Journal of Educational Psychology
2012, Vol. 104, No. 4, 987–1005

0022-0663/12/$12.00 DOI: 10.1037/a0028163

Two-Year Follow-Up of a Kindergarten Phonics Intervention for English Learners and Native English Speakers: Contextualizing Treatment Impacts by Classroom Literacy Instruction

Patricia F. Vadasy
Washington Research Institute

Elizabeth A. Sanders
University of Washington

The current study follows a sample of lower skilled language minority (LM) and native English-speaking (non-LM) students who participated in an efficacy trial of a kindergarten phonics-based intervention. Follow-up procedures allowed 93% of the original sample to be retained for simple treatment effects modeling (N = 78 LM and N = 59 non-LM) and 72% to be retained for classroom instruction modeling (N = 62 LM and N = 44 non-LM). Simple treatment effects on longer term outcomes were detected on word reading, spelling, and comprehension outcomes for LM students (approximate effect sizes averaged 0.27); comparatively, treatment effects for non-LM students were detected on all outcomes, including fluency (approximate effect sizes averaged 0.54). Instructional model results showed that greater time in Grade 1 word study instruction and Grade 2 meaning instruction was associated with higher reading scores for LM students at the end of second grade, irrespective of experimental condition. For non-LM children, greater time in Grade 1 meaning instruction was connected with higher reading scores at the end of Grade 2, irrespective of experimental condition. Finally, kindergarten intervention effects tended to be greater for both LM and non-LM students who received more Grade 1 word study instruction and more Grade 2 meaning instruction. Limitations and practical implications are discussed.

Keywords: English learner, reading, intervention, longitudinal studies, randomized trial

Acquisition of early reading skills is important for a successful trajectory of reading development. Many studies have shown that children who begin school as poor readers are likely to remain poor readers (e.g., Foorman, Francis, Shaywitz, Shaywitz, & Fletcher, 1997; Juel, 1988; Pianta, Belsky, Vandergrift, Houts, & Morrison, 2008; Stanovich, 1986), although there is also evidence that the initial gap between young children with good and poor reading skills can be narrowed (e.g., Parrila, Aunola, Leskinen, Nurmi, & Kirby, 2005; Phillips, Norris, Osmond, & Maynard, 2002). Students known to be at greater risk for difficulty acquiring beginning reading skills include children from low socioeconomic status (Kaplan & Walpole, 2005) and language minority backgrounds (August & Shanahan, 2006; Rueda & Windmueller, 2006), and these children are likely to benefit from interventions that help them establish reading foundation skills, in particular alphabetic and decoding skills. Alphabet instruction warrants careful study in light of the solidly established importance of alphabetic knowledge in word reading and spelling development (e.g., Catts, Fey, Tomblin, & Zhang, 2002; Ehri, 1987, 1998; Scarborough, 1998). Systematic phonics instruction, beginning in kindergarten, enables students to apply their knowledge of the alphabetic code to read and spell words (Ehri, Nunes, Stahl, & Willows, 2001), the early literacy skills that most often challenge beginning readers.

A robust body of research highlights features and components of effective early reading interventions in alphabetic, phonemic awareness, decoding, and word identification skills. Features include explicitness and intensity to establish accurate decoding skills (Foorman, Francis, Fletcher, & Schatschneider, 1998; Foorman & Torgesen, 2001; Torgesen et al., 2001), which may be more difficult to acquire in opaque orthographies like English (Goswami, Gombert, & de Barrera, 1998; Landerl, Wimmer, & Frith, 1997). Children arriving in kindergarten with limited early literacy skills often require intensive intervention to learn these skills. Meta-analyses of studies on phonological awareness (Bus & van IJzendoorn, 1999) and phonics (Ehri et al., 2001) instruction support the effectiveness of teaching phonemic and alphabetic skills in kindergarten (Bus & van IJzendoorn, 1999) and the benefits of explicit phonics instruction, in particular when instruction is provided before first grade (Ehri et al., 2001). There is a strong body of research guiding early reading intervention for at-risk kindergarteners who are native English speakers.

This article was published Online First April 30, 2012.

Patricia F. Vadasy, Washington Research Institute, Seattle, Washington; Elizabeth A. Sanders, Department of Educational Psychology, University of Washington.

The curriculum evaluated in this research has been published, and the first author thus could have a vested interest in the results. This research was supported with funding from the Institute of Education Sciences, U.S. Department of Education, Grant R305A070324. The opinions expressed are our own and do not represent the view of the U.S. Department of Education. We thank Sueanne Sluis and Becky Logan Herrera for coordinating the intervention and assessments and Kathryn Compton for assessment and data support. We are especially grateful to participating schools, tutors, and students for making this research possible.

Correspondence concerning this article should be addressed to Patricia F. Vadasy, Washington Research Institute, 150 Nickerson Street, Suite 305, Seattle, WA 98109. E-mail: pvadasy@wri-edu.org

Two Lacunae in the Research

The research base is less well developed in two areas concerning reading interventions for the increasing number of young English learners in U.S. schools (National Clearinghouse for English Acquisition and Language Instruction Educational Programs, 2006). First, there is less research about the benefits of phonics-based early reading interventions for these students. In their original review of intervention studies for young English learners, Shanahan and Beck (2006) found that the small number of available studies of phonological awareness and phonics interventions demonstrated reading benefits similar to those summarized in the National Reading Panel's (2000) review of 90 studies for native English-speaking students. However, Shanahan and Beck (2006) concluded there remained limited high-quality studies to draw specific conclusions for English learners. In a research update on reading instruction for English learners that included 20 studies published since 2002, August and Shanahan (2010) reviewed 12 additional studies on phonological awareness and phonics instruction. Similar to the Shanahan and Beck review, theirs found that systematic and explicit instruction has a positive impact on a range of student reading outcomes. English learners develop word-level skills in a similar manner to native English speakers, and they benefit from instructional features that have been found effective for native English-speaking children (Shanahan & Beck, 2006). Several early reading interventions for English learners with a strong phonics component have reported benefits for word-level skills (Denton, Anthony, Parker, & Hasbrouck, 2004; Gunn, Biglan, Smolkowski, & Ary, 2000). This growing body of research suggests that young English learners benefit from the same instructional features and phonics components that support early reading development in their native English-speaking peers (Ehri et al., 2001). Yet important questions remain to be addressed regarding early intervention for English learners, including components of classroom literacy instruction, optimal timing for supplemental intervention, and precursor reading and language skills that predict treatment response. Comparative research can inform differences in response for English learners and native English speakers to widely used intervention approaches (Lesaux & Kieffer, 2010).

The second lacuna in the research concerns the long-term benefits of early reading interventions for at-risk children, including English learners. Added to the typical challenges of longitudinal research (e.g., cost, effort, attrition) are the difficulties of retaining adequate samples of English learners and at-risk students from populations characterized by greater mobility. Vadasy, Sanders, and Abbott (2008) earlier summarized the limited follow-up studies of early reading interventions for native English-speaking children and found that beginning readers at risk and with reading disabilities who received a code-oriented intervention maintained growth in word-level skills, with a moderate effect size at 1-year follow-up. Others have reported on the treatment effects at follow-up for cohorts of English learners receiving early reading interventions (Cirino et al., 2009; Gunn et al., 2000; Gunn, Smolkowski, Biglan, & Black, 2002; Gunn, Smolkowski, Biglan, Black, & Blair, 2005; Vaughn et al., 2008). The present study is designed to address these two lacunae in research on the increasing numbers of young language minority (LM) students in U.S. public schools. Below, we briefly review research on early interventions for young LM children. Next we review research on child and instruction variables found to moderate intervention outcomes. Finally, we summarize treatment outcomes from the supplemental kindergarten intervention that we follow up in the present study.

Educationally Effective Early Reading Interventions for Young LM Students

As we noted earlier, features of effective early reading interventions have been clearly identified. The major challenges for beginning readers, including LM students, are mastery of the alphabetic system and learning to decode words (Lesaux, Koda, Siegel, & Shanahan, 2006), the most common obstacles to early reading development. In its review, the National Reading Panel (2000) clearly supported the benefits of explicit instruction in phonemic awareness and phonics skills. In their subsequent meta-analysis of systematic phonics instruction, Ehri et al. (2001) examined more closely the benefits of phonics instruction for specific groups of students, although not for LM students. They found it was most effective in kindergarten and first grade and was effective in varied delivery systems, including individual instruction. Research indicates that most young LM students develop proficient word-level skills similar to those of first language-speaking peers (Lesaux et al., 2006), in particular when provided with explicit and carefully designed instruction. English language proficiency and vocabulary knowledge begin to have a strong influence on reading comprehension for English learners by first grade (Carlisle & Beeman, 2000; Carlisle, Beeman, Davis, & Spharim, 1999; Verhoeven, 1990). However, limited research on early reading development of English learners does not clearly identify instructional features that may benefit students with different levels of literacy and language skill (Shanahan & Beck, 2006).

Supplemental Intervention Mediators and Moderators

Classroom Instructional Practices

Student response to supplemental early reading interventions is influenced by child and environment characteristics including literacy skills at school entry as well as other reading instruction and experience, in particular the classroom reading environment (Connor, Morrison, Fishman, Schatschneider, & Underwood, 2007; Connor, Morrison, & Katch, 2004; Connor, Morrison, & Petrella, 2004; Morrison, Connor, & Bachman, 2006). Sonnenschein, Stapleton, and Benson (2010) recently examined the relation of reading development through fifth grade to the classroom reading instructional focus, using data from the Early Childhood Longitudinal Study—Kindergarten cohort (ECSL-K). They found that meaning-oriented reading instruction benefitted more the children who entered kindergarten and first grade with more advanced phonics skills, and the influence of type of classroom instruction was limited to these early school years. Their focus on both type and amount of classroom reading instruction is similar to the instructional emphases we measured in our classroom observations described below. In the intervention year study, we reported that treatment effect of the supplemental instruction was influenced by classroom reading instruction: phonics instruction had significant positive effects on kindergarten phonological awareness and comprehension outcomes. We earlier speculated that the phonics instructional time influence on end of kindergarten comprehension may have been a time-sensitive effect observed in the early grades when word reading is more highly correlated with reading comprehension (e.g., Aunola, Nurmi, Niemi, Lerkkanen, & Rasku-Puttonen, 2002; Parrila, Kirby, & McQuarrie, 2004; Wagner et al., 1997) and that oral language skills more strongly influence comprehension in later school years (Proctor, Carlo, August, & Snow, 2005).

In the current study's follow-up classroom observations, our primary focus was on the balance between word study-focused and meaning level-focused classroom instruction. Others have reported interactions between classroom instruction practices and student skill levels (Connor, Morrison, & Katch, 2004; Connor, Morrison, & Petrella, 2004). In the present study we concentrate on word study and meaning instruction in Grades 1 and 2 classrooms. We hypothesized that an instructional focus on meaning would have increasing positive benefits for English learners, in particular as they move beyond the initial stage of reading.

LM Status

As noted above, beginning reading skills strongly predict later reading performance (Juel, 1988; Pianta et al., 2008; Torgesen & Burgess, 1998), and low-income and language status are associated with lower levels of initial literacy skills and later literacy development (Biemiller & Slonim, 2001; Kaplan & Walpole, 2005; Rueda & Windmueller, 2006). A robust body of research supports the contributions of oral language skills to later reading achievement (e.g., Bishop & Adams, 1990; Scarborough, 1989; Share, Jorm, Maclean, & Matthews, 1984). Early literacy skills that are well-established predictors of reading outcomes include alphabet knowledge (Share et al., 1984; Tunmer, Herriman, & Nesdale, 1988) and phonological awareness (Bus & van IJzendoorn, 1999; Shankweiler & Liberman, 1989; Share, 1995; Snowling, 1991), skills that also predict reading difficulties in English learners (Chiappe & Siegel, 1999; Chiappe, Siegel, & Wade-Wooley, 2002; Droop & Verhoeven, 2003). Most recently Boscardin, Muthen, Francis, and Baker (2008) used a growth mixture modeling approach to examine the developmental profiles of children with a range of precursor reading skills. Developmental profiles at kindergarten predicted later reading development, and membership in the lowest performing group was associated with minority status.

