Journal of Educational Psychology
2010, Vol. 102, No. 4, 786–803

0022-0663/10/$12.00 DOI: 10.1037/a0019639

Efficacy of Supplemental Phonics-Based Instruction for Low-Skilled Kindergarteners in the Context of Language Minority Status and Classroom Phonics Instruction

Patricia F. Vadasy and Elizabeth A. Sanders
Washington Research Institute

This study tested the efficacy of supplemental phonics instruction for 84 low-skilled language minority (LM) kindergarteners and 64 non-LM kindergarteners at 10 urban public schools. Paraeducators were trained to provide the 18-week (January–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 their language 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 for those in higher phonics classrooms. Finally, when we examined LM students alone, we found that pretest English receptive vocabulary positively predicted most posttests and interacted with treatment only on phonological awareness. In general, pretest vocabulary did not moderate kindergarten LM treatment response.

Keywords: English language learner, intervention, paraeducators, randomized trial, multilevel modeling

Supplemental materials: http://dx.doi.org/10.1037/a0019639.supp

The benefits of explicit code-oriented instruction for children at risk for reading difficulties are well established (e.g., National Institute of Child Health and Human Development, 2000; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001). A sound body of research on carefully conducted classroom instruction (e.g., Blachman et al., 2004; Byrne & Fielding-Barnsley, 1991; Foorman et al., 1997; Hatcher, Hulme, & Ellis, 1994; Torgesen et al., 1999) and supplemental programs (Elbaum, Vaughn, Hughes, & Moody, 2000; Hatcher et al., 2006; Savage & Carless, 2005) is available to guide practitioners in the components of effective instruction. Yet because language minority (LM) students have often been excluded from these studies, it is not clear whether these instructional practices are equally effective for the increasing numbers of LM students whose home language is not English (Lesaux, 2006). In their recent meta-analysis on the effects of explicit instruction in components of literacy for LM learners, Shanahan and Beck (2006) identified only 17 studies that met inclusion criteria.

Instructional decisions for LM students are further complicated by the significant diversity and variation in their English language proficiency and how this influences student response to reading interventions. Others (August & Erickson, 2006; August & Hakuta, 1998; Lesaux, 2006; Shanahan & Beck, 2006) have outlined the need for research on specific instructional practices and their interaction with LM student characteristics and instructional contexts. Teachers use varied approaches in their whole-class literacy instruction that may influence LM student literacy outcomes (Foorman et al., 2006; Juel & Minden-Cupp, 2000). Considerable variation may characterize literacy instruction in classrooms that include LM students, due to differences in teacher preparation, student languages, and bilingual supports (McCardle & Chhabra, 2006). In the present study, we had three goals. First, we examined the benefits of supplemental phonics-based instruction for LM kindergarten students at risk for reading difficulties. Second, we considered the influence of classroom time afforded to phonics instruction on student reading outcomes. Third, we investigated whether English vocabulary knowledge influences treatment response for LM students in particular. As suggested by the lexical restructuring model (Metsala & Walley, 1998), early vocabulary growth promotes increasingly segmental representations of spoken words. LM students' initial vocabulary knowledge may

This article was published Online First October 11, 2010.

Patricia F. Vadasy and Elizabeth A. Sanders, Washington Research Institute, Seattle, WA.

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 for coordinating the intervention and assessments and Kathryn Compton for project assistance. 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

therefore influence development of phonemic awareness and early word reading skills.

Instruction of LM Students in Public Schools

The U.S. Census term *Limited English Proficient* (LEP) is often used in reports we reference to describe an individual whose primary language is not English and who has limited English skills. In this study, similar to others (Lesaux & Siegel, 2003; Lipka & Siegel, 2007), children who are LM were defined as students whose parents report that they speak a language other than English at home. Schools across the United States serve a growing number of LM students. Between 1979 and 2007, the number of LM students in the United States increased from 3.8 to 10.8 million, or from 9% to 12% of the school-age populations (Planty et al., 2009). Over 44% of these students are enrolled in pre-K through Grade 3. Students who are LM often come from families with limited income and parent education: Sixty-eight percent of LEP children pre-K to Grade 5 were low income in 2000, and almost half of LEP elementary children have parents with less than high school education (Capps et al., 2005). Young children from low-income and nonnative-English-speaking backgrounds are considered at higher risk for reading difficulties (Snow, Burns, & Griffin, 1998), and social inequalities in academic outcomes increase as children advance through the elementary grades (Entwisle, Alexander, & Olson, 1997; Kieffer, 2008; Phillips, Crouse, & Ralph, 1998). The gap between what is recommended as effective literacy instruction for elementary LM students and what schools are able to provide is often considerable. For example, in a practice guide that summarized research on effective educational practices for LM students (Gersten et al., 2007), recommendations for improving reading achievement and language development include formative assessments in early reading skills, intensive small-group interventions, extensive and high-quality vocabulary instruction, instruction in academic English, and structured peer-assisted reading and language activities. These resources are limited in many schools. For example, when teachers were surveyed in California, a state serving nearly one third of LM students in the country, they reported the need for more resources for second-language reading and writing instruction, including increased paraprofessional assistance (Gandara, Maxwell-Jolly, & Driscoll, 2005).

Because districts serve LM students from an average of eight different native language backgrounds (Hopstock & Stephenson, 2003), schools often have difficulty recruiting staff to provide reading instruction in the home language of students (Goldenberg, 2008). Efforts to address the oral language proficiency of LM students must often be balanced with English literacy instruction (Saunders, Foorman, & Carlson, 2006), which is a challenge for the most experienced and skilled teachers. In a study of the influence of teacher characteristics on the oral language and literacy outcomes of kindergarten bilingual students, teacher quality was significantly related to the use of instructional time and to end-of-year student performance in language and literacy skills (Cirino, Pollard-Durodola, Foorman, Carlson, & Francis, 2007). In the present study we examined the time afforded to literacy content in students' classrooms, including the time allocated to code-oriented early literacy skills.

Effective Interventions for LM Students

The No Child Left Behind Act (NCLB, 2001) requires states to be accountable for the achievement of LM students and to provide that this group of students meets state proficiency levels in academic standards by the 2013–2014 school year (Rossell, 2005). In order to meet this challenge for reading standards, schools need information on effective literacy interventions for LM students. Research on efficacy of specific literacy programs and interventions for LM students remains limited (Lesaux, 2006), and challenges that researchers must address include nonrandom assignment to services, student mobility and dropout, and variability in definition of LEP and former-LEP status (Hopstock, 2003).

Two recent reviews examine the efficacy of specific literacy interventions for LM students. The U. S. Department of Education's What Works Clearinghouse (2007) issued a report on reading interventions tested for their efficacy with LM students. The report assigned effectiveness ratings to 10 interventions, ranging from explicit reading instruction that emphasized decoding (*Read Well* and *Reading Mastery*) to reading comprehension and vocabulary interventions. Overall, eight programs were rated as potentially positive for use with LM students. In another review, Shanahan and Beck (2006) examined 17 studies of literacy interventions for LM students. They found that these studies yielded findings similar to those reported for native speakers, although effect sizes observed for LM speakers were generally smaller than for native speakers. Shanahan and Beck concluded that the research base was too limited for them to draw conclusions for policy and practice.

Interventions for at-risk kindergarten students commonly focus on decoding and word reading skills, which are strong predictors of reading success for native speakers. The most common obstacle for children with reading disabilities is accurate and fluent word recognition (Rack, Snowling, & Olson, 1992; Stanovich, 1980, 1988). Research from intervention studies clearly supports the benefits of systematic phonics instruction to develop accurate and fluent word reading skills, in particular for students at risk for reading difficulties due to limited home literacy or phonological awareness skills (see Rayner et al., 2001).

The effectiveness of systematic phonics instruction for beginning non-LM readers was examined in a meta-analysis by Ehri, Nunes, Stahl, and Willows (2001). The meta-analysis included experimental studies only, and effect sizes were calculated on outcomes reported in 38 studies. For the kindergarten and first-grade studies, the average effect size for phonics instruction was .96 for decoding regular words, .70 for decoding pseudowords, .57 for reading miscellaneous words, and .79 for spelling words. In order to compare these effects for non-LM and LM students, we consider the two most carefully conducted reviews of literacy interventions for LM students. In Shanahan and Beck's (2006) review, four studies evaluated the efficacy of phonics instruction. The first, an evaluation of 2 years of supplemental *Reading Mastery* instruction for Hispanic students Grades K–3 reading below grade level (Gunn, Biglan, Smolkowski, & Ary, 2000), reported significant effects for Hispanic students in Word Attack and Word Identification at the end of 2 years of intervention. A follow-up study (Gunn, Smolkowski, Biglan, & Black, 2002) found that Hispanic students who did not speak English at the start of the intervention experienced benefits similar to those for the other students, although these findings are qualified due to lack of power. Only two studies in Shanahan and Beck's review reported spelling outcomes, and only one of these was published in a peer-reviewed journal. In a British study (Stuart, 1999), LM learners 4 and 5 years of age were provided with 12 weeks of either *Jolly Phonics* or a *Big Book* shared oral reading approach with enlarged storybooks. Children in the phonics group were significantly higher at posttest in reading nonwords and in spelling (overall treatment effect of .46), and at 1-year follow up, children in the *Jolly Phonics* group continued to significantly outperform the *Big Book* group in spelling and reading accuracy.

In the What Works Clearinghouse review of literacy interventions for LM students (U.S. Department of Education, 2007), three studies that included phonics instruction reported outcomes for Word Attack and Word Identification. The first study was the randomized controlled trial of the *Reading Mastery* program for Spanish-speaking students in Grades 1 or 2 (Gunn et al., 2000). Small-group supplemental instruction in *Reading Mastery* included phonemic awareness, alphabetic, and decoding. At the end of 2 years of intervention, effect sizes for letter–word identification and Word Attack were .55 and .70, respectively. The second quasi-experimental study evaluated the effects of 1 year of the Success for All program for Hispanic LM students in kindergarten and first grade (Chambers, Slavin, Madden, Cheung, & Gifford, 2004). Intervention effect sizes were .26 for Word Identification and .45 for Word Attack. Finally, the third study (Denton, Anthony, Parker, & Hasbrouck, 2004) evaluated *Read Well* tutoring for low-skilled LM students in Grades 2–5. Effect sizes were .40 for Word Identification and .35 for Word Attack.

These two reviews allow us to draw tentative conclusions on phonics-based interventions for young English-language learners. The studies by Gunn et al. (2000, 2002) and Stuart (1999) suggest that systematic instruction in phonics produces positive outcomes in word reading, although smaller effect sizes compared to similar interventions for non-LM students. The single finding on spelling outcomes in Stuart (1999) provides limited support for intervention transfer to encoding skills.

Research Questions

The current study tested the efficacy of a code-oriented reading intervention with low-skilled LM and non-LM kindergarteners in the context of time spent on classroom phonics instruction. We randomly assigned LM and non-LM kindergarteners performing in the bottom half of their classrooms to experience either a carefully described supplemental phonics intervention or their regular classroom literacy instruction for half the school year. The intervention used in the current study is delivered one-to-one by paraeducator tutors and was evaluated previously with three samples of primarily non-LM kindergarten students at risk for reading difficulties (Vadasy & Sanders, 2008a, 2008b; Vadasy, Sanders, & Peyton, 2006a). In the first study (Vadasy et al., 2006a), kindergarten students who averaged in the lowest 13th percentile in pretest phonemic and alphabetic skills were randomly assigned either to their regular classroom literacy instruction (business as usual) or to 18 weeks of individual supplemental tutoring in phonics-based skills. At posttest, students who received treatment made significantly more growth than controls in phonemic segmentation and nonsense word fluency and averaged higher than their peers in the control condition (in the 45th percentile in reading accuracy and the 32nd percentile in reading efficiency). Tutored students also maintained significantly higher levels of growth at follow-up at the end of first grade. The subsequent two studies (Vadasy & Sanders, 2008a, 2008b) replicated these results.

