	88.45 (14.76)	91.82 (9.81)	92.81 (14.53)	85.34 (10.93)
	Fall 2007	88.95 (10.85)	92.71 (10.83)	88.18 (13.43)	89.60 (11.17)	89.56 (16.77)	84.78 (10.93)
	Spring 2008	91.39 (12.30)	94.89 (9.51)	85.64 (16.30)	91.37 (11.31)	92.76 (17.48)	85.59 (11.22)
Comparison	Spring 2007	86.17 (13.04)	91.39 (10.48)	83.51 (14.80)	88.37 (12.30)	91.41 (15.85)	79.98 (12.35)
	Fall 2007	85.64 (12.68)	90.11 (12.44)	84.58 (14.93)	88.22 (12.56)	89.36 (16.30)	79.86 (14.00)
	Spring 2008	87.49 (12.47)	91.46 (11.53)	82.73 (18.74)	88.60 (12.03)	91.17 (19.30)	78.06 (11.76)

Note. Group means with standard deviations in parentheses. TOWRE = Test of Word Reading Efficiency. WJ = Woodcock-Johnson test.

TABLE 3

Moderating Effects of Treatment for Special Education and Limited English Proficiency (LEP)

Outcomes	Special Education Language	Minority Status
WJ Letter Word Identification		
Treatment—individual	−16.72 (.001)[a]	3.70 (.375)[a]
Treatment—standard	−7.99 (.158)[b]	3.14 (.335)[a]
Comparison	−13.57 (.001)[a,b]	.002 (.999)[a]
WJ Word Attack		
Treatment—individual	−11.90 (.001)[a]	4.62 (.226)[a]
Treatment—standard	−8.76 (.001)[a]	1.38 (.643)[a]
Comparison	−10.99 (.001)[a]	2.02 (.524)[a]
WJ Spelling		
Treatment—individual	−15.44 (.001)[a]	−3.15 (.464)[a]
Treatment—standard	−10.45 (.001)[a]	−2.61 (.587)[a]
Comparison	−18.05 (.001)[a]	−1.82 (.711)[a]
TOWRE Sight Word		
Treatment—individual	−6.59 (.026)[a]	2.99 (.407)[a]
Treatment—standard	−8.57 (.001)[a]	4.50 (.116)[a]
Comparison	−11.32 (,001)[a]	.680 (.864)[a]
TOWRE Phonemic Decoding		
Treatment—individual	−13.64 (,001)[a]	11.46 (.016)[a]
Treatment—standard	−12.33 (.001)[a]	11.79 (.004)[a]
Comparison	−20.06 (.001)[a]	.53 (.388)[a]
WJ Passage Comprehension		
Treatment—individual	−7.97 (.002)[a]	−3.68 (.291)[a]
Treatment—standard	−4.14 (.118)[a,b]	−7.28 (.016)[a]
Comparison	1.90 (.684)[b]	−9.57 (.014)[a]

Note. Values are unstandardized coefficients for model-estimated intercept regressed on the moderators in question. Estimates are group differences in standard score points for the relationship of special education and language minority status and the various outcome measures. Non-special education and non-LEP were used as the comparisons (i.e., dummy-coded as 0), meaning that a negative coefficient within a given condition indicates lower performance for special education students versus students who were not in special education and LEP students versus non-LEP students. Estimates with the same superscripted letters do not differ significantly. Estimates with different superscripts differ significantly at $p < .05$ or greater. For example, although the average reading comprehension score for special education students in the standardized protocol condition did not differ from the comparison or from the individualized condition, the mean for the individualized condition did differ from the comparison mean. TOWRE = Test of Word Reading Efficiency. WJ = Woodcock-Johnson test.