Original Intervention Study

In earlier research Vadasy and Sanders (2010) tested the efficacy of supplemental code-oriented instruction for $n = 84$ low-skilled LM and $n = 64$ non-LM kindergarteners at 10 urban public schools. Paraeducators were trained to provide the 18-week (January to May) intervention. Students performing in the bottom half of their classroom language group (LM and non-LM) were randomly assigned either to individual supplemental instruction (treatment) or to classroom instruction only (control). Irrespective of LM status, treatment students ($n = 67$) significantly outperformed controls ($n = 81$) at posttest in alphabetics, word reading, spelling, passage reading fluency, and comprehension (average treatment $d = 0.83$); nevertheless, LM students tended to have lower posttest performance than non-LM students (average LM $d = -0.30$) and were significantly less responsive to treatment on word reading. When we examined the contribution of classroom phonics time to student outcomes, we found that the treatment effect on spelling was greater for students in lower phonics classrooms whereas the treatment effect on comprehension was greater in higher phonics classrooms. Finally, when we examined LM students alone, we found that pretest English receptive vocabulary positively predicted most posttests and interacted only with treatment on phonological awareness. In general, pretest vocabulary did not moderate kindergarten LM treatment response.

Current Study

The present study tests longer term treatment effects for LM and non-LM students 2 years after kindergarten posttest. Additionally, we test the unique contributions of average daily meaning and word study instruction students received to students' final outcomes. Finally, we test whether longer term treatment effects depend on kindergarten pretest alphabetic knowledge and instructional time variables. Research questions for the current study are as follows:

1. Do treatment effects on LM and non-LM students' literacy outcomes remain 2 years after kindergarten posttest?

2. Do kindergarten pretest and instructional time variables uniquely contribute to LM and non-LM students' final outcomes?

3. Do kindergarten pretest and instructional time variables moderate longer term treatment effects for LM and non-LM students?

Method

Participants

Original sample. Because it is important to understand sample characteristics for this follow-up study, we review original study procedures. In October of 2007–2008, all students in full-day kindergarten classrooms at 10 urban public elementary schools were invited to participate in the study. All participating schools were designated Title I and had student enrollments averaging 85% minority, 75% free or reduced lunch, 33% bilingual, and 18% special education during the intervention year. At study onset, students were defined as LM if a parent reported the primary home language as other than English on the school registration record (Lesaux & Kieffer, 2010; Lesaux & Siegel, 2003). All invitations and consent forms were sent home in English and in translations for the following languages: Spanish (60% of LM students), Vietnamese (13%), Somali (11%), Cambodian (5%), Chinese (5%), Cham (2%), Tigrigna (2%), Samoan (1%), and Tagalog (1%).

Once consents were received (62% return rate), students were screened on early literacy skills with three measures: two measures of alphabetic knowledge (letter name and sound accuracy) and a norm-referenced test of phonological awareness. We used a composite mean of the screening measures to rank order students, within LM and non-LM subgroup within classroom, from lowest to highest; students in the lower half of their classroom's LM or non-LM group were then randomly assigned to treatment (supplemental reading tutoring) or control (regular classroom instruction, no tutoring) conditions. This process ensured approximately equal representation of classrooms and LM and non-LM groups in each of the two experimental conditions; simultaneously, this meant that the students in the sample were at relative, rather than cut-score, risk for reading difficulties. After attrition, the kindergarten sample included $n = 84$ LM students (38 treatment) and $n = 64$ non-LM students (29 treatment), from 24 classrooms across 10 schools.

Follow-up sample. We attempted to follow up all 148 students in the original sample at every test wave occurring in the fall and spring of the 2 years subsequent to kindergarten. Our guideline to minimize attrition was to test any student who moved within a 30-mile radius of his or her original kindergarten school. Any student with at least two follow-up tests out of the four test waves (not necessarily consecutive) was included in our first set of simple treatment effect analyses, yielding a follow-up sample of 137 students (78 LM and 59 non-LM; 93% of the original sample). With only one exception, simple t tests showed no significant difference on kindergarten pretests or posttests between those who were included in the follow-up sample and those who were not, for both LM and non-LM students ($ps > .05$). The exception was that, at kindergarten posttest, LM students who moved out of testing range had significantly lower phonological awareness ($n = 6$, $M = 79.50$, $SD = 5.54$) than LM students who we were able to follow up ($n = 78$, $M = 90.09$, $SD = 10.43$), $F(1, 82) = 6.01$, $p < .05$. Although we sometimes refer to the two follow-up years in this study as first and second grades for brevity, we also note that eight of the students we followed were retained either in kindergarten or in first grade (four LM and four non-LM). All of them were in the control condition.

Finally, we were able to capture follow-up classroom literacy block observation data (procedures described below) for 106 (62 LM and 44 non-LM) of the 137 students who had at least two test waves completed (including seven of the eight retained students). Thus, observation data were available for 77% of the follow-up sample for instructional model analyses. Without exception, simple t tests showed no significant differences between students who had follow-up classroom observation data and those who did not for both LM and non-LM students ($ps > .05$). Demographic information for the follow-up sample is provided in Table 1.

Original Study Intervention Procedures

Original study intervention procedures are briefly reviewed here. Students assigned to treatment received individual systematic and explicit phonics tutoring instruction in English, which included letter-sound correspondences, phonemic decoding, spelling, and assisted oral reading practice in decodable texts. Instruction occurred during the school day for 30 min/day, four days/week, over a period of 18 weeks (January–May). Each tutor ($n = 23$ paraeducators) was provided with a set of 70 scripted lessons (with 7–8 activities per lesson) matched to decodable texts for oral reading practice. In a typical tutoring session, paraeducators spent 20 min on phonics activities and 10 min scaffolding students' oral reading practice in decodable texts. Research staff provided ongoing coaching and modeling to support paraeducators' instruction (Foorman & Torgesen, 2001; Juel, 1996). By the end of intervention, treatment students received $M = 27.68$ hr of tutoring ($SD = 2.74$ hr) and completed $M = 64.99$ lessons ($SD = 13.69$).

Six research staff were trained to conduct onsite fidelity observations using a 5-point rating scale ranging from 1 (*never implements*) to 5 (*always implements*) for each instructional component. After training but prior to field observations, research staff viewed videotaped tutoring sessions of paraeducators implementing instruction with students; the internal consistency (Cronbach's alpha) of the observers' mean implementation ratings for the videotaped sessions was .97. By the end of the study, researchers had conducted a total of 156 fidelity observations for the 23 paraeducators; fidelity ratings averaged $M = 4.41$ ($SD = 0.57$).

Follow-Up Study Student Assessments

During the follow-up period, students were assessed each fall and spring for two consecutive years (first and second grades for

Table 1
Sample Demographic Characteristics

	Original kindergarten sample				Two-year follow-up sample							
					Students in simple models				Students in instructional models			
	LM (n = 84)		Non-LM (n = 64)		LM (n = 78)		Non-LM (n = 59)		LM (n = 62)		Non-LM (n = 44)	
Characteristic	N	%	N	%	N	%	N	%	N	%	N	%
Male	47	56	35	55	44	56	32	54	32	52	21	48
Bilingual	80	95	0	0	73	94	0	0	58	94	0	0
FRL	78	93	43	67	72	92	40	68	57	92	29	66
SPED	1	1	4	6	1	1	3	5	1	2	2	5
Minority	84	100	43	67	78	100	40	68	62	100	29	66
Asian	20	24	10	16	19	24	10	17	15	24	7	16
Black	13	15	21	33	13	17	19	32	11	18	14	32
Hispanic	49	58	10	16	44	56	9	15	35	56	7	16
Mixed/other	2	2	2	3	2	3	2	3	1	2	1	2

Note. LM = language minority; non-LM = native English speaking; Bilingual = receives bilingual services; FRL = eligible for free or reduced lunch; SPED = receives special education services.

most students) on measures of word reading, spelling, reading fluency, and reading comprehension. All tests were individually administered by testers unaware of group assignment. Standard scores were used for all measures except alphabetic knowledge. Test descriptions and sample reliabilities (Cronbach's alpha) are provided below.

1. *Alphabetic knowledge* was assessed in the late fall of kindergarten using the mean of two naming measures: letter names and sounds correctly produced in 1 min. (For each measure, each of the 26 letters of the alphabet in upper- and lowercase was randomly sorted and presented on a single page in six rows. Students were asked to name as many letters or to produce as many of the letter sounds as they could. The number correctly named or produced, respectively, was divided by the number of seconds the students took to finish the 52 items. This value was then multiplied by 60 to obtain names or sounds correct per minute. Internal consistency was .98 for both measures, for both LM and non-LM students.

2. *Word reading* was measured at each test wave using the mean standard score of the Word Attack and Word Identification subtests from the *Woodcock Reading Mastery Test—Revised/Normative Update* (WRMT-R/NU; Woodcock, 1987/1998). The Word Attack subtest includes 50 nonwords that increase in difficulty. For this sample, internal consistencies were .90, .95, .95, and .95 for LM students at each respective test wave and .95, .96, .97, and .96 for non-LM students. The Word Identification subtest consists of 106 words that increase in difficulty. Internal consistencies were .95, .97, .98, and .97 for LM students at each respective test wave and .97, .98, .97, and .98 for non-LM students.

3. *Spelling* was assessed at each test wave using the standard score from the *Wide Range Achievement Test—Fourth Edition* (WRAT-4; Wilkinson & Robertson, 2006) Spelling subtest, which requires students to write their own name, copy letters, and spell dictated, increasingly difficult words. Internal consistencies were .70, .80, .87, and .87 for LM students at each respective test wave and .76, .86, .91, and .91 for non-LM students.

4. *Reading fluency* was assessed at each test wave using the composite Passage Fluency score comprising the Rate and Accuracy subtests from the Gray Oral Reading Test-4 Form B (GORT; Wiederholt & Bryant, 2001). For these subtests, students read increasingly difficult passages aloud, beginning at their grade level. The Rate raw score is determined by the pattern of students' time to complete passages, and the Accuracy raw score is determined by the pattern of errors students make during passage reading. These scores are combined into a norm-referenced composite Fluency score. Rate subtest internal consistencies were .71, .89, .88, and .92 for LM students at each respective test wave and .93, .92, .94, and .95 for non-LM students. For the Accuracy subtest, internal consistencies were .71, .87, .78, and .86 for LM students and .94, .89, .86, and .91 for non-LM students.

5. *Reading comprehension* was assessed at each test wave with the GORT Comprehension subtest. Students read passages aloud, beginning at their grade level, and responded to a series of comprehension questions for each. To reduce possible "passageless" performance (cf. Keenan & Betjemann, 2006; Pierce, Katzir, Wolf, & Noam, 2007), we employed an adapted scoring of the GORT to reflect only responses on items pertaining to passages the students could read with some degree of fluency (i.e., the responses we used in the adapted scoring are those that students would be less likely to guess compared with passages the student could not read fluently). Internal consistencies were .62, .76, .84, and .84 for LM students at each respective test wave and .83, .85, .82, and .92 for non-LM students.

Classroom Literacy Instruction Observations

We hypothesized that two aspects of classroom literacy instruction could interact with treatment condition: the amount of time students received for word study-focused instruction (phonics and spelling) and time received for meaning-focused instruction (vocabulary and comprehension). Each year we invited teachers of students from our original study to complete literacy block instruction observations on three occasions (approximately 2.5 months apart in November/December, February/March, and May/June). If a student moved, we attempted to follow the student and his or her teacher. The only limit was that we observed classrooms with participating students who had not moved from the original school district (for which we had human subjects approval).