These findings align with prior research on effective instructional practices to develop early literacy skills in low-performing primary-age students (e.g., Hatcher et al., 1994; Schneider, Roth, & Ennemoser, 2000). In the current study, we recruited both LM and non-LM students so that we could directly compare each subgroup's responsiveness to treatment. Further, we explicitly incorporated time spent on classroom phonics instruction into our models so that we could test for interactions with treatment response. Our research questions are as follows.

1. What are the effects of phonics intervention and LM status on kindergarten student outcomes? Are treatment effects qualified by LM status?

2. Does treatment response depend on classroom time afforded to phonics instruction?

3. For LMs, does pretest receptive (English) vocabulary moderate treatment response?

Method

Participants

Initial sample. In October of 2007–2008, all students in full-day kindergarten classrooms at 12 urban public elementary schools known for relatively large proportions of language minority (LM) student enrollment were invited to participate in our research study. For this study, we defined a student as LM if the student's parent reported the primary home language as other than English on the student's school registration record. (This definition is consistent with that of August and Shanahan, 2006.) Otherwise, the student was defined as non-LM. Students receiving extra services from the school or district, such as special education, Title I (a federally funded program for schools with large concentrations of low-income students), or bilingual services, were not excluded from participation.

Of the 827 students invited, 479 had assenting consents returned (243 of whom were LM) and 34 (12 of whom were LM) had parents who declined their participation. Thus, our consent return rate was 62%, which is in alignment with the mean 66% active consent rate found across 124 published school-based intervention studies (Blom-Hoffman et al., 2009). All invitations and consent forms were sent home in English, and for the top 10 most frequent languages in the district we also sent translated invitations and consent forms. (Nevertheless, we recognize that some parents may not be literate in their home language.) In our initial sample, there were at least 28 languages represented (some were other African languages that were not specified); the top five most frequent languages in our initial sample of LM students, in rank order, were Spanish (49% of LMs), Vietnamese (15%), Somali (11%), Chinese (6%), and Tagalog (3%). This distribution is similar to the distribution of the top languages spoken by LM students who attend U.S. public schools (Center for Public Education, 2007).

Screening. Each participating classroom was required to have sufficient student sample sizes for random assignment to experimental groups (within LM/non-LM subgroups within classrooms). This required removing 10 classrooms (and thus two schools) from study participation prior to screening due to low numbers of LM students. Further, some students had moved from their schools or were persistently absent during screening. As such, 317 students (n = 178, 56% of whom were LM) were actually screened. Screening occurred in November, and assessments were administered individually by trained testers. The screen included three measures. The first two, the number of letter sounds and letter names produced out of 52 randomly ordered uppercase English letters (Fuchs et al., 2001), were measures of alphabetic knowledge. The third was a test of phonological awareness (Sound Matching subtest from the Comprehensive Test of Phonological Processing; Wagner, Torgesen, & Rashotte, 1999). Both alphabetic knowledge (Chall, 1967; Share, Jorm, Maclean, & Matthews, 1984; Tunmer, Herriman, & Nesdale, 1988) and phonological awareness (Adams, 1990; Bus & van IJzendoorn, 1999; Shankweiler & Liberman, 1989; Share, 1995; Snowling, 1991; Stanovich, 2000) are well-established early predictors of reading outcomes in monolingual English speakers, and phonological awareness and alphabetic skills are related to English word reading in LM children (Chiappe, Siegel, & Wade-Wooley, 2002; Lesaux, Koda, Siegel, & Shanahan, 2006; Lipka & Siegel, 2007).

Assignment to experimental conditions. We identified LM and non-LM students in the lower half of their classrooms for random assignment to experimental conditions. First, we separated students by classroom. Then, we separated LM and non-LM students within each classroom. Next, we computed a composite z score for each student, based on the mean z score of each of the three screening measures already described (all z scores were computed within LM/non-LM subgroup, within classroom). Students were then rank ordered, within LM/non-LM subgroup within classroom, from lowest to highest; students in the upper half of their classroom's LM or non-LM group were then removed from further study participation. Finally, students in the lower half of their classroom's LM or non-LM group were randomly assigned to treatment (supplemental tutoring) or control (regular classroom instruction, no tutoring) conditions.

Final sample. Student attrition included nine students who moved (five treatment and four controls), two students removed due to severe behavior difficulties during pretesting (one in each condition), and 10 treatment students (three LM and seven non-LM) who had been randomly selected out of the study due to a lack of enough tutoring time slots at their school (this occurred only at larger schools). After attrition, the final sample included 67 treatment students (38 of whom were LM) and 81 controls (46 LMs) from 24 classrooms across 10 schools. These schools had student enrollments averaging 85% minority, 75% free or reduced lunch, 33% bilingual, and 18% special education during the intervention year. All of the schools were designated Title I. Table 1 summarizes observed student characteristics for each condition by LM/non-LM group; chi-square tests of independence showed no evidence that treatment and control conditions differed on any student characteristic (all p values $>$ .10).

Paraeducators. All paraeducator tutors (paraeducators) were recruited from their school communities based on their interest in working with children, prior tutoring and school volunteer experience, and scheduling flexibility. The 23 participating paraeducators were mostly nonminority (74%) and female (83%) and varied in their age, educational levels, general tutoring experience, and experience working with kindergartners. Tutors averaged a mode of 35–44 years old (ranging from 18 to over 55 years) and ranged in educational level from high school diploma to master's degree, with a modal level of a bachelor's degree (39%). (The average education level of paraeducators in this study is similar to that recommended under NCLB for supplemental education services.) Prior to the study, paraeducator tutoring experience ranged from 0 (22%) to 10 or more years (26%), with a mean of 4.52 years (SD = 5.06). Most paraeducators (72%) had at least one year previous experience working with early grade levels (K–2; M = 3.22 years, SD = 4.04, range = 0 to 15 years). All paraeducators were hired as district employees and were paid by the schools with funds provided by the research grant. The assignment of students to tutors was wholly determined by a combination of classroom scheduling, paraeducator availability, and the number of eligible students within classrooms within sites.

Table 1
Student Characteristics

	Treatment (n = 67)				Control (n = 81)			
	LM (n = 38)		Non-LM (n = 29)		LM (n = 46)		Non-LM (n = 35)	
Characteristic	N	%	N	%	N	%	N	%
Male	19	50	14	48	28	61	21	60
Bilingual	37	97	0	0	43	93	0	0
FRL	35	92	21	72	43	93	22	63
SPED	0	0	1	3	1	2	3	9
Minority	38	100	21	72	46	100	22	63
Asian	9	24	4	19	11	24	6	27
Black	7	18	14	67	6	13	7	32
Hispanic	20	53	3	14	29	63	7	32
Mixed/other	2	5	0	0	0	0	2	9

Note. N = 148 students from 24 classrooms and 10 schools; LM = language minority; Bilingual = receives bilingual services; FRL = eligible for free or reduced-price lunch; SPED = receives Special Education services.

Intervention

Students assigned to the treatment condition received individual systematic and explicit phonics instruction, including letter–sound correspondences, phonemic decoding, spelling, and assisted oral reading practice in decodable texts. Each paraeducator was provided with a set of 70 scripted lessons in a three-ring binder (with 7–8 activities per lesson) that were matched to decodable texts for oral reading practice (see online supplemental materials). Tutors and students shared and worked from the same lesson pages. All tutoring was conducted during the school day, outside the classroom in a quiet nearby school space. In a typical tutoring session, paraeducators spent 20 min on phonics activities and 10 min scaffolding students' oral reading practice in decodable texts. Although each phonics lesson was designed to be completed within 20 min, paraeducators adjusted the rate of progress through the lessons to meet students' needs. Instruction occurred 30 min a day, 4 days per week, over a period of 18 weeks (January–May). Lesson activities included the following.

1. Letter–sound correspondence. Letters were introduced at a rate of about one new letter every two lessons. Both letter names and letter sounds were explicitly taught and practiced. Students practiced by pointing to the letters, saying the sound, and writing the letters that matched the sounds spoken by the instructor. Lessons 1–27 included practice identifying 12 letters per lesson, and later lessons included practice in 16 letters per lesson. Instruction design featured cumulative review of all letters and added review on vowel sounds. When a student needed more practice to learn letter names, and particularly letter sounds, the paraeducator and the student practiced with a letter-sound card for a few extra minutes each lesson. During this practice, the student pointed and matched the letter name/sound to the printed letter and pictured key word on the card, in a procedure described by Berninger (1998). To assist tutors working with LM students, coaches provided tutors with information on letter sounds that do not occur in Spanish.

2. Segmenting. Students learned to segment two-part compound words, two-syllable words, two-phoneme nonwords, three-phoneme words, and, finally, four-phoneme words with consonant blends. Paraeducators modeled each item and then orally presented four items for the student to segment using Elkonin boxes (squares drawn on a piece of paper with one square for each speech segment). Students repeated each speech item, pointed to each box as they spoke the syllable or phoneme, and then swept their finger under the boxes and said the word fast.

3. Word reading and spelling. In the first 20 lessons the paraeducator modeled phoneme blending: pointing to the sample word in each lesson, stretching out the sounds without stopping between phonemes, and then saying the word fast. Students then orally blended six words per lesson, with scaffolding and assistance. Words for decoding and spelling were composed of letters that had already been introduced. Students received added practice on weak letter sounds if needed by identifying the sound in the initial, final, and medial position in a spoken word. The paraeducator dictated three words for the student to spell (words including the new sound, a difficult sound, and ending with an easy word) and provided explicit instruction in how to map letters to phonemes. Students repeated each word before they attempted to spell it, learned to segment each word into phonemes, and reread each word they spelled. Tutors directed students to a handwriting chart with numbered arrows to guide letter strokes and form letters efficiently. Students also fingerpoint-read short sentences constructed with previously taught words. At Lesson 33 students learned to read and spell words with plurals. Coaches encouraged tutors, when needed, to reinforce successful decoding attempts by LM students who may not recognize the words they successfully sounded out.

4. Irregular word instruction. Beginning in Lesson 16, the tutor introduced high-frequency irregular words that appeared in the decodable texts. The paraeducator read the word and the student pointed to the word, spelled it aloud, and read the word again. One irregular word was introduced every one to two lessons, with ongoing cumulative review of previously introduced words.

5. Phoneme blending. To add practice in recognizing orally blended words, the paraeducator asked the student to guess the word (i.e., say it fast) that the tutor said in a slow, stretched out way (i.e., without stopping between phonemes, just as the student was learning to do in the word-reading activity). This practice was included in Lessons 1–20.

6. Alphabet naming practice. The paraeducator asked the student, based on the student's level of alphabetic knowledge, to do one of the following activities: (a) say the alphabet (letter names) while pointing to the letters on the letter sound card; (b) say the alphabet (letter names) without looking at the letters/chart; (c) point to the letters that the tutor names; or, (d) name the letters to which the tutor points.

7. Assisted oral reading practice. During the last 10 min of each session, beginning in Lesson 9, students practiced reading aloud in the decodable *Bob Books* (Maslen, 2003), which were matched to the lessons for their instructional consistency (Hoffman, Sailors, & Patterson, 2002; Mesmer, 2004). Students read each book twice the first time it was introduced and reread other books if time was available. Most students read independently, with tutor assistance, and coaches helped tutors offer more reading support when needed. In these cases the students read the story with the paraeducator (partner reading), or reread a line of text after the tutor read the same line (echo reading). Supplemental decodable titles were provided when students needed more texts for reading practice.

Instructional scaffolding. Research staff were teachers or experienced tutors with backgrounds in reading instruction. They provided ongoing coaching and modeling of appropriate scaffolding to help paraeducators provide the type of support at-risk students often require to accomplish phonemic segmenting, decoding, and encoding tasks (Foorman & Torgesen, 2001; Juel, 1996). Tutors were instructed that, before students moved beyond Lesson 10 (when modeling of the phonemic decoding task decreases), each student should demonstrate at least 70% mastery of all letter sounds introduced as well as an understanding of phoneme decoding (although not necessarily at full mastery). Tutors who worked with LM students were instructed to provide judicious incidental vocabulary instruction without compromising the intensity of the phonics instructional time.

Intervention coverage. Throughout the intervention, each tutor recorded their students' daily attendance (tutoring sessions) and lesson coverage (intervention lessons completed). By the end of intervention, treatment students received a mean of 55.36 tutoring sessions (or $M = 27.68$ hr of tutoring, $SD = 2.74$ hr), completed a mean of 64.99 lessons ($SD = 13.69$), and had a mean lesson coverage rate of 1.18 lessons per session ($SD = 0.25$).

Paraeducator training. Researchers provided an initial 2-hr training session to describe each lesson activity and model paraeducator/student behaviors, errors, and error correction strategies. Trainees were paired together to practice each activity while trainers provided feedback and responded to questions. Follow-up training was provided throughout the intervention, with added coaching for paraeducators with less experience and/or low initial intervention fidelity ratings. Less experienced tutors received from 0.5 to 3.0 hr of coaching during the intervention, averaging 1 hr of additional on site coaching. All coaches also conducted fidelity observations, described below.

Treatment fidelity. Six research staff were trained to conduct on-site fidelity observations of paraeducators with their assigned students. Fidelity observations involved a 5-point rating scale ranging from 1 (*never implements correctly*) to 5 (*always implements correctly*) for each of the instructional components. After training but prior to field observations, researcher staff viewed six videotaped tutoring sessions of paraeducators implementing instruction with students; students in the videotaped sessions were selected to represent a range of reading skills we expected to see in the field; their ages ranged from 4 to 7 years. To determine interrater reliability prior to onsite treatment fidelity observations, we calculated the internal consistency of the fidelity observers' mean implementation ratings for the videotaped sessions (using observers' ratings as items and each videotape as a subject): Cronbach's alpha was .97. After reliability was established, researchers conducted a total of 156 fidelity observations for the 23 paraeducators over the course of the intervention, averaging 6.78 observations per tutor. Fidelity ratings had a mean of 4.41 (SD = 0.57).

Student Assessments

Abilities hypothesized to contribute to or correlate with early word reading skills were assessed at screening (November) or pretest (December) and included measures of receptive vocabulary, alphabetic knowledge, phonological awareness, word reading, and spelling. Receptive vocabulary was assessed at pretest to investigate its potential influence on treatment response for LM learners. Students were posttested at the end of the intervention on measures of alphabetic knowledge, phonological awareness, word reading, spelling, passage reading fluency, and comprehension (students are not expected to be able to perform on the latter two measures at the beginning of kindergarten, and these measures were therefore not administered at pretest).

Tests were individually administered by testers who were unaware of student group assignment. Testers were primarily former teachers with testing experience in similar intervention research. They were trained and supervised by research staff to administer assessments according to protocols. Training included explaining, modeling, and supervised independent practice on each measure. In the measure descriptions that follow, published reliabilities for each measure, as well as sample reliabilities (internal consistencies reported are Cronbach's alpha), are provided for kindergarten-age students. Standard scores were used when available (exceptions are alphabetics, spelling, and passage reading fluency).

1. *Receptive vocabulary* was measured at pretest only with the norm-referenced Peabody Picture Vocabulary Test IIIA (Dunn & Dunn, 1997). Students select a picture that best illustrates the meaning of an orally presented stimulus word. Testing is discontinued after the student misses eight out of 12 items. The raw score (number of items correct) is converted to a standard score. Reliability reported in the test manual is .94 for 5-year-olds. For this sample, internal consistency was .97 for all students and .95 for LM students.

2. *Alphabetic knowledge* was measured at screening and posttest as the mean of two naming measures: letter names and letter sounds correctly produced in 1 min. We developed our own measures of alphabetics; however, they are highly similar to the letter name task in Fuchs et al. (2001) and the letter name fluency subtest from the Dynamic Indicators of Basic Early Literacy Skills (Good & Kaminski, 2002). Both our letter names and letter sounds measures use all 26 letters of the alphabet twice (once in uppercase and once in lowercase). Letters are randomly sorted (upper- and lowercase together) and presented on a single page in six rows in Comic Sans font, which allows students to better differentiate between the lowercase letter *l* and uppercase letter *I* in particular. For the letter names task, students are asked to name as many letters on the page as they can. For the letter sounds task, students are asked to produce the sounds that each letter on the page represents. The number correctly named or produced, respectively, is divided by the number of seconds the students took to finish the 52 items. This value is then multiplied by 60 to obtain letter names or sounds correct per minute. For letter sounds, we considered only hard consonants and short vowels as correct. For our sample, internal consistency at pretest was .98 (for all students and LM learners only) and at posttest was .96 and .97 for letter names and sounds, respectively (for LM learners, .97 for both).

3. *Phonological awareness* was measured at screening/pretest and posttest with the composite standard score of three subtests from the Comprehensive Test of Phonological Processing (Wagner et al., 1999): Blending Words (pretest), Elision (pretest), and Sound Matching (screening). During the Blending Words subtest, the student is asked to listen to parts of words and blend them together to make a whole word. This subtest has 20 items, and testing is discontinued after the student misses three items in a row or when the student completes all items. Test–retest reliability reported in the manual for 5- to 7-year-olds is .88. For this sample, internal consistency at pretest was .78 and at posttest was .83 (correspondingly, .74 and .82 for LMs). During the Elision subtest, the student is asked to listen to the sounds in a spoken word and is then asked to say the word without one or more of its sounds, creating a new word (e.g., the student is asked, "Say the word *spider* without saying *der*"). This subtest has 20 items, and testing is discontinued after the student misses three items in a row or when all items are completed. Test–retest reliability reported in the test manual for 5- to 7-year-olds is .88. For this sample, internal consistency at pretest was .79 and at posttest was .83 (for LM learners, we found .65 and .80 at pretest and posttest, respectively). The Sound Matching subtest has two parts: In Part I, the tester says a word and asks the student to say, out of three word choices, the word that starts with the same sound as the initial word (e.g., the student is asked, "Which word starts with the same sound as *sock*? *Sun*, *cake*, or *bear*?"). Part II of this subtest asks the student to say, out of three word choices, the word that ends with the same sound as the initial word. This subtest has 20 items, and testing is discontinued after the student gets four out of seven items incorrect. Test–retest reliability reported in the manual for 5- to 7-year-olds is .83. For our sample, internal consistency was .71 at screening (.66 for LMs) and was .90 (.88 for LMs) at posttest.

4. *Word reading* was measured at pretest and posttest with 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, 1998). The Word Attack subtest includes 50 nonwords that increase in difficulty. Testing is discontinued after six consecutive incorrect responses. Split-half reliability for first graders reported in the manual is .94. For this sample, internal consistency at posttest was .91 (.86 for LM learners). Sample reliability could not be computed at pretest due to extreme floor effects. The Word Identification subtest consists of 106 words that increase in difficulty. Testing is discontinued after six consecutive incorrect responses. Split-half reliability reported in the manual for first graders is .98. For this

sample, internal consistency at pretest was .88 (.85 for LM learners) and at posttest was .95 (.93 for LM learners).

5. *Spelling* was assessed at pretest and posttest with developmental raw scores of words correctly spelled on the Wide Range Achievement Test—Revised (WRAT–R; Jastak & Wilkinson, 1984) Spelling subtest. This test requires the student to copy marks or symbols, print his or her name, and print a list of dictated words. Testing is discontinued after 10 consecutive incorrect responses. Test–retest reliability reported for ages 7.0–7.5 is .97. As did Fuchs et al. (2001), however, we applied the Tangel and Blachman (1992) developmental scoring rubric to all words attempted (within normal test administration guidelines). This rubric allowed us to credit students for partial and less phonemically sophisticated responses. Items were scored from 0 (random string of letters) to 6 (entire word correctly spelled). Internal consistency for this sample was .89 and .96 at pretest and posttest, respectively. Reliability for LM students was similar: .85 at pretest and .95 at posttest.

6. *Passage reading fluency* was assessed at posttest only using the mean words correctly read in 1 min on two grade-level story passages. Both stories are from the *Primary Phonics* series (Makar, 1996). One passage was considered more decodable ("Mac and Tab"), and the other was considered less decodable ("Ben Bug"). Students read each passage aloud for one minute. Words omitted, substituted, and hesitations of more than three seconds are scored as errors. Words self-corrected within three seconds were scored as accurate. The score is calculated as the number of words correctly read in one minute. For our kindergarten sample, internal consistency was .97 for both the more and less decodable passages (for all students as well as LM learners only). The correlation between the two passages (which may be considered as a form of alternate-form reliability) across all students was .91 (for LM learners, .90).

7. *Comprehension* was assessed at posttest only with the standard score of the WRMT–R/NU Passage Comprehension subtest. The student is asked to silently read a short passage and then orally provide a missing key word. (For each blank, the student is asked to supply a word appropriate in the context of the passage.) Acceptable responses are listed on the examiner's easel page, and testing is discontinued after six consecutive incorrect responses. Internal consistencies reported by test developers range from .94 to .97. For this sample, internal consistency was .86 (for LM learners, .70).

Classroom Literacy Instruction Observations

We hypothesized that a specific aspect of classroom literacy instruction would interact with treatment condition: time spent on phonics/word study instruction. For example, teachers who spend more time on phonics may bolster control students' reading and spelling skills in the same manner as that expected for treatment students. As such, treatment may benefit only students whose teachers spend less time on phonics instruction. Further, this may be more or less the case depending on LM status. We therefore endeavored to quantify features of classroom literacy instruction for all students in the intervention study. All 24 kindergarten classroom teachers granted permission for us to observe their literacy block instruction on three occasions (approximately one month apart at the end of January, March, and May).

Observation tool. We used an adapted version of the Instructional Content Emphasis—Revised (ICE–R; Edmonds & Briggs, 2003) for measuring time afforded to dimensions of classroom literacy instruction. The standard ICE–R includes the following four dimensions: "main instructional category," "instructional subcategory," "grouping arrangement," and "materials used." We simplified the measure because establishing interobserver reliability across all dimensions would be enormously challenging and because instructional content was our primary interest. We therefore collected data on two dimensions only: main instructional category and grouping arrangement.

The main instructional category measures instructional time spent on 10 mutually exclusive literacy activities: print concepts, phonological awareness, alphabetic knowledge, word-study/phonics, spelling, oral language development, fluency building, text reading, comprehension, and writing/language arts. After reviewing sample classroom reading instruction videotapes, we identified the need for two additional codes. First, a "vocabulary" code was designed to capture all types of vocabulary instruction, including direct instruction in word meanings, strategy instruction in deriving word meanings from context, and vocabulary instruction embedded in text reading. Second, an "other" instructional activity code was designed to capture behavior management, evaluative feedback, transition time, and other types of nonliteracy instruction (i.e., math instruction). The grouping arrangement dimension, which overlaps with the main instructional category activities, measured the amount of time the teacher grouped students for instruction in whole class, small group, or other arrangements (pair, independent, and individualized).

Recording process. Observations were conducted for the entire duration of teachers' literacy blocks. Observers began timing at the beginning of the observation. When 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. Time entries and codes were entered into a database, and the time spent in minutes on each code was computed automatically for each teacher for each observation.