During the first follow-up year, we observed 45 classrooms; in the second follow-up year, there were 48 classrooms observed. Collapsed across both follow-up years, 71 different classroom combinations were observed. Given the large number of classroom combinations relative to the sample size of each subgroup (i.e., most students had unique classroom combinations), we treat observation data at the student level in our analyses.

Observation tool description. We used an adapted version of the *Instructional Content Emphasis—Revised* (ICE-R; Edmonds & Briggs, 2003) for measuring time spent on literacy content, which includes concepts of print, phonological awareness, alphabetic knowledge, word study/phonics, spelling, oral language development, fluency building, text reading, comprehension, and writing/language arts. We added two other codes: vocabulary and "other." Vocabulary captured all types of vocabulary instruction, including direct instruction in word meanings, translations of word meanings for LM students, strategy instruction in deriving word meanings from context, and vocabulary instruction embedded in text reading. "Other" captured nonliteracy content such as behavior management, evaluative feedback, transition time, and other content areas (e.g., math instruction).

Time recording process. Observations were conducted for the entire duration of teachers' literacy blocks. Observers began timing at the beginning of the observation. When classroom instruction began, observers coded the teacher's first instructional activity and thereafter, recorded each clock-time (running forward from zero) associated with instructional change, along with the appropriate instructional codes.

Observer reliability. In the first follow-up year, six certificated teachers were trained to conduct observations. In the second follow-up year, seven certificated teachers were trained (five of whom were observers from the first year). Each observer studied the ICE-R manual and coding instructions and participated in several training sessions to review the measure prior to onsite observations. To establish observer reliability (prior to onsite observations), observers coded videotapes of literacy instruction in primary grade classrooms (seven videotapes were used in the first follow-up year and eight new videotapes were used in the second year). Each year, reliabilities (internal consistencies) were calculated for each content category by treating each videotape as an observation and each classroom observer's computed time per instructional category as an item. Estimated observer reliabilities (Cronbach's alpha) are given in Table 2 for each content category, which average .96 in the first follow-up year and .95 in the second follow-up year. We note that none of the videotapes exhibited the concepts of print category; and the second-year set of videotapes did not exhibit oral language; hence, we use strikethrough font to make explicit that observation estimates for these categories should be treated with caution, as we were unable to establish reliabilities prior to field observations.

Mean minutes for content categories. For each follow-up year's three observation periods, we computed the mean daily literacy block time observed for each student's teacher or combination of teachers on a given content category. Descriptive statistics for the observation data are reported in Table 2. We note that the composite instruction variables (minutes spent on instruction with word study or meaning foci) were created by first summing the minutes spent on the respective categories at each observation occasion (for word study this was phonics and spelling time), and then these sums were averaged across the three observations for each year. Hence, the composite minutes shown in Table 2 do not directly sum from the category averages. It is also noteworthy that, for both LM and non-LM subgroups, the composite variables were not significantly correlated with each other within a given year, indicating that students who received more time in meaning-focused instruction did not necessarily receive more or less time in word study-focused instruction. In both follow-up years (Grades 1 and 2), the highest allocations of time students received during their literacy blocks were on phonics, text reading, and comprehension.

Analytic Strategy

We adopted multilevel modeling as our primary analytic tool for testing our research questions. Compared with traditional unilevel methods (e.g., analysis of variance and multiple regression), this more complex analysis method accounts for dependencies among measurement occasions within students and within schools. Throughout all analyses, we estimated models separately for LM and non-LM students in order to test our research questions without the complexity of three-way interactions among treatment, LM status, and the other predictor variables in the models. Models were estimated in HLM 6.08 (Raudenbush, Bryk, & Congdon, 2004), and basic descriptive statistics were computed in SPSS 13.0.

General growth models. Prior to testing our research questions, we assessed the general growth trajectories for LM and non-LM subgroups separately by specifying three-level hierarchical growth models for each outcome in which the four measurement occasions (Level 1) were nested within students (Level 2), which in turn were nested within their last known schools (Level 3). With this specification, we note that classroom membership was ignored because (a) students changed classrooms at least once during the 2-year follow-up period (and many had changed classrooms on multiple occasions), whereas there was less change in school membership; (b) with the small number of students relative to the large number of classroom combinations spread among many schools, our data do not support cross-classified modeling of classrooms (see Classroom Literacy Instruction Observations section); and (c) we reasoned that classroom membership would have less impact on final student outcomes (i.e., dependencies) than last known school (especially for children who moved classrooms but stayed within their school). Any student with at least two follow-up tests was included in our analyses (see Participants section), and for students with less than four completed tests (i.e., missing data), we assumed missingness at random (MAR) (i.e., that student attrition was not related to test scores). HLM employs a full information maximum likelihood approach to estimating

Table 2
Literacy Block Observations for Students With Follow-Up Tests

	Year 1					Year 2				
		LM		Non-LM			LM		Non-LM	
Content category	α	*M*	*SD*	*M*	*SD*	α	*M*	*SD*	*M*	*SD*
Word study focus		25.88	10.70	21.61	10.75		14.41	10.29	11.61	10.08
Phonics	.983	22.02	9.37	18.33	9.86	.993	14.94	12.31	12.68	13.10
Spelling	.953	3.60	4.84	3.29	4.78	.918	1.76	4.24	3.03	5.82
Meaning focus		17.52	9.68	19.01	12.24		27.55	16.26	27.06	15.88
Comprehension	.955	16.82	9.60	19.82	12.07	.980	26.68	16.23	28.48	13.45
Vocabulary	.956	2.38	3.07	2.36	3.35	.990	4.69	6.43	3.14	4.95
Other areas										
Print concepts	—	~~0.65~~	~~1.05~~	~~0.92~~	~~1.49~~	—	~~0.14~~	~~0.41~~	~~0.45~~	~~1.92~~
Phono aware	.969	4.51	4.66	2.52	3.51	.978	0.56	2.10	0.38	1.55
Alphabetics	.988	1.22	2.65	3.54	4.92	.983	0.24	0.93	0.34	1.03
Oral language	.947	5.13	5.67	4.15	5.29	—	~~1.39~~	~~2.82~~	~~1.05~~	~~2.10~~
Fluency building	.996	2.22	4.01	1.41	2.19	.979	1.43	3.24	1.01	3.23
Text reading	.954	20.46	7.96	19.49	10.28	.962	22.36	10.56	23.60	7.86
Writing/language	.898	8.66	8.71	6.69	8.19	.740	4.74	8.16	7.24	7.77
Non-literacy content	.980	28.47	10.15	25.01	11.59	.976	29.09	10.94	26.53	9.17

Note. $N = 62$ LM students and $N = 44$ non-LM students. Observation instrument is Instructional Content Emphasis—Revised. Category means and standard deviations are reported in mean daily minutes observed, computed by averaging across three literacy block observations each year; composite variables (word study and meaning foci) do not directly sum from categories; these were created by first summing the categories at each observation period and then averaging across the sums. Minutes observed for categories in which reliabilities were not established prior to field observations are reported in strikethrough font to emphasize that these estimates should be considered with caution. LM = language minority; non-LM = native English speaking.

model parameters using the covariance structure of all available data. This analytic approach for data with MAR is preferable to casewise deletion (eliminating scores for any student with any missing data) and mean imputation (replacing missing values with the group mean) because (a) both latter methods distort variances and covariances and (b) casewise deletion decreases statistical power (Schafer & Graham, 2002). In addition, Collins, Schafer, and Kam (2001) demonstrated that, in practice, minor violations of MAR have a negligible impact on parameter estimates and standard errors.

Because our emphasis in this paper is on testing effects on students' end-of-study outcomes (rather than change over time), we fixed the intercept at the final measurement occasion (end of second grade for the majority of children) and coded time for each test wave to reflect the average length between assessments (Biesanz, Deeb-Sossa, Papadakis, Bollen, & Curran, 2004). As such, linear change over time was coded as -19 months $=$ first-grade fall follow-up, $-12 =$ first-grade spring, $-7 =$ second-grade fall, $0 =$ second-grade spring. The linear values were squared (the quadratic term allows for one bend in the growth trajectory) to test for quadratic change over time. We purposefully tested both a linear-only model (simpler) and a linear plus quadratic model (more complex; allows for one bend in the growth trajectory) so that we could compare model fit using chi-square likelihood ratio tests (if the more complex model does not fit significantly better than the simpler model, the simpler linear model is retained) (Raudenbush & Bryk, 2002, pp. 60–61). Although we did not expect these growth models to capture all nuances in each student's trajectory, we did wish to use a model that accurately predicted final outcomes. Finally, with no a priori assumptions about student and school variance, we allowed students and schools to vary at random on the intercept and slopes.

Simple treatment effects models. After testing general growth model fit, we retained the best fitting model for each outcome for testing simple treatment differences on the final test score (end of Grade 2). To this end we simply added treatment condition as a predictor (effect coded, $1 =$ treatment, $-1 =$ control) on the intercept (end of second grade score). Again, LM and non-LM subgroups were modeled separately. As for the general growth models, any student who was tested at least twice during the four follow-up test waves was included in these models.

Instructional effects models. For our instructional effects models, we added 10 new covariates to the aforementioned simple treatment effects model, including five main effects (students' kindergarten pretest [fall] alphabetic knowledge, mean daily minutes allocated to meaning- and word study-focused instruction during first grade, mean daily minutes allocated to meaning- and word study-focused instruction during second grade) and five two-way treatment interaction terms. For ease of interpretation (particularly for interactions) and comparability of effects to future studies, the five predictors were standardized (i.e., transformed into z scores) prior to creation of the respective treatment interaction terms (had we kept the pretest covariate in its raw score form or the instructional covariates in minutes, the significance of effects would not change). Finally, although it would have been interesting to test all possible interactions, the sample size in each model precluded us from doing so.

Results

Means and standard deviations at each test wave for LM and non-LM students are presented in Table 3 by outcome measure. Simple zero-order correlations among predictor and end of second-grade outcomes are provided in the Appendix. In the following we report our general growth model results, simple treatment effects model results, and, finally, our instructional model results.

General Growth Models

Using likelihood ratio chi-square tests to compare linear growth models (additive change over time) with quadratic growth models (change that accelerates or decelerates over time) for LM and non-LM students within schools, we found that a linear fit was better suited to word reading, spelling, and comprehension outcomes, whereas a quadratic model was best for fluency. The predicted and observed growth trajectories for each model are compared in Figure 1 (Panel A shows models for LM students; Panel B illustrates models for non-LM students). As can be seen, a drop occurs between Months 12 and 18 (i.e., the period between spring and fall test waves) for many of the outcomes, which is likely due to the lack of instruction students receive during summer break. This drop would indicate that a cubic growth model (i.e., with two points of exponential change over time) may have truly fit each of the outcome models best; however, with only four measurement occasions a cubic model was not possible. Nevertheless, because our focus for this follow-up study is predicting the final measurement occasion (end of Grade 2 for most children), there is confidence that the growth models employed here sufficiently predict the observed final follow-up occasion.

Simple Treatment Effect Model Results

Language minority (LM) students. The model results presented in Table 4 show that there was significant linear growth on spelling (0.25 points per month since end of kindergarten), fluency (0.74 points per month growth), and comprehension (0.56 points per month growth) but not word reading. The quadratic growth term was significant for fluency, indicating an acceleration of 0.02 points per month over time. Translating standard scores at the final measurement occasion (end of Grade 2), LM students averaged near the 50th percentile on word reading, near the 37th percentile on spelling, near the 16th percentile on fluency, and near the 25th percentile on comprehension.

Recall that our first research question asks whether intensive kindergarten phonics-based intervention effects remain 2 years postintervention. These model results, which test the simple effect of treatment ($1 =$ treatment, $-1 =$ control) on the final measurement occasion (2 years postintervention), showed that treatment effects were present for LM students at the end of follow-up (i.e., end of Grade 2 for most children) on word reading (estimated advantage of 5.58 points), spelling (estimated advantage of 5.66 points), and comprehension (estimated advantage of 2.68 points) but not on fluency.