Establishing reliability. Seven certificated teachers were trained to conduct observations of classroom literacy blocks. Each classroom observer studied the ICE–R manual and coding instructions and participated in several training sessions to review the measure. Classroom observers were trained by the first author, and observers used the ICE–R to code eight videotapes of literacy instruction in kindergarten and first-grade classrooms to establish reliability (prior to onsite observations). Reliabilities (internal consistencies) were calculated for each content category and for each grouping arrangement. To calculate reliabilities, we treated each videotape as an observation and each classroom observer's computed time (in minutes) per instructional category as an item. In other words, for "phonics/word study," eight subjects (videotapes) and seven items (observers' times) were used to compute a reliability for phonics/word study. Estimated classroom observation reliabilities (Cronbach's alpha) ranged from .74 (phonological awareness) to .99 (phonics/word study) for content categories (Mdn = .95 and M = .93) and .99 for all grouping arrangement categories. We note that, for five content categories, the high interrater reliabilities are due in part to zero inflation (right skew);

these categories included concepts of print, alphabetic knowledge, spelling, fluency building, and writing/language arts.

Results

Analytic Strategy

We adopted multilevel modeling as our primary analytic tool (also known as hierarchical linear modeling and random effects modeling) for testing hypotheses for data collected in this study. As compared with unilevel methods (e.g., analysis of variance or multiple linear regression), the more complex analysis method accounts for dependencies among student scores due to classrooms and schools, allowing for valid inferences to be drawn about relationships between outcomes and predictors without violating the assumption of independence.

Classroom growth models. Prior to being able to answer our second research question (as well as to simply describe kindergarten classroom literacy instruction practices), we analyzed our classroom observation data using three-level growth models in which observation occasions (January, March, and May; Level 1, $n = 72$) were nested within classroom teachers (Level 2, $n = 24$), which in turn were nested within schools (Level 3, $n = 10$). Time was coded as 0, 1, and 2, corresponding to each observation such that the model intercept would represent minutes in January and the slope would represent linear change per observation period. In all models, we estimated random effects for both the intercept and slope (i.e., we hypothesized that teachers would vary in both the initial amount of time afforded to instructional categories and the rate at which they changed instruction over the latter half of the kindergarten year). The general mixed-model equation for these classroom growth models is thus as follows:

$$\text{Minutes}_{tij} = \gamma_{000} + \gamma_{100}(\text{Time}_{tij}) + U_{0ij} + U_{00j} + \text{Time}(U_{1ij}) + \text{Time}(U_{10j}) + r_{tij}.$$

In the model above, the number of minutes allocated to a particular instructional activity at time t for classroom teacher i in school j is equal to the intercept γ_{000} (mean minutes spent on instruction in January), plus the fixed effect of slope γ_{100} (linear change in minutes spent on the instruction over the latter half of the kindergarten year), plus the residuals between classroom teachers and schools in January (U_{0ij} and U_{00j}, respectively) plus the residuals between classroom teachers and schools on linear change over time (Time $\times$ U_{1ij} and Time $\times$ U_{10j}, respectively), plus the residual error between predicted and observed scores at each observation occasion, r_{tij}.

Pretest models. For student pretests, we employed three-level models in which student scores (Level 1, $n = 148$) were nested within classrooms (Level 2, $n = 24$), which were in turn nested within schools (Level 3, $n = 10$). The t test of the treatment effect slope (as well as LM and interaction slopes) on a given outcome in this framework is nearly identical to the classical t and F tests, except that variance associated with the nesting structure (i.e., variance between classrooms and schools) is explicitly estimated and accounted for in the predicted values and residual errors. In all analyses, treatment condition (Txt) and LM status were effect coded (i.e., 1 = treatment, −1 = control; 1 = LM, −1 = non-LM) and then were multiplied together to create the Txt $\times$ LM interaction term. The general mixed-model equation for our pretest models is as follows:

$$\text{Score}_{ijk} = \gamma_{000} + \gamma_{100}(\text{Txt}_{ijk}) + \gamma_{200}(\text{LM}_{ijk}) + \gamma_{300}(\text{Txt})(\text{LM}_{ijk}) + U_{0jk} + U_{00k} + r_{ijk}.$$

In the model above, the pretest score for student i in classroom j in school k is equal to the intercept γ_{000} (mean score across all students, holding treatment and LM status constant); plus the fixed effect of slope γ_{100} (treatment effect); plus the fixed effect of slope γ_{200} (LM effect); plus the fixed effect of slope γ_{300} (Treatment $\times$ LM interaction); plus the residual between the student's classroom mean and the grand mean of classrooms, U_{0jk}; plus the residual between the student's school mean and the grand mean of schools, U_{00k}; plus the student residual, r_{ijk}.

Posttest models. To test our first and second research questions, we employed a three-level model that included tests of treatment status, language status, and classroom phonics time (Phon), as well as corresponding interactions. For ease of interpretation, we grand-mean centered (i.e., standardized) classroom phonics time. The general mixed-model equation was as follows:

$$\begin{aligned}\text{Posttest}_{ijk} = {} & \gamma_{000} + \gamma_{100}(\text{Txt}_{ijk}) + \gamma_{200}(\text{LM}_{ijk}) + \gamma_{300}(\text{Txt})(\text{LM}_{ijk}) \\ & + \gamma_{010}(\text{Phon}_{jk}) + \gamma_{110}(\text{Phon})(\text{Txt}_{ijk}) \\ & + \gamma_{210}(\text{Phon})(\text{LM}_{ijk}) + \gamma_{310}(\text{Phon})(\text{Txt})(\text{LM}_{ijk}) \\ & + U_{0jk} + U_{00k} + r_{ijk}.\end{aligned}$$

In the model above, the posttest score for student i in classroom j in school k is equal to the intercept γ_{000} (mean posttest across all students, holding treatment, LM, and phonics time constant); plus the treatment effect (slope γ_{100}); the LM effect (slope γ_{200}); the interaction between treatment and LM (slope γ_{300}); the effect of classroom phonics time and corresponding interactions (slopes in standard deviations, γ_{010}–γ_{310}); and classroom, school, and unaccounted-for residuals on posttests (U_{0jk}, U_{00k}, and r_{ijk}, respectively).

Finally, to answer our third research question (whether initial vocabulary moderated treatment response for LM learners in particular), we employed our posttest model described above but with two changes: We took out language status as a predictor, because this model was applied only to LM students' data, and we added pretest receptive vocabulary (grand-mean centered and denoted VOC) and corresponding interactions. Thus, to test our third research question, we used the following model.

$$\begin{aligned}\text{Posttest}_{ijk} = {} & \gamma_{000} + \gamma_{100}(\text{Txt}_{ijk}) + \gamma_{200}(\text{VOC}_{ijk}) + \gamma_{300}(\text{VOC})(\text{Txt}_{ijk}) \\ & + \gamma_{010}(\text{Phon}_{jk}) + \gamma_{110}(\text{Phon})(\text{Txt}_{ijk}) \\ & + \gamma_{210}(\text{Phon})(\text{OC}_{ijk}) + \gamma_{310}(\text{Phon})(\text{Txt})(\text{OC}_{ijk}) \\ & + U_{0jk} + U_{00k} + r_{ijk}.\end{aligned}$$

In the model above, the posttest score for LM student i in classroom j in school k is equal to the intercept γ_{000} (mean posttest across all students, holding treatment, vocabulary, and classroom phonics time constant), plus the treatment effect (slope γ_{100}); the vocabulary effect (slope in standard deviations, γ_{200}); the interaction between treatment and vocabulary (slope γ_{300}); the effect of classroom phonics time and corresponding interactions (slopes in standard deviations, γ_{010}–γ_{310}); and classroom, school, and unaccounted-for residuals on posttests (U_{0jk}, U_{00k}, and r_{ijk}, respectively).

We note that *HLM* was used for all multilevel analyses (Raudenbush, Bryk, & Congdon, 2004); *SPSS* (1989–2004) was used to compute all other statistics.

Classroom Observation Growth Model Results

Table 2 shows teachers' time spent on each activity by observation period. On average, teachers spent the majority of their literacy blocks on "other," noncontent instruction (30%–34% of the average time), phonics/word study (18%–22% of the average time), and text reading (13%–18% of the average time). Teachers spent the least amount of their literacy block time on print concepts (<1% for all observation occasions) and fluency (<1%–2% of the average time). Finally, teachers primarily grouped students for whole-class instruction (50%–64% of their literacy block time) compared with other grouping arrangements.

In order to test whether teachers were systematically changing their instructional allocations over the latter half of the kindergarten year (during which time treatment students were receiving intervention), we employed three-level linear growth models. Results of our models (which are summarized here for brevity) revealed, first, that kindergarten teachers did not afford print concepts or fluency any significant amount of time (intercept estimate *t*-test *p* values $> .05$). Moreover, results showed that kindergarten teachers did not systematically change instructional activity time allocations during the latter half of the school year (slope estimate *t*-test *p* values $> .05$). The only exception to this finding was for writing/language arts: for this instructional category, teachers increased their instructional time by an average of 2.51 min per observation period, $t(9) = 2.79$, $p < .05$. Finally, the growth models also revealed significant variation between schools at the initial (January) observation for time spent on comprehension, $Var = 12.89$, $\chi^2(9) = 25.70$, $p < .01$; whole-class instruction, $Var = 175.23$, $\chi^2(9) = 25.07$, $p < .001$; and small-group instruction, $Var = 226.82$, $\chi^2(9) = 33.01$, $p < .001$. Classrooms also systematically varied at the initial observation on phonological awareness, $Var = 13.48$, $\chi^2(14) = 29.03$, $p < .01$; alphabetics, $Var = 4.94$, $\chi^2(14) = 31.98$, $p < .01$; phonics/word study, $Var = 39.75$, $\chi^2(14) = 41.05$, $p < .001$; vocabulary, $Var = 5.52$, $\chi^2(14) = 44.10$, $p < .001$; and other noncontent-instruction minutes, $Var = 36.51$, $\chi^2(14) = 26.51$, $p < .05$. There was less variation in changes over time to instructional categories. Schools varied significantly on changes to small-group instruction only, $Var = 33.24$, $\chi^2(9) = 16.98$, $p < .05$. Classrooms varied slightly more: there was significant between-classroom variance on changes to phonics/word study, $Var = 9.76$, $\chi^2(14) = 25.11$, $p < .05$; spelling, $Var = 0.71$, $\chi^2(14) = 26.71$, $p < .05$; fluency, $Var = 3.30$, $\chi^2(14) = 151.58$, $p < .001$; and writing/language arts, $Var = 17.50$, $\chi^2(14) = 38.17$, $p < .001$.

Given the overwhelming lack of change in instructional minute allocations across observation periods, particularly for phonics/word study, we used each classroom's simple mean phonics/word study minutes (across the three observations) in our student outcomes models (see next section).

Pretests

As shown in Table 3, students in the sample were typically in the lower quartile of skills: conversion of standard scores to percentile ranks shows that the sample averaged in the 10th percentile on receptive vocabulary, 22nd percentile on phonological awareness, and 32nd percentile on word reading. Results from our pretest multilevel models (see Table 4) revealed no evidence of differences between the treatment and control groups; however, as expected, there were large differences between LM learners and non-LM learners on receptive vocabulary and alphabetics but not

Table 2
Minutes Observed During Classroom Literacy Instruction Observations

	January		March		May		Average	
Activity	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Instructional category								
Print concepts	0.08	0.29	0.15	0.51	0.10	0.36	0.11	0.25
Phonological awareness	5.30	6.13	4.50	6.98	2.65	3.46	4.15	3.89
Alphabetic knowledge	2.67	4.04	1.15	1.75	2.13	3.95	1.98	2.67
Word study/phonics	17.66	8.32	14.00	9.55	13.58	9.03	15.08	6.80
Spelling	0.48	1.24	0.95	2.10	0.92	2.92	0.78	1.51
Oral language development	3.64	4.69	5.75	7.55	3.54	3.49	4.31	4.07
Fluency building	0.07	0.36	0.16	0.77	1.51	4.15	0.58	1.66
Text reading	10.14	6.38	10.55	7.57	13.66	10.54	11.45	5.93
Comprehension	6.63	7.09	6.38	6.09	6.05	5.05	6.35	4.07
Writing/language arts	3.10	3.85	4.98	6.84	8.10	10.06	5.39	4.80
Vocabulary	2.75	3.06	1.89	2.66	1.43	1.83	2.02	1.92
Other (nonliteracy)	27.07	10.45	25.01	8.94	23.05	8.35	25.04	6.70
Grouping arrangement								
Whole class	40.07	24.19	47.98	19.67	42.30	22.21	43.45	17.43
Small group	30.12	25.23	17.19	14.84	24.96	15.90	24.11	15.04
Other	9.40	11.88	10.29	7.79	9.45	9.11	9.72	5.51
Total	79.59	19.91	75.46	17.47	76.71	19.45	77.26	15.27

Note. $N = 24$ classroom teachers from 10 schools; observation instrument is the Instructional Content Emphasis—Revised.