Approximate treatment effect size estimates were calculated as the model-estimated treatment slope coefficient divided by the approximate standard deviation; the approximate standard deviation is computed by dividing the model-implied standard error by

Table 3
Student Assessment Descriptive Statistics

Measure	LM						Non-LM					
	Treatment			Control			Treatment			Control		
	N	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Alphabetics												
Fall, Grade K	34	8.80	8.83	44	6.64	7.94	28	13.57	6.79	31	13.27	9.84
Word reading												
Fall, Grade 1	32	105.34	11.34	44	99.41	10.14	26	110.63	12.99	29	100.12	11.17
Spring, Grade 1	34	108.87	13.13	44	102.40	13.13	27	113.69	14.18	31	105.95	13.11
Fall, Grade 2	34	103.57	12.19	42	97.39	11.56	27	106.39	14.58	29	101.88	14.97
Spring, Grade 2	34	105.63	13.81	43	100.78	13.10	24	108.02	16.15	29	104.26	12.83
Spelling												
Fall, Grade 1	32	93.25	10.30	44	87.43	8.13	26	98.65	12.05	29	88.28	8.96
Spring, Grade 1	34	99.65	15.81	44	93.30	14.09	27	103.11	15.88	31	97.29	13.58
Fall, Grade 2	34	93.85	14.49	42	88.38	11.40	27	100.26	16.05	29	92.69	15.10
Spring, Grade 2	34	100.03	16.31	43	94.09	13.45	24	102.50	18.25	29	97.86	15.45
Fluency												
Fall, Grade 1	32	79.53	5.59	44	78.30	4.93	26	84.42	13.22	29	80.00	7.32
Spring, Grade 1	34	84.56	13.62	44	80.00	10.57	27	92.78	17.39	31	87.42	15.38
Fall, Grade 2	34	83.68	12.57	42	80.24	11.15	27	88.70	19.88	29	87.59	17.30
Spring, Grade 2	34	88.53	15.74	43	87.44	27.74	24	95.83	19.60	29	93.28	21.10
Comprehension												
Fall, Grade 1	32	82.19	4.57	44	80.45	3.55	26	87.12	11.59	29	82.76	7.97
Spring, Grade 1	34	88.82	11.42	44	83.86	8.27	27	94.26	13.85	31	89.84	12.35
Fall, Grade 2	34	87.79	10.74	42	85.71	9.73	27	95.93	14.08	29	94.31	15.16
Spring, Grade 2	34	93.68	12.81	43	91.74	13.13	24	101.88	14.58	29	98.45	16.53

Note. Except for alphabetics, all measures are in standard scores, with a normative mean of 100 and a standard deviation of 15 (alphabetics is in letters correct per minute). LM = language minority; non-LM = native English speaking; Alphabetics = mean of letter names correct per minute and letter sounds correct per minute; Word reading = mean of Woodcock Reading Mastery Test—Revised/Normative Update Word Identification and Word Attack subtests; Spelling = Wide Range Achievement Test—Fourth Edition Spelling subtest; Fluency = Gray Oral Reading Tests (GORT) Passage Fluency composite of Rate and Accuracy subtests; Comprehension = adapted GORT Comprehension subtest.

the square root of the reciprocal number of students in the analysis. (The simple pooled standard deviation from the observed data was not used, as it does not account for the nonindependence of students' scores within schools or for missing data.) We term this predicted effect size estimate as d^* for convenience, which is similar in interpretation to simple Cohen's *d* (Cohen, 1988). For LM students, approximate treatment effect sizes were $d^* = 0.34$ for word reading, 0.36 for spelling, 0.08 for fluency (which was not significant), and 0.28 for comprehension. Hence, the effect sizes for LM students were fairly small, averaging 0.27, or about a third of a standard deviation in the differences between conditions.

Native English-speaking (non-LM) students. The model results presented in Table 5 for non-LM students show that there was significant linear growth on spelling (0.28 points per month since end of kindergarten), fluency (0.62 points per month growth), and comprehension (0.79 points per month growth) but not word reading. The quadratic growth term was not significant for fluency; however there was significant random variation among students in acceleration/deceleration over time. Translating standard scores at the final measurement occasion (end of Grade 2), non-LM students averaged near the 60th percentile on word reading, near the 50th percentile on spelling, near the 37th percentile on fluency, and near the 50th percentile on comprehension.

Importantly, the model results for non-LM students also showed that treatment effects were present at the end of second grade on all outcomes, including fluency. For non-LM students, the estimated treatment advantages were 8.86 points in word reading ($d^* = 0.59$), 9.18 points in spelling ($d^* = 0.75$), 4.80 points in fluency ($d^* = 0.47$), and 3.94 points in comprehension ($d^* = 0.33$). Hence, the approximate effect sizes for non-LM students 2 years after kindergarten intervention were moderate, averaging 0.54, or about a half of a standard deviation in the differences between conditions.

Instructional Model Results

LM students. When we folded in our covariates for testing our second and third research questions (i.e., kindergarten pretest alphabetic knowledge, instructional variables, and treatment × covariate interaction terms), the previously detected simple treatment effect on comprehension for LM students was no longer statistically significant ($p > .05$; results shown in Table 6). Nevertheless, treatment status uniquely predicted the two word-level outcomes (word reading and spelling). Further, holding all other things constant, kindergarten pretest alphabetic knowledge also positively predicted each outcome but did not interact with treatment status (i.e., LM longer term treatment effects did not depend on kindergarten pretest alphabetic knowledge).

Grade 1 word study instruction time positively predicted both word-level outcomes for LM students, irrespective of experimental condition. Holding all other things constant, children who received

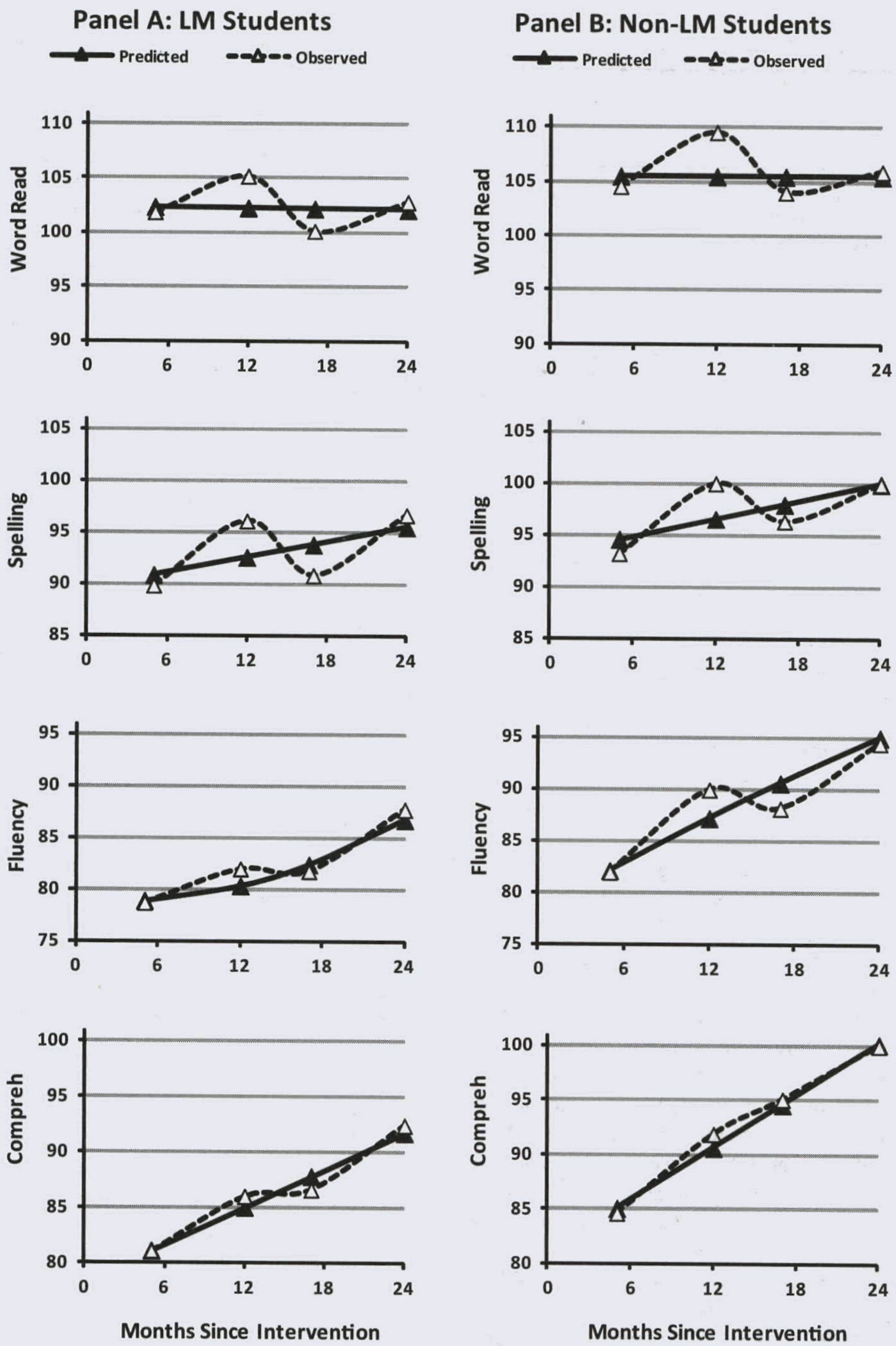

Figure 1. Model-predicted (solid line) and observed (dotted line) change over time. $N = 78$ LM and $N = 59$ non-LM students in growth analyses. LM = language minority; non-LM = native English speaking; Word Read = mean of Woodcock Reading Mastery Test—Revised/Normative Update Word Identification and Word Attack subtests; Spelling = Wide Range Achievement Test—Fourth Edition Spelling subtest; Fluency = Gray Oral Reading Tests (GORT) Passage Fluency composite of Rate and Accuracy subtests; Compreh = adapted GORT Comprehension subtest.

Grade 1 word study instruction time that was one standard deviation more than average (i.e., 10.70 more daily min above average, which was 25.88 min; see Table 2) were predicted to have a 5.31-point advantage on word reading and spelling at the end of Grade 2. Although not significant, a trend for the same pattern of association can be seen for fluency and comprehension as well (see Table 6). For the spelling outcome only, we found that Grade 1 word study moderated the treatment effect: All other things held

Table 4
Simple Treatment Effects Models for LM Students

Fixed effects	Coefficient estimates			
	Word reading	Spelling	Fluency	Comprehension
Intercept: End of Grade 2 mean	102.50*	95.85*	86.85*	91.90*
Linear change/month since fall, Grade 1	−0.02	0.25*	0.74*	0.56*
Quadratic change/month since fall, Grade 1	—	—	0.02*	—
Main effect on intercept				
Treatment condition	2.79*	2.83*	0.36	1.34*

Random effects	Variance estimates			
	Word reading	Spelling	Fluency	Comprehension
Between students				
Intercept: End of Grade 2 mean	164.03*	169.88*	444.02*	128.23*
Linear change	0.16*	0.10	9.70*	0.22*
Quadratic change	—	—	0.01*	—
Between schools				
Intercept: End of Grade 2 mean	0.12	15.01	23.12	0.12
Linear change	0.02	0.03	0.06	<0.01
Quadratic change	—	—	<0.01	—
Residual error	23.55	58.47	24.10	36.62

Note. $N = 78$ LM students in analysis. For all analyses, treatment condition was effect coded (1 = treatment, −1 = control). LM = language minority; Word reading = mean of Woodcock Reading Mastery Test—Revised/Normative Update Word Identification and Word Attack subtests; Spelling = Wide Range Achievement Test—Fourth Edition Spelling subtest; Fluency = Gray Oral Reading Tests (GORT) Passage Fluency composite of Rate and Accuracy subtests; Comprehension = adapted GORT Comprehension subtest.
* $p < .05$, two-tailed.

constant, LM control children who received greater time in Grade 1 word study instruction were expected to outperform treatment students by 3.54 points at the end of Grade 2.