Table 3
Student Assessment Simple Means and Standard Deviations

	Treatment (n = 67)								Control (n = 81)							
	LM (n = 38)				Non-LM (n = 29)				LM (n = 46)				Non-LM (n = 35)			
	Pretest		Posttest		Pretest		Posttest		Pretest		Posttest		Pretest		Posttest	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Receptive vocabulary	69.89	13.82			96.10	14.23			72.30	14.81			92.86	11.74		
Alphabetics	8.72	9.09	44.82	16.77	14.76	7.94	48.51	14.13	6.47	7.82	31.61	14.95	14.38	13.52	35.80	17.33
Phonological awareness	81.42	7.07	89.50	10.55	89.17	9.72	98.21	13.22	83.13	6.54	89.20	10.57	86.11	7.47	92.60	10.37
Word reading	92.70	4.55	108.83	8.41	93.62	7.10	112.47	9.75	92.71	5.35	103.17	9.78	93.10	4.97	102.84	11.25
Spelling	4.61	7.34	59.05	33.55	8.34	14.63	77.72	34.68	4.54	7.62	45.98	36.83	8.29	10.66	44.89	34.90
PRF			20.54	10.93			27.19	15.94			9.82	12.54			12.27	11.99
Comprehension			97.50	9.64			103.00	10.43			93.24	8.40			95.20	9.73

Note. N = 148 students from 24 classrooms and 10 schools; LM = language minority; Receptive vocabulary = Peabody Picture Vocabulary Test IIIA standard scores; Alphabetics = mean letters correct per minute; Phonological awareness = Comprehensive Test of Phonological Processing Phonological Awareness standard scores; Word reading = mean of Woodcock Reading Mastery Test—Revised/Normative Update (WRMT–R/NU) Word Identification and Word Attack subtest standard scores; Spelling = Wide Range Achievement Test—Revised Spelling subtest developmental scores of words dictated; PRF = passage reading fluency words correct per minute; Comprehension = WRMT–R/NU Passage Comprehension standard scores.

on word reading and spelling. Finally, although we found no differences between classrooms, we did find significant variation between schools on phonological awareness, word reading, and spelling pretests (chi-square test p values $< .05$).

Zero-Order Correlations

The zero-order correlations between treatment condition, LM status, classroom phonics time, and each student assessment are provided in the Appendix. Although these correlations do not take into account the hierarchical nature of our data (i.e., students' outcomes are nested within classrooms and schools), these data do provide some indication of what we may expect in our results: namely, the positive correlations associated with treatment indicate that treatment has a positive effect on posttests. The Appendix also shows that, as expected, LMs tend to have lower scores at pretest and posttest. Finally, classroom phonics time appears to have a small positive relationship with student outcomes.

Research Questions 1 and 2: What Are the Impacts of Treatment on Student Outcomes, and Are These Impacts Qualified by LM Status and Classroom Time Afforded to Phonics Instruction?

Results from our posttest models (see Table 5) show, first, that treatment effects largely replicate the results of previous kinder-

Table 4
Three-Level Pretest Model Results

	Receptive vocabulary		Alphabetics		Phonological awareness		Word reading		Spelling	
Effect type	*Coeff*	*SE*	*Coeff*	*SE*	*Coeff*	*SE*	*Coeff*	*SE*	*Coeff*	*SE*
Fixed										
Pretest mean[a]	94.57	1.91***	14.53	1.01***	87.70	0.76***	93.00	0.85***	8.16	2.19**
Treatment[b]	0.22	1.18	0.60	0.62	0.30	0.61	0.13	0.44	0.13	0.81
LM[b]	−23.52	2.92***	−6.92	2.12**	−6.05	1.05***	−0.31	0.44	−3.57	2.32
Txt × LM[b]	−1.42	1.28	0.46	0.68	−1.14	0.62	−0.19	0.23	0.00	0.47
	Variance									
Random										
Classrooms[c]	3.27		2.10		0.03		1.99		1.07	
Schools[d]	0.05		2.26		4.49**		5.64***		5.08*	
Residual	181.12		88.86		50.95		21.73		91.74	

Note. N = 148 students from 24 classrooms and 10 schools; Coefficient = coefficient; Receptive vocabulary = Peabody Picture Vocabulary Test IIIA standard score; Alphabetics = mean letters correct per minute; Phonological awareness = Comprehensive Test of Phonological Processing Phonological Awareness standard scores; Word Reading = mean of Woodcock Reading Mastery Test—Revised/Normative Update Word Identification and Word Attack subtest standard scores; Spelling = Wide Range Achievement Test—Revised Spelling subtest developmental scores of words dictated; Treatment = treatment (Txt), effect-coded 1 = treatment, −1 = control; LM = language minority, effect-coded 1 = LM, −1 = non-LM.
[a] t-test $df = 9$. [b] t-test $df = 144$. [c] chi-square test $df = 14$. [d] chi-square test $df = 9$.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 5
Three-Level Posttest Model Results for All Students

Effect type	Alphabetics Coeff	Alphabetics SE	Alphabetics d	Phonological awareness Coeff	Phonological awareness SE	Phonological awareness d	Word reading Coeff	Word reading SE	Word reading d
Fixed									
Mean posttest[a]	41.96	2.05***		95.02	1.54***		107.60	1.69***	
Treatment[b]	6.77	1.23***	0.91	1.44	1.05	0.23	3.91	0.63***	1.03
LM[b]	−4.22	2.39	−0.29	−6.04	2.24**	−0.45	−1.99	2.04	−0.16
Phonics time[c]	0.44	0.30	0.25	0.52	0.23*	0.38	0.42	0.25	0.28
Txt × LM[b]	0.39	0.88	0.07	−1.12	0.65	−0.29	−0.73	0.31*	−0.39
Phon × Txt[b]	−0.21	0.19	−0.18	−0.05	0.11	−0.08	−0.04	0.11	−0.07
Phon × LM[b]	−0.03	0.47	−0.01	−0.31	0.27	−0.19	0.04	0.28	0.03
Phon × Txt × LM[b]	0.16	0.23	0.12	−0.10	0.15	−0.12	0.02	0.07	0.04
					Variance				
Random									
Classrooms[d]		20.01			0.10			21.02***	
Schools[e]		8.60			2.65			0.06	
Residual		203.62			108.47			64.96	
	Spelling			**PRF**			**Comprehension**		
Fixed									
Mean posttest[a]	61.17	6.60***		19.52	2.12***		98.87	1.34***	
Treatment[b]	12.25	3.47***	0.58	6.59	0.86***	1.27	3.00	0.52***	0.96
LM[b]	−11.74	7.55	−0.26	−4.20	2.83	−0.25	−3.46	1.46*	−0.39
Phonics time[c]	0.67	0.62	0.18	0.25	0.20	0.21	0.25	0.17	0.24
Txt × LM[b]	−4.64	3.67	−0.21	−0.96	0.75	−0.21	−0.64	0.42	−0.25
Phon × Txt[b]	−0.79	0.32*	−0.41	−0.04	0.12	−0.05	0.22	0.10*	0.35
Phon × LM[b]	−0.11	0.93	−0.02	0.02	0.27	0.01	0.04	0.21	0.03
Phon × Txt × LM[b]	0.12	0.52	0.04	0.15	0.16	0.15	−0.06	0.05	−0.18
					Variance				
Random									
Classrooms[d]		4.33			6.01			5.81	
Schools[e]		122.80***			0.07			0.32	
Residual		989.76			148.77			75.95	

Note. N = 148 students from 24 classrooms and 10 schools. Coeff = coefficient; d = approximate Cohen's d; Alphabetics = mean letters correct per minute; Phonological awareness = Comprehensive Test of Phonological Processing Phonological Awareness standard scores; Word Reading = mean of Woodcock Reading Mastery Test—Revised/Normative Update (WRMT–R/NU) Word Identification and Word Attack subtest standard scores; Spelling = Wide Range Achievement Test—Revised Spelling subtest developmental scores of words dictated; PRF = passage reading fluency words correct per minute; Comprehension = WRMT–R/NU Passage Comprehension standard scores; Treatment = treatment (Txt) effect, coded 1 = treatment, −1 = control; LM = language minority effect, coded 1 = LM, −1 = non-LM; Phon = mean daily minutes on classroom phonics, in standard deviations.
[a] t-test $df = 9$. [b] t-test $df = 140$. [c] t-test $df = 22$. [d] chi-square test $df = 13$. [e] chi-square test $df = 9$.
* $p < .05$. *** $p < .001$.

garten studies (Vadasy & Sanders, 2008a, 2008b; Vadasy et al., 2006a). Holding all other variables constant, treatment students significantly outperformed controls on every posttest except phonological awareness: treatment students averaged 13.54 more letters correct per minute, 7.82 more standard score points on word reading, 24.50 more points on developmental spelling, 13.18 more words correct per minute on passage reading fluency, and 6.00 more standard score points on comprehension (recall that treatment status was effect coded). Across measures (including phonological awareness), the approximate treatment effect size was large by Cohen's (1988) standard, averaging $d = 0.83$. The approximate Cohen's d for treatment was calculated as the model-estimated treatment slope coefficient divided by the approximate pooled standard deviation; the approximate pooled standard deviation was computed as the square root of the squared model estimate of the standard error divided by the sum of the fractional weights of the two group sizes. The approximate effect size for LM effects was calculated similarly, as the LM coefficient divided by the approximate pooled standard deviation; finally, the pretest effect size was calculated as the difference between the predicted value for one standard deviation above average (the coefficient multiplied by 1) with the predicted value for average (the coefficient multiplied by 0), divided by the approximate pooled standard deviation. The simple pooled standard deviation was not used, as it does not account for the nonindependence of students' scores within classrooms and schools.

LM students performed significantly lower than non-LM students on phonological awareness (12.08 points lower) and comprehension (6.92 points lower), and overall LM students tended to perform lower than non-LM students (average approximate $d = -0.30$). Only one treatment effect was significantly moderated by LM status, and this was on word reading. Upon examination, the interaction is ordinal in nature: The model-implied word reading posttest score for treatment non-LM learners is 114.23 points

[107.60 + (+1)(3.91) + (−1)(−1.99) + (−1)(−0.73)], which is 9.28 points higher than control non-LM learners at 104.95 points [107.60 + (−1)(3.91) + (−1)(−1.99) + (+1)(−0.73)]. Comparatively, LM treatment students were predicted to average 108.79 points and LM controls were predicted to average 102.43 points on posttest word reading, yielding a difference of 6.36 points. Thus, the treatment effect is approximately 1.5 times greater for non-LM than for LM students on word reading.

Model results also show significant unique, positive effects of classroom phonics instruction time on posttest phonological awareness, holding all other things constant. Teachers who spent one standard deviation more minutes on phonics instruction had students predicted to be 0.52 standard score points higher (across all outcomes, the average classroom phonics effect is small; approximate $d = 0.26$). More important, we found that the treatment effects for spelling and comprehension were qualified by classroom time afforded to phonics instruction, although in opposite ways. Figure 1 depicts the observed linear relationships between phonics time (in minutes) and posttests, by experimental condition and LM status.