Grade 1 meaning instruction time did not uniquely predict any outcome, nor did Grade 2 word study instruction time. Additionally, neither of these variables moderated treatment effects. However, Grade 2 meaning instruction time positively predicted spelling and fluency: Irrespective of experimental condition, LM children who received one standard deviation more than average time on Grade 2 meaning instruction (16.26 more daily min above average, with average time at 27.55 min; see Table 2) were predicted to score 2.22 and 0.92 more points on second-grade spelling and fluency, respectively. No interaction between Grade 2 meaning instruction and treatment was detected.

Non-LM students. Instructional model results (see Table 7) for non-LM students are in contrast with those observed for LM children. First, treatment effects were still significant on all four end of Grade 2 outcomes, just as they were in the simple treatment effects model (see Table 5). Additionally, unlike for LM students, kindergarten alphabetic knowledge did not uniquely predict outcomes; instead, it negatively interacted with treatment status on word reading and comprehension. Figure 2 displays predicted treatment effect sizes by levels of kindergarten pretest alphabetics knowledge at average and at one standard deviation above average (effect sizes were computed as d^* as an approximate Cohen's d). As shown in the figure, non-LM students who entered the study with somewhat higher alphabetic knowledge (at approximately 22 letters correct per minute) were predicted to have much lower treatment effects than non-LM students who entered the study with average knowledge (at approximately 13 letters correct per minute); the approximate effect size averaged 0.14 for the relatively higher skilled participant and 0.49 for the more typical student in the study. In contrast, LM students' effect sizes were more uniform across kindergarten alphabetics, averaging 0.34 for relatively higher skilled students and 0.26 for the more typical student (recall that there was no significant interaction between pretest and treatment status for LM students).

Again in contrast to the LM model results, for non-LM students, neither Grade 1 nor Grade 2 word study instruction time uniquely predicted second-grade outcomes. (The exception is a small negative effect of Grade 1 word study instruction on spelling: Non-LM students who received more of this instruction were lower by 2.74 points in spelling at the end of second grade, all other things held constant.)

However, Grade 1 and Grade 2 meaning instruction time uniquely contributed to each of the follow-up outcomes but in opposite ways. Grade 1 meaning instruction was positively associated with every outcome: Non-LM students who received one standard deviation more than average time on Grade 1 meaning instruction (12.24 more daily min above average, with average time at 19.01 min; see Table 2) were predicted to score 3.82 more points than average on word reading, 3.49 more points on spelling, 3.15 more points on fluency, and 2.64 more points on comprehension. Students who received higher amounts of Grade 2 meaning instruction (i.e., 15.88 more daily min than the average time of 27.06 min), on the other hand, were expected to have lower second-grade scores: 4.41 fewer points on word reading, 3.81 fewer points on spelling, 4.53 fewer points on fluency, and 4.29 fewer points in comprehension. It is readily noted that the negative

Table 5
Simple Treatment Effects Models for Non-LM Students

Fixed effects	Coefficient estimates: Word reading	Spelling	Fluency	Comprehension
Intercept: End of Grade 2 mean	105.42*	100.06*	95.26*	100.31*
Linear change/month since fall, Grade 1	<0.01	0.28*	0.62*	0.79*
Quadratic change/month since fall, Grade 1	—	—	<0.01	—
Main effect on intercept				
Treatment condition	4.43*	4.59*	2.40*	1.97*

Random effects	Variance estimates: Word reading	Spelling	Fluency	Comprehension
Between students				
Intercept: End of Grade 2 mean	187.39*	239.72*	298.46*	218.02*
Linear change	0.12*	0.11*	0.73*	0.26*
Quadratic change	—	—	0.01*	—
Between schools				
Intercept: End of Grade 2 mean	7.34	2.70	39.18	0.79
Linear change	<0.01	<0.01	0.59	<0.01
Quadratic change	—	—	<0.01	—
Residual error	27.03	42.78	38.79	35.91

Note. $N = 59$ non-LM students in analysis. For all analyses, treatment condition was effect coded (1 = treatment, −1 = control). non-LM = native English speaking; Word reading = mean of Woodcock Reading Mastery Test—Revised/Normative Update Word Identification and Word Attack subtests; Spelling = Wide Range Achievement Test—Fourth Edition Spelling subtest; Fluency = Gray Oral Reading Tests (GORT) Passage Fluency composite of Rate and Accuracy subtests; Comprehension = adapted GORT Comprehension subtest.
* $p < .05$, two-tailed.

effects of Grade 2 meaning instruction are larger in estimated value than Grade 1 meaning instruction's positive effects.

Time received in meaning instruction was also found to moderate treatment effects for word reading, fluency, and comprehension. Students who received more Grade 1 meaning instruction (which had an overall positive effect) were predicted to have increased treatment benefits on fluency (these students were expected to have an additional 3.84-point advantage above and beyond the 4.76-point treatment advantage). Although not significant, this pattern held across the other three outcomes as well. Negative interactions between Grade 2 meaning instruction and treatment status, in contrast, showed that the treatment effect was diminished for three of the four outcomes for children who received greater Grade 2 meaning instruction time. For example, for word reading, the predicted treatment advantage is 1.58 points for children who received one standard deviation higher than average Grade 2 meaning instruction time compared with 9.62 points for average Grade 2 meaning instruction.

Summary of instructional treatment moderators for LM and non-LM students. To put the treatment moderator findings for both samples in perspective, predicted treatment effect sizes for LM (Panel A) and non-LM (Panel B) students are illustrated in Figure 3 by instruction time levels (Grade 1 is shown in the upper row; Grade 2 is shown in the lower row). Across outcomes and across LM and non-LM students, longer term treatment benefits were predicted to be greatest for students who received, in Grade 1, increased meaning instruction time (time on meaning instruction at average). Predicted effect sizes averaged $d^* = 0.43$ for LM and 0.77 for non-LM students. The Grade 1 instructional condition that was least amenable to longer term treatment effects were for students who received high amounts of word study instruction (average predicted effect size was −0.09 for LM students and 0.44 for non-LM students).

Interactions between treatment and Grade 2 instruction time are less clear cut. LM students were predicted to have better longer term treatment effects if their Grade 2 instruction reflected average levels of both word study and meaning instruction time (average predicted effect size across measures was 0.26), whereas non-LM children were predicted to have the greatest longer term treatment effects if their Grade 2 instruction reflected higher amounts of time on word study instruction (average predicted effect size across measures was 0.65).

Discussion

This study extends the current reading intervention literature base several ways: (a) for 2 years after kindergarten, we closely followed students who had been randomized to supplemental kindergarten phonics-based treatment or no-treatment control conditions; (b) the language minority (LM) subgroup comprises a highly diverse array of first languages; (c) the original intervention was delivered by paraeducator instructors rather than teachers or research assistants; (d) observations of classroom literacy blocks for most students were conducted during each of the follow-up years; and (e) attrition rates were minimized such that 93% of the original sample was available for analysis in the simple treatment effects models (73% for instructional models).

Table 6
Instructional Models for LM Students

Fixed effects	Coefficient estimates: Word reading	Spelling	Fluency	Comprehension
Intercept: End of Grade 2 mean	103.70*	96.75*	87.45*	91.61*
Linear change/month since fall, Grade 1	−0.01	0.25*	0.58*	0.51*
Quadratic change/month since fall, Grade 1	—	—	0.01	—
Main effects on intercept				
Treatment condition	1.81*	1.72*	0.31	0.79
Pretest Grade K alphabet (Grade K pretest)	4.27*	2.20*	2.42*	2.51*
Grade 1 instruction, word study (WSI)	5.31*	5.31*	0.43	0.47
Grade 1 instruction, meaning (MI)	1.54	0.89	0.36	0.20
Grade 2 WSI	0.42	0.86	−0.21	−0.50
Grade 2 MI	1.61	2.22*	0.92*	0.66
Treatment moderators on intercept				
Treatment × Grade K Pretest	−0.22	1.33	0.14	0.05
Treatment × Grade 1 WSI	−2.02	−3.49*	0.16	−0.77
Treatment × Grade 1 MI	0.97	1.14	0.17	0.75
Treatment × Grade 2 WSI	−0.27	0.38	−0.96	−0.82
Treatment × Grade 2 MI	−1.42	−1.35	−0.68	−0.72

Random effects	Variance estimates: Word reading	Spelling	Fluency	Comprehension
Between students				
Intercept: End of Grade 2 mean	145.76*	145.92*	510.20*	117.22*
Linear change	0.19*	0.13	11.72*	0.23*
Quadratic change	—	—	0.02*	—
Between schools				
Intercept: End of Grade 2 mean	3.96	5.59	8.11	2.01
Linear change	0.02	0.03	0.06	0.02
Quadratic change	—	—	0.00	—
Residual error	24.40	56.89	28.12	34.63

Note. $N = 62$ LM students in analysis. For all analyses, treatment condition was effect coded (1 = treatment, −1 = control), and continuous predictors were standardized (z scores). LM = language minority; Pretest Grade K = kindergarten pretest alphabet knowledge; Word reading = mean of Woodcock Reading Mastery Test—Revised/Normative Update Word Identification and Word Attack subtests; Spelling = Wide Range Achievement Test—Fourth Edition Spelling subtest; Fluency = Gray Oral Reading Tests (GORT) Passage Fluency composite of Rate and Accuracy subtests; Comprehension = adapted GORT Comprehension subtest.
* $p < .05$, two-tailed.

Simple Treatment Effects

This current study's findings show that kindergarten supplemental phonics intervention continued to have advantages for LM and non-LM students 2 years postintervention. For LM children, the advantages, after controlling for covariates, were for word level outcomes (i.e., word reading and spelling). For non-LM children, advantages were significant for word level, fluency, and comprehension outcomes. Translating the average standard score performance of students into percentile ranks, we observed that non-LM students ended the 2-year follow-up near the 50th percentile on word reading, spelling, and comprehension, whereas LM children neared 50th percentile only on word reading (i.e., standard scores of 100 are 50th percentile). Two years after kindergarten intervention, LM students still lagged behind non-LM peers in comprehension: At posttest of kindergarten in the original study, the disadvantage for LM students was approximately 3 standard score points—but 2 years later, the lag had grown to about 10 standard score points. From the longer term viewpoint, these results suggest that English language proficiency is a necessary foundation for (English) kindergarten phonics instruction to transfer to outcomes beyond those directly targeted by the intervention. The majority of students in this study were from minority and low-income backgrounds associated with more limited language and oral reading interactions. The LM students entered the study with significantly lower vocabulary knowledge, and classroom instruction typically does not close this gap (Biemiller & Slonim, 2001; Bus, van IJzendoorn, & Pellegrini, 1995; Roberts, 2008). Whereas a focus on the decoding component of reading enabled the non-LM students to develop more advanced reading skills, the LM students required more support to catch up in oral language and vocabulary skills that support fluent reading with comprehension.