For spelling, the interaction shows that the treatment effect was strongest in lower phonics classrooms (i.e., there appears little treatment effect in classrooms in which more phonics time was allocated). For comprehension, on the other hand, the treatment effect is higher for students in higher phonics classrooms, all other things constant.

Research Question 3: For LM Students, Does Pretest (English) Receptive Vocabulary Moderate Treatment Response?

Main effects. To test our final research question we analyzed LM student outcomes in isolation and added pretest receptive vocabulary and corresponding interactions into our model. Results revealed that LM students' pretest vocabulary positively predicted all but two outcomes: overall, students who were one standard deviation higher on pretest vocabulary were predicted to have an average score $d = 0.86$ higher at posttest. Results from these analyses (see Table 6) also reflect what was observed for our all-student analyses (see Table 5), with the caveat that the overall treatment effects were slightly smaller for LM students: overall mean approximate treatment effect was $d = 0.71$. Further, LM students in classrooms that spent more time on phonics, instruction had a $d = 0.34$ advantage on posttest, on average, all other things constant.

Interactions. Most important, the results shown in Table 6 reveal that LM treatment response was moderated by receptive vocabulary only for posttest phonological awareness (two-way

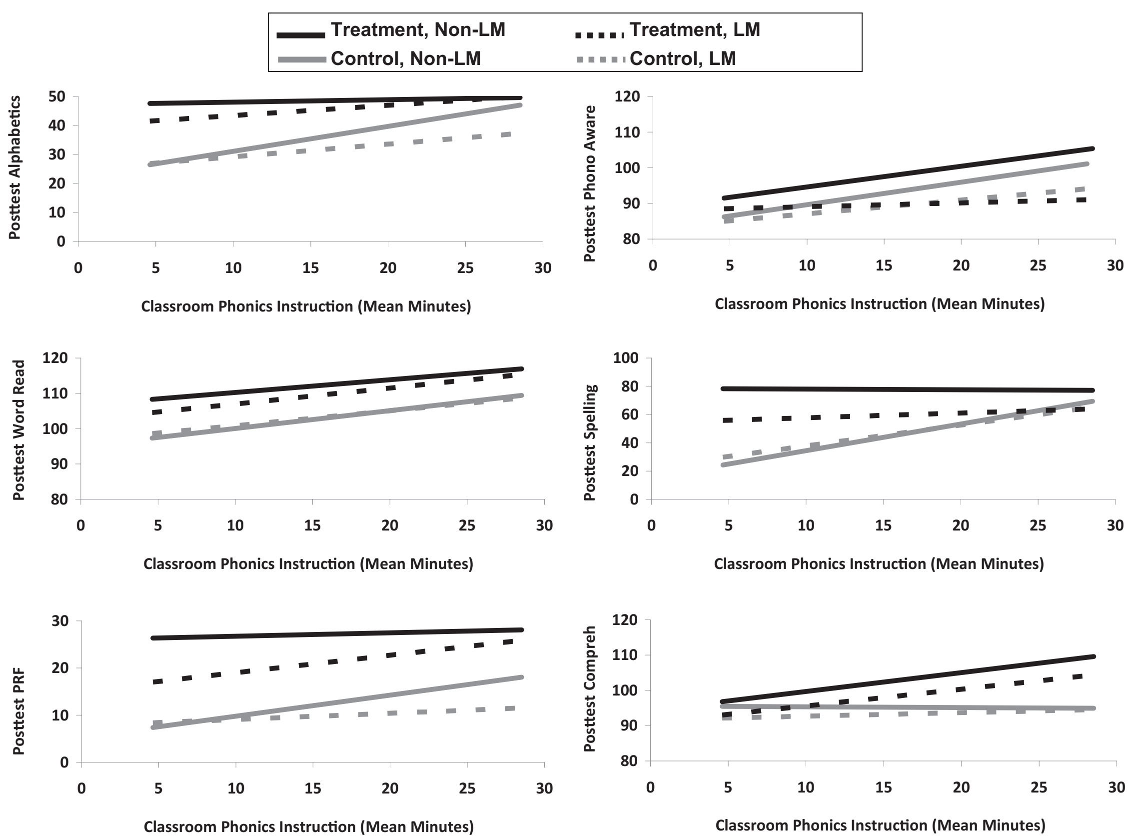

Figure 1. Observed linear relationships between daily average classroom minutes afforded to phonics instruction and posttests, by treatment and language minority (LM) status. $N = 148$ students; significant effect of classroom phonics time on phonological awareness; significant interactions between phonics and treatment on spelling and comprehension (Compreh). PRF = passage reading fluency.

Table 6
Three-Level Posttest Model Results for LM Students Only

Effect type	Alphabetics Coeff	Alphabetics SE	Alphabetics d	Phonological awareness Coeff	Phonological awareness SE	Phonological awareness d	Word reading Coeff	Word reading SE	Word reading d
Fixed									
Mean posttest[a]	37.61	2.22***		88.95	1.16***		105.62	1.26***	
Treatment[b]	7.16	1.55***	1.01	0.68	1.32	0.11	3.11	0.75***	0.91
Vocabulary[b]	2.16	1.40	0.34	3.60	0.60***	1.32	1.87	0.50***	0.82
Phonics time[c]	0.36	0.29	0.27	0.11	0.07	0.34	0.43	0.14**	0.69
Txt × Vocab[b]	−2.78	2.01	−0.30	2.38	0.78**	0.67	−0.09	1.03	−0.02
Phon × Txt[b]	−0.06	0.15	−0.08	−0.22	0.16	−0.30	−0.03	0.09	−0.08
Phon × Vocab[b]	−0.29	0.20	−0.32	−0.01	0.18	−0.01	−0.11	0.12	−0.20
Phon × Txt × Vocab[b]	−0.30	0.21	−0.31	−0.22	0.07**	−0.64	−0.01	0.14	−0.02
					Variance				
Random									
Classrooms[d]		33.58*			0.11			22.17***	
Schools[e]		11.22			3.83			2.21	
Residual		174.94			81.28			46.00	

Effect type	Spelling Coeff	Spelling SE	Spelling d	PRF Coeff	PRF SE	PRF d	Comprehension Coeff	Comprehension SE	Comprehension d
Fixed									
Mean posttest[a]	49.66	4.78***		15.27	1.67***		95.52	0.78***	
Treatment[b]	7.93	5.35	0.32	5.61	1.29***	0.95	2.46	0.57***	0.95
Vocabulary[b]	9.80	1.40***	1.52	1.54	0.94	0.36	2.47	0.67***	0.80
Phonics time[c]	0.46	0.57	0.17	0.14	0.17	0.18	0.27	0.14	0.41
Txt × Vocab[b]	−0.69	4.01	−0.04	−1.86	1.52	−0.27	0.15	1.09	0.03
Phon × Txt[b]	−0.73	0.49	−0.32	0.11	0.18	0.13	0.16	0.11	0.34
Phon × Vocab[b]	−0.29	0.47	−0.13	−0.30	0.12**	−0.57	−0.01	0.23	−0.01
Phon × Txt × Vocab[b]	−0.41	0.38	−0.24	−0.17	0.13	−0.28	0.10	0.20	0.11
					Variance				
Random									
Classrooms[d]		70.47			0.11			5.53	
Schools[e]		108.10*			11.46*			0.02	
Residual		870.73			116.56			63.39	

Note. $N = 84$ language minority (LM) students from 24 classrooms and 10 schools; Coeff = coefficient; d = approximate Cohen's d; Alphabetics = mean letters correct per minute; Phonological awareness = Comprehensive Test of Phonological Processing Phonological Awareness standard scores; Word reading = mean of Woodcock Reading Mastery Test—Revised/Normative Update (WRMT–R/NU) Word Identification and Word Attack subtest standard scores; Spelling = Wide Range Achievement Test—Revised Spelling subtest developmental scores of words dictated; PRF = passage reading fluency words correct per minute; Comprehension = WRMT–R/NU Passage Comprehension standard scores; Treatment = treatment (Txt), effect-coded 1 = treatment, −1 = control; Vocab = pretest Peabody Picture Vocabulary Test IIIA standard scores, in standard deviations; Phon = mean daily minutes on classroom phonics, in standard deviations.
[a] t-test $df = 9$. [b] t-test $df = 140$. [c] t-test $df = 22$. [d] chi-square test $df = 13$. [e] chi-square test $df = 9$.
* $p < .05$. ** $p < .01$. *** $p < .001$.

interaction with treatment and three-way interaction with treatment and classroom phonics). To aid our understanding of this interaction, we created plots (see Figure 2) using LM students' observed linear relationships among pretest receptive vocabulary and posttests by experimental condition (treatment and control) and classroom phonics time; for brevity, these were dichotomized as classrooms at or below mean phonics time (lower phonics) compared with classrooms with greater-than-average time afforded to phonics (higher phonics). As shown in the upper right panel of Figure 2, treatment appears to positively influence phonological awareness for LMs who have higher pretest vocabulary and who are in lower phonics classrooms; by comparison, treatment does not seem to affect phonological awareness for LMs in higher phonics classrooms, irrespective of pretest vocabulary.

Finally, classroom phonics time interacted with pretest vocabulary on passage reading fluency. Inspection of the lower left panel of Figure 2 shows that there is virtually no relationship between pretest vocabulary and posttest fluency for students in higher phonics classrooms. However, there is a positive relationship between vocabulary and fluency in lower phonics classrooms.

Discussion

This study tested the benefits of supplemental phonics instruction for low-skilled LM and non-LM kindergarten students. We found significant positive treatment effects for kindergarten students who averaged in the lower quartile in language and literacy skills at pretest, holding LM status constant. These effects replicate findings from earlier research on supplemental, one-to-one code-oriented instruction for non-LM at-risk kindergarteners (Vadasy & Sanders, 2008a, 2008b; Vadasy et al., 2006a), first graders (Jen-

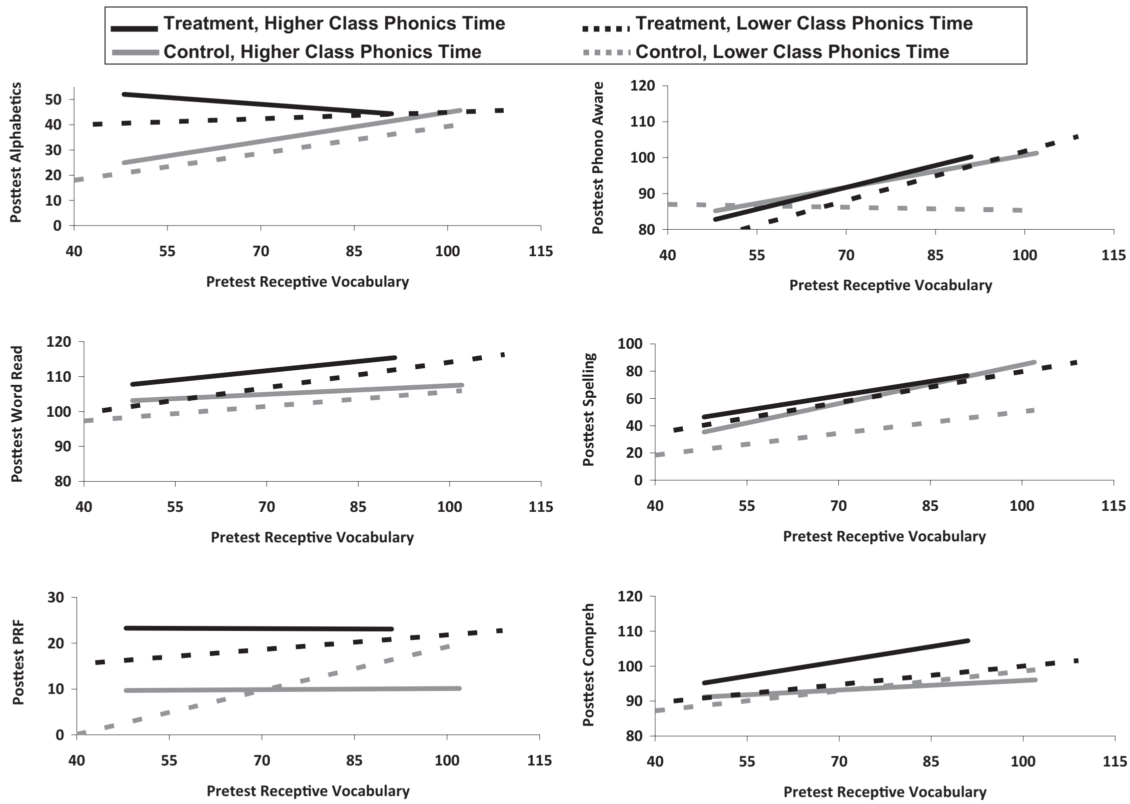

Figure 2. Observed linear relationships between language minority (LM) students' pretest (English) receptive vocabulary and posttests, by treatment status and higher/lower daily average classroom minutes afforded to phonics instruction. $N = 84$ LM students; significant effect of pretest on phonological awareness (Phono Aware), word reading, spelling, and comprehension (Compreh); significant interaction between pretest and treatment on phonological awareness and pretest and phonics on passage reading fluency (PRF).