Alphabetic Knowledge

Kindergarten pretest (late fall) alphabetic knowledge uniquely predicted all longer term outcomes for LM students but not for non-LM students. However, we note that this finding was in the presence of instructional variables. When instructional variables are omitted from our models, early alphabetic knowledge does predict outcomes for both LM and non-LM students, which is consistent with prior research (e.g., McBride-

Table 7
Instructional Models for Non-LM Students

Fixed effects	Coefficient estimates: Word reading	Spelling	Fluency	Comprehension
Intercept: End of Grade 2 mean	105.86*	100.12*	94.89*	101.15*
Linear change/month since fall, Grade 1	0.04	0.28*	0.48	0.79*
Quadratic change/month since fall, Grade 1	—	—	−0.01	—
Main effects on intercept				
Treatment condition	4.81*	4.88*	2.38*	2.18*
Pretest Grade K alphabet (Grade K pretest)	1.31	2.49	0.88	0.99
Grade 1 instruction, word study (WSI)	−2.15	−2.74*	−1.12	−1.78
Grade 1 instruction, meaning (MI)	3.82*	3.49*	3.15*	2.64*
Grade 2 WSI	1.59	1.18	0.85	−0.35
Grade 2 MI	−4.41*	−3.81*	−4.53*	−4.29*
Treatment moderators on intercept				
Treatment × Grade K Pretest	−4.14*	−1.47	−1.65	−2.61*
Treatment × Grade 1 WSI	−1.71	−1.22	1.15	−0.67
Treatment × Grade 1 MI	0.56	0.80	3.84*	1.20
Treatment × Grade 2 WSI	0.55	0.63	1.59	1.11
Treatment × Grade 2 MI	−4.02*	−2.04	−1.93*	−2.89*

Random effects	Variance estimates: Word reading	Spelling	Fluency	Comprehension
Between students				
Intercept: End of Grade 2 mean	160.96*	216.07*	298.52*	204.14*
Linear change	0.14*	0.14*	1.11*	0.29*
Quadratic change	—	—	0.004*	—
Between schools				
Intercept: End of Grade 2 mean	5.79	0.10	104.70	0.13
Linear change	0.02	0.00	1.03	0.00
Quadratic change	—	—	0.00	—
Residual error	26.87	39.62	32.95	37.00

Note. $N = 44$ non-LM students. For all analyses, treatment condition was effect coded (1 = treatment, -1 = control), and continuous predictors were standardized (z scores). non-LM = native English speaking; Pretest Grade K = kindergarten pretest alphabet knowledge; Word reading = mean of Woodcock Reading Mastery Test—Revised/Normative Update Word Identification and Word Attack subtests; Spelling = Wide Range Achievement Test—Fourth Edition Spelling subtest; Fluency = Gray Oral Reading Tests (GORT) Passage Fluency composite of Rate and Accuracy subtests; Comprehension = adapted GORT Comprehension subtest.
* $p < .05$, two-tailed.

Chang, 1999; Muter, Hulme, Snowling, & Stevenson, 2004). Notably, treatment effects for LM students did not depend on pretest alphabetic knowledge, whereas for non-LM students we observed negative treatment interactions: Non-LM students with lower pretests reaped greater treatment benefits on word reading and comprehension (the pattern was also evident albeit not significant for spelling and fluency). However, because LM children started the study lower than non-LM peers on alphabetic knowledge, when taken together, both sets of findings suggest that treatment generally has greater impacts on outcomes for kindergarteners presenting with lower alphabetic knowledge. This replicates earlier research on supplemental phonics intervention (Vadasy & Sanders, 2008, 2010; Vadasy, Sanders, & Peyton, 2006).

Instruction Main Effects

When literacy block time spent on word study- and meaning-focused activities students received was taken into account, the pattern of results was remarkable and differed for LM and non-LM students. For LM students (who were in the bottom half of their kindergarten LM peers), irrespective of experimental condition, we found that increased word study instruction during Grade 1 and increased meaning instruction in Grade 2 had the strongest positive benefits on their second-grade outcomes. It may indeed be the case that the typical pedagogical emphases on word study activities in Grade 1 classrooms and/or meaning-focused activities in Grade 2 classrooms offer similar benefits as the kindergarten intervention for LM students who enter kindergarten in the bottom half of their classrooms. These findings, considered with the earlier end of kindergarten findings, suggest that an effective balance of phonics and meaning in beginning reading instruction for lower skilled students may include sensitive adjustment across the K–2 period (Xue & Meisels, 2004).

For non-LM children (who were in the bottom half of their kindergarten non-LM peers at study onset), irrespective of treatment condition, it was optimal for children to receive relatively more time on meaning instruction during Grade 1 and relatively more time on word study instruction in Grade 2. Given the lack of a consistent interaction between treatment and Grade 1 meaning instruction, time on Grade 1 meaning instruction appeared to add

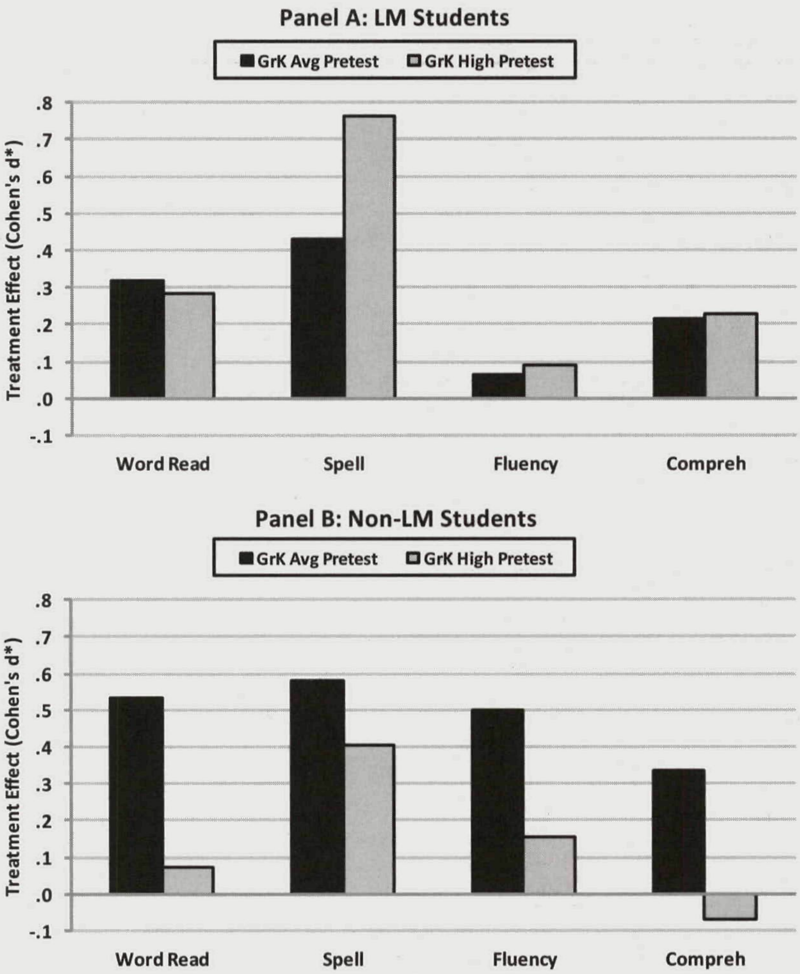

Figure 2. Model-predicted end of Grade 2 treatment effect sizes by kindergarten pretest alphabetic knowledge. $N = 62$ LM and $N = 44$ non-LM students in instructional model analyses. LM = language minority; non-LM = native English speaking; GrK Pretest = kindergarten alphabetic knowledge (Avg = mean and High = $+1SD$); Word Read = mean of Woodcock Reading Mastery Test—Revised/Normative Update Word Identification and Word Attack subtests; Spelling = Wide Range Achievement Test—Fourth Edition Spelling subtest; Fluency = Gray Oral Reading Tests (GORT) Passage Fluency composite of Rate and Accuracy subtests; Compreh = adapted GORT Comprehension subtest.

to, rather than compensate for, kindergarten intervention effects on longer term outcomes. The pattern is not surprising: Reading fluency in particular has been shown to be strongly influenced by knowledge of the meanings of words as well as their lexical representations (see, e.g., Fuchs, Fuchs, Hosp, & Jenkins, 2001), and vocabulary and background knowledge begin to predict reading comprehension in the primary grades (e.g., Muter et al., 2004; Shapiro, 2004; Snow, Tabors, Nicholson, & Kurland, 1995).

This said, Grade 2 meaning instruction time was significantly negatively predictive of outcomes for non-LM children and, although not significant, exhibited the same negative pattern for LM children. It could be case that children who were receiving increased Grade 2 meaning instruction were simultaneously receiving lower amounts of word study instruction. However, the simple correlations among instructional variables show no consistent or significant pattern among Grade 2 word study and Grade 2 meaning instruction times for either language group. It seems more likely that some unmeasured aspect of time on meaning instruction (e.g., quality of instruction or types of embedded instructional activities) is the underlying mechanism for this negative finding.

Treatment Moderator Effects

When we tested the interaction between instructional variables and treatment status, we found that longer term kindergarten treatment impacts for both LM and non-LM students were greater for students who received higher amounts of Grade 1 meaning instruction time. Conversely, treatment impacts were small to nonexistent for students who received higher amounts of Grade 1 word study instruction time. Grade 2 was a different case: Non-LM students reaped greater treatment impacts if their Grade 2 literacy instruction included higher amounts of word study time, but LM students had better longer term treatment

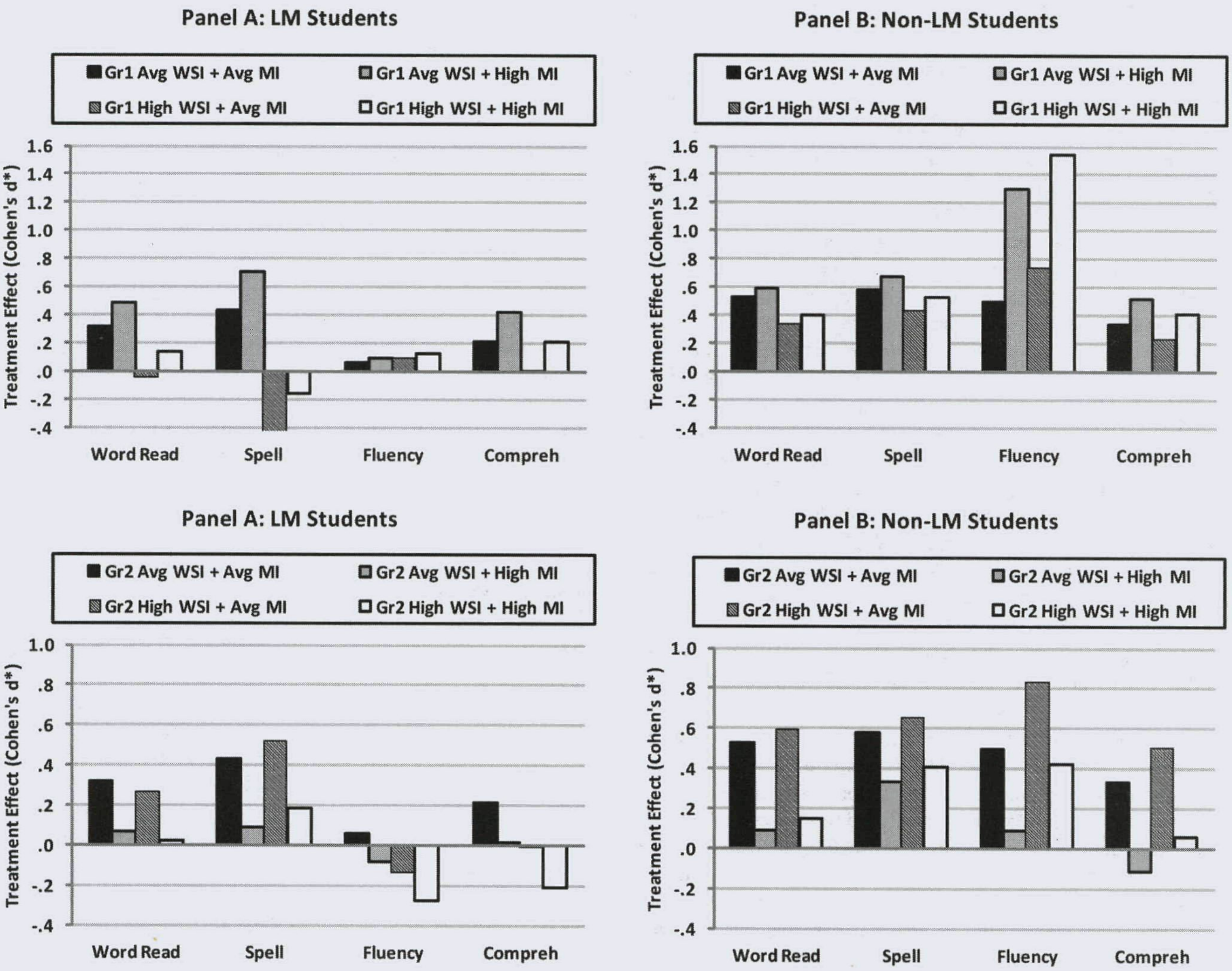

Figure 3. Model-predicted end of Grade 2 treatment effect sizes by instruction time (Grade 1: upper row, Grade 2: lower row). N = 62 LM and N = 44 non-LM students in instructional model analyses. LM = language minority; non-LM = native English speaking; WSI = word study-focused instruction time (daily minutes received, standardized); MI = meaning-focused instruction time (daily minutes received, standardized); for both WSI and MI, Avg = mean and High = +1*SD*; Word Read = mean of Woodcock Reading Mastery Test—Revised/Normative Update Word Identification and Word Attack subtests; Spelling = Wide Range Achievement Test—Fourth Edition Spelling subtest; Fluency = Gray Oral Reading Tests (GORT) Passage Fluency composite of Rate and Accuracy subtests; Compreh = adapted GORT Comprehension subtest.

effects with average amounts of word study and meaning instruction.