kins, Peyton, Sanders, & Vadasy, 2004; Vadasy, Jenkins, Antil, Wayne, & O'Connor, 1997a, 1997b; Vadasy, Jenkins, & Pool, 2000), and older children in Grades 2–3 (Vadasy, Sanders, & Peyton, 2005, 2006b; Vadasy, Sanders, & Tudor, 2007). Significant treatment impacts ranged from an approximate $d = 0.58$ for spelling to approximate $d = 1.27$ for passage reading fluency; across all six posttests, treatment effects averaged .83. With respect to LM students, only for word reading was the magnitude of treatment effects significantly smaller than for non-LM students, indicating that the treatment offers an effective approach to boosting most early reading skills for at-risk kindergarten LMs who come from diverse language backgrounds similar to those represented in this study. In their study of English learners and native English-speaking students who were instructed in an intensive literacy curriculum from kindergarten through Grade 5, D'Anguilli, Siegel, and Maggi (2004) found that with strong instruction, the word-reading trajectories for both groups of students became increasingly similar, and similar findings on the "catch-up" in early reading skills for English learners have been reported by Chiappe and Siegel (2006), Chiappe et al. (2002), and Lipka and Siegel (2007). In these studies, although English learners started out with lower phonological and early literacy skills than native English-speaking children, the learning trajectories of these two groups of children rapidly converge when they are provided with classroom literacy instruction that includes training in phonological and phonics skills.

Contribution of Classroom Phonics Time on Student Outcomes

In order to test the impacts of classroom phonics instruction time on at-risk kindergarten student outcomes, we observed the kindergarten literacy blocks of participating students three times during the latter half of their kindergarten year, capturing amount of time spent on a comprehensive set of instructional dimensions. Although we did not attempt to directly assess teacher quality, we were able to capture time spent on phonics/word study instruction (the focus of the intervention), which turned out to be the second most frequently observed dimension of kindergarten literacy instruction. Because we found no significant change in teachers' time spent on phonics between January and May, we used the mean minutes spent on phonics/word study as a classroom-level predictor in our posttest models. Results showed significant positive effects of classroom phonics instruction time on phonological awareness and comprehension, holding all other things constant (across all measures, the average approximate classroom phonics effect was $d = 0.26$). Further, we found that the treatment effects detected for developmental spelling and comprehension were qualified by classroom time afforded to phonics instruction. Controls benefited from being in higher phonics classrooms on spelling, whereas treatment students in higher phonics classrooms had an advantage on comprehension. Taken together, it appears that higher amounts of kindergarten classroom phonics time benefits spelling, whereas a double dose of phonics instruction (in the classroom and in a pull-out phonics-based intervention) benefits comprehension. In other words, additive phonics instruction may have comprehension benefits for treatment students in this early stage of reading development when decoding problems represent a major, word-level obstacle to comprehension.

Role of English Vocabulary on LM Student Outcomes

When we examined LM kindergarten outcomes separately from non-LM students, we found that LM treatment response was generally not moderated by pretest English receptive vocabulary. The exception to this was for phonological awareness: Treatment was beneficial to LM learners' phonological awareness if they (a) were at least somewhat proficient in English vocabulary and (b) were situated in lower phonics classrooms. Findings suggest that vocabulary may play a role in the development of phonological awareness, consistent with the lexical restructuring model (Metsala & Walley, 1998).

Limitations

The findings of the present study must be considered in light of several limitations. First, our definition of LM status was based on parent self-report of language primarily spoken in the home rather than students' actual English language proficiency. Nevertheless, this definition has been used in previous studies (e.g., Lipka & Siegel, 2007; Silverman & Hines, 2009). Second, we did not have information on language proficiency levels in students' primary home language, or on early home literacy practices that might account for differences in L2 reading development, and we did not differentiate LM students on the basis of English proficiency levels. It is noteworthy, however, that the language diversity of this sample precluded obtaining home language proficiency levels. Third, our models do not account for bilingual services students received, which varied across schools in part based on the proportion of LM students enrolled. Nevertheless, we did account for between-school and between-classroom variation in the dependent measures in all of our models. Fourth, children in our study were drawn from 10 urban public elementary schools; the results of this study, as with any study, generalize only to similar populations. Finally, the treatment effects sizes found in the current study apply to individual, 1:1, tutored instruction, a delivery model most appropriate for at-risk students and expensive for many schools to adopt. Future research could examine the benefits of small-group instruction for LM learners.

Practical Implications

LM students. Our findings provide support for benefits of supplementing classroom reading instruction with phonics-based tutoring for both LM and non-LM kindergarten students screened to be at risk for poor reading outcomes. The fact that this tutoring intervention was effectively implemented by paraeducators extends options available to schools attempting to augment classroom kindergarten literacy instruction. We have reported previously on the effectiveness of similar instruction for native English-speaking students provided by paraeducators with skill levels similar to those in this study. Tutors implemented the explicitly specified phonics-based instruction with a high degree of fidelity with kindergarten LM students with widely varied levels of English vocabulary. Tutors noted that many simple words needed to be explained to LM students and suggested the word lists include picture stimuli that could be quickly referenced. This would allow tutors to easily add incidental vocabulary instruction.

Although our results show that pretest English vocabulary in general did not moderate treatment response for LMs, the paraeducators in the present study on occasion requested added support from research staff coaches to work with LM students who were very recent arrivals with minimal oral English language skills. In a few cases, research staff helped tutors obtain translation assistance in the building to make the directions for the phonics lesson activities clear for the students. The language diversity of the children in the study limited the extent to which we could offer this support. In other cases, tutors modeled both directions and responses in the phonics activities for students with very limited English skills until the students were able to respond. Overall, these types of assistance and scaffolds were used infrequently and briefly. Tutors were generally able to use the explicit, scripted phonics lessons without significant adaptations. Tutors added incidental vocabulary instruction as time permitted and were cautioned not to divert the focus of their instruction. Had additional time been available in this study for the pull-out instruction, tutors could have effectively provided added instruction in word meanings.

Classroom phonics time. Classroom observations across three time points in the latter half of the academic year revealed that kindergarten teachers allocated substantive time to phonics/word study instruction (18%–22% of average time) and text reading (13%–18% of average time). This primary content focus of the classroom literacy instruction is closely aligned with the content focus of the tutoring intervention. Phonics and word reading instruction provided in the classroom had general, albeit small, positive benefits across reading outcomes. The proportion of time observed for "other" nonliteracy activities during the reading block may reflect the time teachers must make when they serve large numbers of students who are pulled out for special services. The positive effects of classroom phonics time on student outcomes in general, as well as the advantage of more classroom phonics time for controls on spelling, and for treatment students on comprehension, suggest that a closer examination of instructional time allocation in the literacy block is warranted. It may be possible for many teachers to reduce time spent on "other" content and increase the instructional intensity of the literacy blocks. We expect that many teachers were not aware of how much time they spent on transitions, student management, and classroom organization. Others have found that teachers respond to professional development on aspects of literacy instruction by adjusting their instructional focus (Correnti, 2007).

Conclusion

This study allowed us to consider the simultaneous influences of supplemental code-oriented intervention, language status, and classroom literacy instruction on kindergarten reading outcomes. The results of our models reflect the collective nature of teaching for many at-risk students (Valli, Croninger, & Walters, 2007).

Findings demonstrate the complexity of instructional interactions for students, both non-LM and LM, at risk due to low levels of phonological and alphabetic skills, in determining literacy outcomes.

References

Adams, M. J. (1990). *Beginning to read.* Cambridge, MA: Bradford.

August, D., & Erickson, F. (2006). Qualitative studies of classroom and school practices. 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. 498–522). Mahwah, NJ: Erlbaum.

August, D., & Hakuta, K. (1998). *Educating language minority children.* Washington, DC: National Academy Press.

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

Berninger, V. W. (1998). *Process assessment of the learner (PAL): Guides for intervention and PAL intervention kit.* San Antonio, TX: Psychological Corporation.

Blachman, B., Schatschneider, C., Fletcher, J., Francis, D. J., Clonan, S. M., Shaywitz, B. A., & Shaywitz, S. E. (2004). Effects of intensive reading remediation for second and third graders and a 1-year follow-up. *Journal of Educational Psychology, 96,* 444–461.

Blom-Hoffman, J., Leff, S. S., Franko, D. L., Weinstein, E., Beakley, K., & Power, T. J. (2009). Consent procedures and participation rates in school-based intervention and prevention research: Using a multi-component, partnership-based approach to recruit participants. *School Mental Health, 1,* 3–15.

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

Byrne, B., & Fielding-Barnsley, R. (1991). Evaluation of a program to teach phonemic awareness to young children. *Journal of Educational Psychology, 83,* 451–455.

Capps, R., Fix, M., Murray, J., Ost, J., Passel, J., & Herwantoro, S. (2005). *The new demography of America's schools: Immigration and the No Child Left Behind Act.* Washington, DC: Urban Institute.

Center for Public Education. (2007). *Preparing English language learners for academic success.* Retrieved from http://www.centerforpubliceducation.org/site/c.lvIXIiN0JwE/b.5057521/k.A64C/Preparing_English_language_learners_for_academic_success.htm

Chall, J. S. (1967). *Learning to read: The great debate.* New York, NY: McGraw-Hill.

Chambers, B., Slavin, R. E., Madden, N. A., Cheung, A., & Gifford, R. (2004). *Effects of Success for All with embedded video on the beginning reading achievement of Hispanic children.* Baltimore, MD: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk.

Chiappe, P., & Siegel, L. S. (2006). A longitudinal study of reading development of Canadian children from diverse language backgrounds. *Elementary School Journal, 107,* 135–152.

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.

Cirino, P. T., Pollard-Durodola, S. D., Foorman, B. R., Carlson, C. D., & Francis, D. J. (2007). Teacher characteristics, classroom instruction, and student literacy and language outcomes in bilingual kindergartners. *Elementary School Journal, 107,* 341–364.

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

Correnti, R. (2007). An empirical investigation of professional development effects on literacy instruction using daily logs. *Educational Research and Policy Analysis, 29,* 262–295.

D'Anguilli, A., Siegel, L. S., & Maggi, S. (2004). Literacy instruction, SES, and word-reading achievement in English-language learners and children with English as a first language: A longitudinal study. *Learning Disabilities Research & Practice, 19,* 202–213.

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.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test–IIIA* (3rd ed.). Circle Pines, MN: American Guidance Service.

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. 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.

Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology, 92,* 605–619.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1997). *Children, schools, and inequality.* Boulder, CO: Westview.

Foorman, B. R., Francis, D. J., Winikates, D., Mehta, P., Schatschneider, C., & Fletcher, J. M. (1997). Early interventions for children with reading disabilities. *Scientific Studies of Reading, 1,* 255–276.

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

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

Fuchs, D., Fuchs, L. S., Thompson, A., Al Otaiba, S., Yen, L., Yang, N. J., . . . O'Connor, R. E. (2001). Is reading important in reading-readiness programs? A randomized field trial with teachers as program implementers. *Journal of Educational Psychology, 93,* 251–267.

Gandara, P., Maxwell-Jolly, J., & Driscoll, A. (2005). *Listening to teachers of English language learners: A survey of California teachers' challenges, experiences, and professional development needs.* Santa Cruz, CA: University of California.

Gersten, R., Baker, S. K., Shanahan, T., Linan-Thompson, S., Collins, P., & Scarcella, R. (2007). *Effective literacy and English language instruction for English learners in the elementary grades* (NCEE Report 2007–4011). Retrieved from http://ies.ed.gov/ncee/wwc/pdf/practicequides/20074011.pdf

Goldenberg, C. (2008). Teaching English language learners: What the research does—and does not—say. *American Educator, 32,* 8–23, 42–43.

Good, R. H., & Kaminski, R. A. (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

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.

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

Hatcher, P. J., Goetz, K., Snowling, M. J., Hulme, C., Gibbs, S., & Smith, G. (2006). Evidence for the effectiveness of the Early Literacy Support programme. *British Journal of Educational Psychology, 76,* 351–367.

Hatcher, P. J., Hulme, C., & Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development, 65,* 41–57.

Hoffman, J. V., Sailors, M., & Patterson, E. U. (2002). *Decodable texts for beginning reading instruction* (Report No. 1–016). Ann Arbor: Center for the Improvement of Early Reading Achievement, University of Michigan.

Hopstock, P. J. (2003). *Descriptive study of services to LEP students and LEP students with disabilities: Special Topic Report #3. Issues in studying learning outcomes for LEP students.* Arlington, VA: Development Associates.

Hopstock, P. J., & Stephenson, T. G. (2003). *Descriptive study of services to LEP students and LEP students with disabilities: Special Topic Report #1. Native languages of LEP students.* Retrieved from http://www.ncela.gwu.edu/files/rcd/BE021196/BE021196DSSST1_NativeLangsofLE.pdf

Jastak, S., & Wilkinson, G. S. (1984). *The Wide Range Achievement Test—Revised.* Wilmington, DE: Jastak Associates.

Jenkins, J. R., Peyton, J. A., Sanders, E. A., & Vadasy, P. F. (2004). Effects of reading decodable texts in supplemental first-grade tutoring. *Scientific Studies of Reading, 8,* 58–86.

Juel, C. (1996). What makes literacy tutoring effective? *Reading Research Quarterly, 31,* 268–289.

Juel, C., & Minden-Cupp, C. (2000). Learning to read words: Linguistic units and instructional strategies. *Reading Research Quarterly, 35,* 458–492.

Kieffer, M. J. (2008). Catching up or falling behind? Initial English proficiency, concentrated poverty, and the reading growth of language minority learners in the United States. *Journal of Educational Psychology, 100,* 851–868.

Lesaux, N. K. (2006). Building consensus: Future directions for research on English language learners at risk for learning disabilities. *Teachers College Record, 11,* 2406–2438.

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., & Siegel, L. S. (2003). The development of reading in children who speak English as a second language. *Developmental Psychology, 39,* 1005–1019.

Lipka, O., & Siegel, L. S. (2007). The development of reading skills in children with English as a second language. *Scientific Studies of Reading, 11,* 105–131.

Makar, B. (1996). *Primary phonics workbook 1.* Cambridge, MA: Educators Publishing Service.

Maslen, B. L. (2003). *Bob books.* New York, NY: Scholastic.

McCardle, P., & Chhabra, V. (2006). Commentary. *Elementary School Journal, 107,* 239–248.

Mesmer, H. (2004, December). *Textual scaffolds and beginning readers: Predicting accuracy and fluency in leveled and decodable text.* Paper presented at the annual meeting of the National Reading Conference, San Antonio, TX.

Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–121). Mahwah, NJ: Erlbaum.

National Institute of Child Health and Human Development. (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* (NIH Publication No. 00–4769). Washington, DC: U.S. Department of Health and Human Services.

No Child Left Behind Act, 20 U.S. C. §§ 6310 et seq. (2001).

Phillips, M., Crouse, J., & Ralph, J. (1998). Does the black/white test score gap widen after children enter school? In C. Jencks & M. Phillips (Eds.), *The black/white test score gap* (pp. 229–272). Washington, DC: Brookings.

Planty, M., Hussar, W., Snyder, T., Kena, G., Kewal Ramani, A., Kemp, J., . . . Dinkes, R. (2009). *The condition of education 2009* (NCES 2009-081). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly, 27,* 29–53.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. T. (2004). *HLM for Windows 6.0.* Lincolnwood, IL: Scientific Software International.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2,* 31–74.

Rossell, C. (2005). *Making uneven strides: State standards for achieving English language proficiency under the No Child Left Behind Act.* Arlington, VA: Lexington Institute.

Saunders, W. M., Foorman, B. R., & Carlson, C. D. (2006). Is a separate block of time for oral English language development in programs for English learners needed? *Elementary School Journal, 107,* 181–198.

Savage, R., & Carless, S. (2005). Learning support assistants can deliver effective reading interventions for "at-risk" children. *Educational Research, 47,* 45–61.

Schneider, W., Roth, E., & Ennemoser, M. (2000). Training phonological skills and letter knowledge in children at risk for dyslexia: A comparison of three kindergarten intervention programs. *Journal of Educational Psychology, 92,* 284–295.

Shanahan, T., & Beck, I. L. (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, MI: University of Michigan Press.

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

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.

Silverman, R., & Hines, S. (2009). The effects of multimedia-enhanced instruction on the vocabulary of English-language learners and non-English language learners in pre-kindergarten through second grade. *Journal of Educational Psychology, 101,* 305–314.

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

Snowling, M. J. (1991). Developmental reading disorders. *Journal of Child Psychology and Psychiatry, 32,* 49–77.

Stanovich, K. E. (1980). Toward an interactive–compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16,* 32–71.

Stanovich, K. E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-differences model. *Journal of Learning Disabilities, 21,* 590–604.

Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers.* New York, NY: Guilford Press.

Stuart, M. (1999). Getting ready for reading: Early phoneme awareness and phonics teaching improves reading and spelling in inner-city second language learners. *British Journal of Educational Psychology, 69,* 587–605.

Tangel, D. M., & Blachman, B. A. (1992). Effect of phoneme awareness instruction on kindergarten children's invented spelling. *Journal of Reading Behavior, 24,* 233–261.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., & Conway, T. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91,* 579–593.

Tunmer, W. E., Herriman, M. L., & Nesdale, A. B. (1988). Metalinguistic abilities and beginning reading. *Reading Research Quarterly, 23,* 134–158.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2007, July 30). *English language learners.* Retrieved from http://ies.ed.gov/ncee/wwc/pdf/ELL_TR_07_30_07.pdf

Vadasy, P. F., Jenkins, J. R., Antil, L. R., Wayne, S. K., & O'Connor, R. E. (1997a). Community-based early reading intervention for at-risk first graders. *Learning Disabilities: Research and Practice, 12,* 29–39.

Vadasy, P. F., Jenkins, J. R., Antil, L. R., Wayne, S. K., & O'Connor, R. E. (1997b). The effectiveness of one-to-one tutoring by community tutors for at-risk beginning readers. *Learning Disability Quarterly, 20,* 126–139.

Vadasy, P. F., Jenkins, J. R., & Pool, K. (2000). Effects of tutoring in phonological and early reading skills on students at risk of reading disabilities. *Journal of Learning Disabilities, 33,* 579–590.

Vadasy, P. F., & Sanders, E. A. (2008a). Code-oriented instruction for kindergarten students at risk for reading difficulties: A replication and comparison of instructional grouping. *Reading and Writing: An Interdisciplinary Journal, 21,* 929–963.

Vadasy, P. F., & Sanders, E. A. (2008b). Individual tutoring for struggling readers: Moving research to scale with interventions implemented by paraeducators. In G. A. Reid, A. Fawcett, F. Manis, & L. Seigel (Eds.), *The Sage handbook of dyslexia* (pp. 337–355). Thousand Oaks, CA: Sage.

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2005). Relative effectiveness of reading practice or word-level instruction in supplemental tutoring: How text matters. *Journal of Learning Disabilities, 38,* 364–380.

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006a). 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.

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006b). Paraeducator-supplemented instruction in structural analysis with text reading practice for second and third graders at risk for reading problems. *Remedial and Special Education, 27,* 365–378.

Vadasy, P. F., Sanders, E. A., & Tudor, S. (2007). Effectiveness of paraeducator-supplemented individual instruction: Beyond basic decoding skills. *Journal of Learning Disabilities, 40,* 508–525.

Valli, L., Croninger, R. G., & Walters, K. (2007). Who (else) is the teacher? Cautionary notes on teacher accountability systems. *American Journal of Education, 113,* 635–662.

Wagner, R., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing.* Austin, TX: Pro-Ed.

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

Appendix

Zero-Order Correlations

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Conditions														
1. Treatment	—													
2. LM	.00	—												
3. Phonics time	−.04	−.07	—											
Pretests														
4. Receptive vocabulary	.00	**−.64**	.07	—										
5. Alphabetics	.07	**−.34**	.10	**.34**	—									
6. Phonological awareness	.02	**−.32**	.11	**.44**	**.29**	—								
7. Word reading	.02	−.06	**.23**	.09	**.48**	**.33**	—							
8. Spelling	.00	**−.18**	**.20**	**.34**	**.56**	**.39**	**.50**	—						
Posttests														
9. Alphabetics	**.38**	−.12	**.17**	.13	**.43**	.10	**.26**	**.30**	—					
10. Phonological awareness	.11	**−.25**	**.25**	**.40**	**.29**	**.61**	**.26**	**.36**	**.44**	—				
11. Word reading	**.35**	−.07	**.27**	**.17**	**.38**	**.33**	**.49**	**.38**	**.68**	**.59**	—			
12. Spelling	**.29**	−.11	**.18**	**.25**	**.32**	**.28**	**.22**	**.39**	**.70**	**.62**	**.66**	—		
13. PRF	**.44**	−.15	.11	.15	**.53**	.13	**.45**	**.42**	**.74**	**.40**	**.74**	**.56**	—	
14. Comprehension	**.29**	**−.18**	**.17**	**.30**	**.36**	**.47**	**.45**	**.33**	**.45**	**.54**	**.73**	**.47**	**.59**	—

Note. $N = 148$ students from 24 classrooms and 10 schools. Correlations in boldface are significant at the .05 level. Pearson's r reported. Treatment dummy coded 1 = treatment, 0 = control; LM = language minority, dummy coded 1 = LM, 0 = non-LM; Phonics time = mean daily minutes on classroom phonics; Receptive vocabulary = Peabody Picture Vocabulary Test IIIA standard scores; Alphabetics = mean letters correct per minute; Phonological awareness = Comprehensive Test of Phonological Processing standard scores; Word reading = mean of Woodcock Reading Mastery Test—Revised/Normative Update Word Identification (WRMT–R/NU) and Word Attack subtest standard scores; Spelling = Wide Range Achievement Test—Revised Spelling subtest developmental scores of words dictated; PRF = passage reading fluency words correct per minute; Comprehension = WRMT–R/NU Passage Comprehension standard scores.

Received July 1, 2009
Revision received January 22, 2010
Accepted March 29, 2010 ■