Taken together, these finding suggests that lower skilled non-LM students who receive kindergarten intervention may get an additional boost in skills at the end of Grade 2 if they receive (a) added meaning instruction in Grade 1 and (b) an additional dose of word study instruction in Grade 2. For LM students, these findings suggest that kindergarten intervention effects (which were relatively small at the end of kindergarten compared to those for non-LM students) may get an additional boost if they receive added meaning instruction in Grade 1 (but more typical instruction in Grade 2).

Limitations

Findings presented in the current study should be considered in light of their limitations. First, as is true for many studies of language diverse students, we had limited information on initial levels and changes in LM students' English language proficiency (and no measures that allowed us to consider the influence of first language oral or reading skills on second language reading outcomes).

Second, original sample selection procedures were based on a relative risk (performers in the bottom half of their classrooms) rather than absolute risk (cut score) criterion; as such, our sample represents a wider range of students than might be seen in a typical intervention study.

A third limitation concerns inferences that may be drawn from the classroom observation analyses. We observed teachers only for schools in which we had human subjects' approval; hence, our sample of observations reflects in some part instructional minutes only for students who have relatively more stable family characteristics (i.e., remaining in schools within the two original districts). Additionally, sample sizes for LM and non-LM subgroups were too small to allow us to test all possible interactions among instructional variables; as such, cross-grade combinations of Grade 1 and Grade 2 meaning and word study effects were not tested and may be a point of exploration for future research. In a similar vein, our sample size precluded us from using cross-classified models to test instructional effects at the classroom level; as such, our instruction effects may be somewhat overstated (we tested effects at the student level). Further, our observations did not capture amount of time spent on incidental literacy instruction occurring outside the regular literacy block or the quality of reading instruction, and they did not include the level of detail that would allow us to contextualize findings with respect to how teachers individualized their reading instruction, particularly for LM students' English language and comprehension skills. Finally, it is important to note that treatment condition was the only randomized variable in this study—not kindergarten pretest or time on classroom instruction. As such, no causal conclusions may be drawn about relationships among instructional variables and student outcomes; instead, only associative links should be inferred. For example, the positive relationship between Grade 1 word study instruction and LM students' longer term outcomes could be due to student skills that were improved as a result of instruction or, conversely, to teachers spending more time on this area to help their students become more successful. A final limitation is that we did not assess students' home literacy environment, which some have found associated with later literacy development (Frijters, Barron, & Brunello, 2000; Hood, Conlon, & Andrews, 2008). However, we have no reason to believe that variation in home literacy environment interacted with treatment, as (a) students were randomly assigned to treatment conditions within classrooms, (b) students likely had the same home literacy environment before study onset as after, and (c) the treatment was a highly structured, word-level intervention that is unlikely to have been mirrored at home or to overlap with learning in the intervention.

Instructional Implications

Several implications can be drawn from the simple treatment effects models (see Tables 4 and 5). A supplemental and explicit phonics intervention that targeted kindergarten students performing in the lowest half of their classrooms in alphabetic and phonological skills had sustained treatment effects across most reading outcomes for both LM and non-LM students 2 years postintervention. Findings add support for benefits of explicit instruction in early reading skills, provided by trained paraeducator tutors, for both LM and non-LM students. Findings suggest that the benefits of providing systematic phonics instruction in kindergarten also extend to young English learners for word-level outcomes in particular (Ehri et al., 2001).

The kindergarten supplemental phonics instruction in this study was not at all coordinated with subsequent classroom reading instruction, which varied in the intensity of word and meaning instruction emphases across the follow-up period. Treatment effects for LM students were significant for longer term word reading and spelling outcomes, whereas fluency and comprehension levels for LM students lagged behind those for non-LM peers. For both LM and non-LM students, treatment effects were observed for word-level outcomes, and these effects were bolstered by increased Grade 1 meaning instruction (although LM students demonstrated no significant simple treatment effects on longer term fluency or comprehension outcomes). This corroborates previous findings that LM students derive similar benefits as non-LM students from supplemental tutoring in early word-level skills, although to a lesser extent. Findings underscore that supplementing instruction in foundation word reading skills did not prevent LM students from falling behind in comprehension. End of Grade 2 comprehension outcomes indicate that more intensive early language comprehension intervention may be necessary for some LM students.

The findings from the current study are consistent with those from the small number of earlier follow-up studies of early reading interventions for LM students. In particular, Gunn et al. (2000, 2002, 2005) reported moderate treatment effects for decoding and oral reading fluency at follow-up, and Cirino et al. (2009) and Vaughn et al. (2008) reported treatment effects for word reading, word reading fluency, comprehension, and spelling at 1-year and 3- to 4-year follow-up. In comparison to these studies, the current study shows that LM treatment students maintained significant advantage over controls for word reading and spelling but not fluency or comprehension (in the presence of instructional variables). Differences in outcomes among the earlier supplemental early reading interventions for LM children Grades K–3 may reflect variations in instructional emphases, group sizes, and instructional intensity that warrant further study.

References

August, D., & Shanahan, T. (Eds.). (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth.* Mahwah, NJ: Erlbaum.

August, D., & Shanahan, T. (2010). Response to a review and update on *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth. Journal of Literacy Research, 42,* 341–348. doi:10.1080/1086296X.2010.503745

Aunola, K., Nurmi, J.-E., Niemi, P., Lerkkanen, M.-K., & Rasku-Puttonen, H. (2002). Developmental dynamics of achievement strategies, reading performance, and parental beliefs. *Reading Research Quarterly, 37,* 310–327. doi:10.1598/RRQ.37.3.3

Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology, 93,* 498–520. doi:10.1037/0022-0663.93.3.498

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods, 9,* 30–52. doi:10.1037/1082-989X.9.1.30

Bishop, D. V. M., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry, 31,* 1027–1050. doi:10.1111/j.1469-7610.1990.tb00844.x

Boscardin, C. K., Muthen, B., Francis, D. J., & Baker, E. L. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories. *Journal of Educational Psychology, 100,* 192–208. doi: 10.1037/0022-0663.100.1.192

Bus, A. G., & van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91,* 403–414. doi:10.1037/0022-0663.91.3.403

Bus, A. G., van IJzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research, 65,* 1–21.

Carlisle, J. F., & Beeman, M. M. (2000). The effects of language of instruction on the reading and writing achievement of first-grade Hispanic children. *Scientific Studies of Reading, 4,* 331–353. doi:10.1207/S1532799XSSR0404_5

Carlisle, J. F., Beeman, M. M., Davis, L. H., & Spharim, G. (1999). Relationship of metalinguistic capabilities and reading achievement for children who are becoming bilingual. *Applied Psycholinguistics, 20,* 459–478. doi:10.1017/S0142716499004014

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairment. *Journal of Speech, Language, and Hearing Research, 45,* 1142–1157. doi:10.1044/1092-4388(2002/093)

Chiappe, P., & Siegel, L. S. (1999). Phonological awareness and reading acquisition in English- and Punjabi-speaking children. *Journal of Educational Psychology, 91,* 20–28. doi:10.1037/0022-0663.91.1.20

Chiappe, P., Siegel, L. S., & Wade-Wooley, L. (2002). Linguistic diversity and development of reading skills: A longitudinal study. *Scientific Studies of Reading, 6,* 369–400. doi:10.1207/S1532799XSSR0604_04

Cirino, P. T., Vaughn, S., Linan-Thompson, S., Cardenas-Hagan, E., Fletcher, J. M., & Francis, D. J. (2009). One-year follow-up outcomes of Spanish and English interventions for English language learners at risk for reading problems. *American Educational Research Journal, 46,* 744–781. doi:10.3102/0002831208330214

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing-data procedures. *Psychological Methods, 6,* 330–351. doi:10.1037/1082-989X.6.4.330

Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007, January 26). Algorithm-guided individualized reading instruction. *Science, 315,* 464–465. doi:10.1126/science.1134513

Connor, C. M., Morrison, F. J., & Katch, E. L. (2004). Beyond the reading wars: Exploring the effect of child-instruction interactions on growth in early reading. *Scientific Studies of Reading, 8,* 305–336. doi:10.1207/s1532799xssr0804_1

Connor, C. M., Morrison, F. J., & Petrella, J. (2004). Effective reading comprehension instruction: Examining Child × Instruction interactions. *Journal of Educational Psychology, 96,* 682–698. doi:10.1037/0022-0663.96.4.682

Denton, C. A., Anthony, J. L., Parker, R., & Hasbrouck, J. E. (2004). Effects of two tutoring programs on the English reading development of Spanish–English bilingual students. *Elementary School Journal, 104,* 289–305. doi:10.1086/499754

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Reading Research Quarterly, 38,* 78–103. doi:10.1598/RRQ.38.1.4

Edmonds, M., & Briggs, K. L. (2003). The Instructional Content Emphasis Instrument: Observations of reading instruction. In S. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: Systems for the observation of teaching and learning* (pp. 31–47). Baltimore, MD: Brookes.

Ehri, L. (1987). Effects of printed language acquisition on speech. In D. Olson, N. Torrence, & A. Hildyard (Eds.), *Literacy, language, and learning* (pp. 333–367). Cambridge, MA: Cambridge University Press.

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

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research, 71,* 393–447. doi:10.3102/00346543071003393

Foorman, B. R., Francis, D. J., Fletcher, J. M., & Schatschneider, C. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90,* 37–55. doi:10.1037/0022-0663.90.1.37

Foorman, B. R., Francis, D. J., Shaywitz, S. E., Shaywitz, B. A., & Fletcher, J. M. (1997). The case for early reading intervention. In B. Blachman (Ed.), *Foundations of reading acquisition and dyslexia* (pp. 243–264). Mahwah, NJ: Erlbaum.

Foorman, B. R., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research & Practice, 16,* 203–212. doi:10.1111/0938-8982.00020

Frijters, J. C., Barron, R. W., & Brunello, M. (2000). Direct and mediated influences of home literacy and literacy interest on prereaders' oral vocabulary and early written language skill. *Journal of Educational Psychology, 92,* 466–477. doi:10.1037/0022-0663.92.3.466

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5,* 239–256. doi: 10.1207/S1532799XSSR0503_3

Goswami, U., Gombert, J. E., & de Barrera, L. F. (1998). Children's orthographic representations and linguistic transparency: Nonsense word reading in English, French, and Spanish. *Applied Psycholinguistics, 19,* 19–52. doi:10.1017/S0142716400010560

Gunn, B., Biglan, A., Smolkowski, K., & Ary, D. (2000). The efficacy of supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. *Journal of Special Education, 34,* 90–103. doi:10.1177/002246690003400204

Gunn, B., Smolkowski, K., Biglan, A., & Black, C. (2002). Supplemental instruction in decoding skills for Hispanic and non-Hispanic students in early elementary school. *Journal of Special Education, 36,* 69–79. doi:10.1177/00224669020360020201

Gunn, B., Smolkowski, K., Biglan, A., Black, C., & Blair, J. (2005). Fostering the development of reading skill through supplemental instruction: Results for Hispanic and non-Hispanic students. *Journal of Special Education, 39,* 66–85. doi:10.1177/00224669050390020301

Hood, M., Conlon, E., & Andrews, G. (2008). Preschool home literacy practices and children's literacy development: A longitudinal analysis. *Journal of Educational Psychology, 100,* 252–271. doi:10.1037/0022-0663.100.2.252

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80,* 437–447. doi:10.1037/0022-0663.80.4.437

Juel, C. (1996). What makes literacy tutoring effective? *Reading Research Quarterly, 31,* 268–289. doi:10.1598/RRQ.31.3.3

Kaplan, D., & Walpole, S. (2005). A stage-sequential model of reading transitions: Evidence from the Early Childhood Longitudinal Study. *Journal of Educational Psychology, 97,* 551–563. doi:10.1037/0022-0663.97.4.551

Keenan, J. M., & Betjemann, R. S. (2006). Comprehending the Gray Oral Reading Test without reading it: Why comprehension tests should not include passage-independent items. *Scientific Studies of Reading, 10,* 363–380. doi:10.1207/s1532799xssr1004_2

Landerl, K., Wimmer, H., & Frith, U. (1997). The impact of orthographic consistency on dyslexia: A German–English comparison. *Cognition, 63,* 315–334. doi:10.1016/S0010-0277(97)00005-X

Lesaux, N. K., & Kieffer, M. J. (2010). Exploring sources of reading comprehension difficulties among language minority learners and their classmates in early adolescence. *American Educational Research Journal, 47,* 596–632. doi:10.3102/0002831209355469

Lesaux, N. K., Koda, K., Siegel, L. S., & Shanahan, T. (2006). Development of literacy. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth* (pp. 75–122). Mahwah, NJ: Erlbaum.

Lesaux, N. K., & Siegel, L. S. (2003). The development of reading in children who speak English as a second language. *Developmental Psychology, 39,* 1005–1019. doi:10.1037/0012-1649.39.6.1005

McBride-Chang, C. (1999). The ABCs of the ABC: The development of letter-name and letter-sound knowledge. *Merrill-Palmer Quarterly, 45,* 285–308.

Morrison, F. J., Connor, C. M., & Bachman, H. J. (2006). The transition to school. In D. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy* (Vol. 2, pp. 375–394). New York, NY: Guilford Press.

Muter, V., Hulme, C., Snowling, M., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40,* 665–681. doi:10.1037/0012-1649.40.5.665

National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs. (2006). The growing numbers of limited English proficient students 1994/95–2005/06. Retrieved from http://www.ncela.gwu.edu/files/uploads/4/GrowingLEP_0506.pdf

National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication 00–4754). Washington, DC: National Institute of Child Health and Development.

Parrila, R., Aunola, K., Leskinen, E., Nurmi, J.-E., & Kirby, J. R. (2005). Development of individual differences in reading: Results from longitudinal studies in English and Finnish. *Journal of Educational Psychology, 97,* 299–319. doi:10.1037/0022-0663.97.3.299

Parrila, R., Kirby, J. R., & McQuarrie, L. (2004). Articulation rate, naming speed, verbal short-term memory, and phonological awareness: Longitudinal predictors of early reading development? *Scientific Studies of Reading, 8,* 3–26. doi:10.1207/s1532799xssr0801_2

Phillips, L. M., Norris, S. P., Osmond, W. C., & Maynard, A. M. (2002). Relative reading achievement: A longitudinal study of 187 children from first through sixth grades. *Journal of Educational Psychology, 94,* 3–13. doi:10.1037/0022-0663.94.1.3

Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal, 45,* 365–397. doi: 10.3102/0002831207308230

Pierce, M. E., Katzir, T., Wolf, M., & Noam, G. G. (2007). Clusters of second and third grade dysfluent urban readers. *Reading and Writing, 20,* 885–907. doi:10.1007/s11145-007-9058-x

Proctor, C. P., Carlo, M., August, D., & Snow, C. (2005). Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology, 97,* 246–256. doi:10.1037/0022-0663.97.2.246

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

Raudenbush, S. W., Bryk, A. S., & Congdon, R. T. (2004). HLM 6 for Windows [Computer software]. Lincolnwood, IL: Scientific Software International.

Roberts, T. A. (2008). Home storybook reading in primary or second language with preschool children: Evidence of equal effectiveness for second-language vocabulary acquisition. *Reading Research Quarterly, 43,* 103–130. doi:10.1598/RRQ.43.2.1

Rueda, R., & Windmueller, M. P. (2006). English language learners, LD, and overrepresentation: A multiple-level analysis. *Journal of Learning Disabilities, 39,* 99–107. doi:10.1177/00222194060390020801

Scarborough, H. S. (1989). Prediction of reading dysfunction from familiar and individual differences. *Journal of Educational Psychology, 81,* 101–108. doi:10.1037/0022-0663.81.1.101

Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities. In B. K. Shapiro, P. J. Accadro, & A. J. Capute (Eds.), *Specific reading disability: A view of the spectrum* (pp. 75–120). Timonium, MD: York Press.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177. doi:10.1037/1082-989X.7.2.147

Shanahan, T., & Beck, I. (2006). Effective literacy teaching for English-language learners. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language Minority Children and Youth* (pp. 415–488). Mahwah, NJ: Erlbaum.

Shankweiler, D., & Liberman, I. Y. (1989). *Phonology and reading disability: Solving the reading puzzle.* Ann Arbor: University of Michigan Press.

Shapiro, A. M. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal, 41,* 159–189. doi:10.3102/00028312041001159

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55,* 151–218. doi:10.1016/0010-0277(94)00645-2

Share, D. L., Jorm, A. F., Maclean, R., & Matthews, R. (1984). Sources of individual differences in reading acquisition. *Journal of Educational Psychology, 76,* 1309–1324. doi:10.1037/0022-0663.76.6.1309

Snow, C. E., Tabors, P. O., Nicholson, P. E., & Kurland, B. F. (1995). SHELL: Oral language and early literacy in kindergarten and first-grade children. *Journal of Research in Childhood Education, 10,* 37–48. doi:10.1080/02568549509594686

Snowling, M. J. (1991). Developmental reading disorders. *Journal of Child Psychology and Psychiatry, 32,* 49–77. doi:10.1111/j.1469-7610.1991.tb00003.x

Sonnenschein, S., Stapleton, L. M., & Benson, A. (2010). The relation between the type and amount of instruction and growth in children's reading competencies. *American Educational Research Journal, 47,* 358–389. doi:10.3102/0002831209349215

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360–407. doi:10.1598/RRQ.21.4.1

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K. S., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities, 34,* 33–58. doi:10.1177/002221940103400104

Torgesen, J. K., & Burgess, S. R. (1998). Consistency of reading-related phonological processes throughout early childhood: Evidence from longitudinal–correlational and instructional studies. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 161–188). Mahwah, NJ: Erlbaum.

Tunmer, W. E., Herriman, M. L., & Nesdale, A. R. (1988). Metalinguistic abilities and beginning reading. *Reading Research Quarterly, 23,* 134–158. doi:10.2307/747799

Vadasy, P. F., & Sanders, E. A. (2008). Code-oriented instruction for kindergarten students at risk for reading difficulties: A replication and comparison of instructional grouping. *Reading and Writing, 21,* 929–963. doi:10.1007/s11145-008-9119-9

Vadasy, P. F., & Sanders, E. A. (2010). Efficacy of supplemental phonics instruction for low-skilled kindergarteners in the context of language-minority status and classroom phonics instruction. *Journal of Educational Psychology, 102,* 786–803. doi:10.1037/a0019639

Vadasy, P. F., Sanders, E. A., & Abbott, R. D. (2008). Effects of supplemental early reading intervention at 2-year follow up: Reading skill growth patterns and predictors. *Scientific Studies of Reading, 12,* 51–89. doi:10.1080/10888430701746906

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trial with paraeducator implementers. *Journal of Educational Psychology, 98,* 508–528. doi:10.1037/0022-0663.98.3.508

Vaughn, S., Cirino, P. T., Tolar, T., Fletcher, J. M., Cardenas-Hagan, E., Carlson, C. D., & Francis, D. J. (2008). Long-term follow-up of Spanish and English interventions for first-grade English language learners at risk for reading problems. *Journal of Research on Educational Effectiveness, 1,* 179–214. doi:10.1080/19345740802114749

Verhoeven, L. (1990). Acquisition of reading in a second language. *Reading Research Quarterly, 25,* 90–114. doi:10.2307/747596

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., . . . Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33,* 468–479. doi:10.1037/0012-1649.33.3.468

Wiederholt, J. L., & Bryant, B. R. (2001). *Gray Oral Reading Tests* (4th ed.). Austin, TX: Pro-Ed.

Wilkinson, G. S., & Robertson, G. J. (2006). *The Wide Range Achievement Test—fourth edition.* Lutz, FL: Psychological Assessment Resources.

Woodcock, R. (1987/1998). *Woodcock Reading Mastery Test—Revised/Normative Update.* Circle Pines, MN: American Guidance Service.

Xue, Y., & Meisels, S. J. (2004). Early literacy instruction and learning in kindergarten: Evidence from the Early Childhood Longitudinal Study—Kindergarten class of 1998–1999. *American Educational Research Journal, 41,* 191–229. doi:10.3102/00028312041001191

Appendix

Zero-Order Correlations Between Predictors and Final Outcomes, Unadjusted for School Membership

Variable	1	2	3	4	5	6	7	8	9	10
Predictors										
1. Treatment	—	−.01	.00	.12	−.09	.16	.18	.17	.03	.13
2. Grade K pretest	.08	—	−.14	.21	−.18	−.04	.26	.36*	.34*	.35*
3. Grade 1 WSI	.12	.04	—	−.16	−.08	−.09	.03	−.08	−.08	−.06
4. Grade 1 MI	.14	.04	−.03	—	−.01	.25	.14	.19	.12	.07
5. Grade 2 WSI	−.10	−.13	.02	−.10	—	.14	.12	.05	−.04	−.14
6. Grade 2 MI	.28*	.34*	−.14	.17	−.05	—	−.15	−.11	−.07	−.02
End of Grade 2 outcomes										
7. Word reading	.17	.43*	.11	.05	.08	.28*	—	.91*	.82*	.74*
8. Spelling	.13	.46*	.07	−.05	−.02	.32**	.86*	—	.75*	.68*
9. Fluency	−.01	.26*	.07	.05	.08	.06	.42*	.52*	—	.86*
10. Comprehension	.06	.43*	.15	.04	.10	.15	.72*	.72*	.66*	—

Note. $N = 62$ LM and $N = 44$ non-LM students represented above (students used in instructional model analyses). LM correlations shown in lower diagonal; non-LM correlations shown in upper diagonal. LM = language minority; non-LM = native English speaking; Grade K pretest = kindergarten alphabetic knowledge; MI = meaning-focused instruction time (daily minutes received); WSI = word study-focused instruction time (daily minutes received); Word reading = mean of Woodcock Reading Mastery Test—Revised/Normative Update Word Identification and Word Attack subtests; Spelling = Wide Range Achievement Test—Fourth Edition Spelling subtest; Fluency = Gray Oral Reading Tests (GORT) Passage Fluency composite of Rate and Accuracy subtests; Comprehension = adapted GORT Comprehension subtest.
* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

Received January 27, 2011
Revision received January 27, 2012
Accepted March 16, 2012 ■