Read Writ (2013) 26:821–843
DOI 10.1007/s11145-012-9393-4

Two-year follow-up of a code-oriented intervention for lower-skilled first-graders: the influence of language status and word reading skills on third-grade literacy outcomes

Patricia F. Vadasy · Elizabeth A. Sanders

Published online: 3 July 2012

Abstract For 2 years we followed lower-performing English learner (EL) and native English speaking (non-EL) students who participated in an efficacy trial of a supplemental first-grade code-oriented intervention implemented by paraeducators. At the end of grade three, across all students ($n = 180$ of the original 187 students), treatment effects were maintained on word reading (approximate $d = .45$), spelling (.36) and reading comprehension (.24). However, treatment effects tended to be smaller for EL students, and were significantly smaller for spelling in particular. While pretest grade one word reading did not moderate treatment response for either ELs or non-ELs, it was found to strongly predict all three end-of- grade-three outcomes, although to a lesser extent for ELs on reading comprehension. Findings add support to previous research on the benefits of early code-oriented tutoring.

Keywords English language learner · Language minority · Reading · Intervention · Longitudinal studies · Multilevel modeling · Randomized trial · Follow-up · First-grade · Third-grade

Introduction

Increasing numbers of children who are English learners in U.S. schools face multiple challenges in learning to read. Like all beginning readers, they must master prerequisite skills of decoding and recognition of printed words. For native English speaking children, these decoding obstacles are the most common cause of reading difficulties, and prerequisite alphabet skills and phoneme awareness are reliable

P. F. Vadasy (✉)
Washington Research Institute, 150 Nickerson Street, Suite 305, Seattle, WA 98109, USA
e-mail: pvadasy@wri-edu.org

E. A. Sanders
University of Washington, Seattle, WA, USA

predictors of word reading skill (Adams, 1990). English learners may be at risk for reading difficulties due to these factors known to influence early reading acquisition for native English speakers. Children from homes in which English is not the first language may have limited exposure to spoken and written English that supports development of early literacy skills (Yeong & Liow, 2011). In particular for children with recent immigration status, their rate of second-language acquisition influences the English vocabulary acquisition and growth that support word-level reading (Paez, Tabors, & Lopez, 2007). Young English learners may often experience poverty and lower levels of family education and literacy associated with poor reading and academic outcomes (Capps et al., 2005; Fry & Gonzales, 2008). Recent National Assessment of Educational Progress (NAEP) data indicate that Hispanic English learners with limited English language skills lag far behind both white students and Hispanic students with stronger English language skills at both fourth and eighth grades (Hemphill & Vannemann, 2011). Although schools in parts of the country that serve primarily Spanish-speaking English learners may often provide bilingual support in reading instruction, there are few data on reading outcomes for English learners enrolled in schools serving large numbers of diverse language groups. In a classroom with large numbers of English learners from diverse backgrounds, teachers are challenged to identify those students whose early literacy, language, and home experiences put them at risk for poor reading outcomes.

As students enter kindergarten and first grade, schools often undertake identification and early literacy interventions to prevent struggling students from falling further behind. Most early literacy interventions focus on developing decoding skills, guided by both strong theory of first language (L1) reading development and the established efficacy of phonics-based early interventions (Bus & van IJzendoorn, 1999; Ehri, Nunes, Stahl, & Willows, 2001; Torgesen, 2002). In the early stages of reading acquisition when listening comprehension typically parallels spoken language, and when the defining deficit for struggling students is phonologically based reading skills (Shankweiler & Liberman, 1972), early intervention for at-risk children is most often characterized by explicit instruction in the alphabetic principle and phonological processing with practice of word-level skills in connected text (Torgesen, Wagner, & Rashotte, 1997).

Evidence for early intervention approaches for native English-speaking children at risk for reading difficulties that develop initial word reading skills is solidly established (Snow, Burns, & Griffin, 1998), although it is less clear if these approaches have similar benefits for English learners (Gersten & Baker, 2003). Recognizing the difficulty of closing the reading achievement gap between English learners and students who are native English speakers, researchers have examined early intervention models similar to those found effective for at-risk English-speaking students (e.g., Denton, Anthony, Parker, & Hasbrouck, 2004; Gunn, Biglan, Smolkowski, & Ary, 2000). Often these interventions are more comprehensive than the systematic and explicit phonics-based approaches used effectively with at-risk K-1 native speakers. There may be added emphasis to engage English learners in comprehending connected text, and opportunities to develop oral language, vocabulary and comprehension skills through structured questions and student–

teacher interchange. In a noteworthy study of a first-grade English language intervention based on these design principles and intervention research Vaughn and her colleagues (Vaughn et al., 2006) identified first-grade English learners (Spanish/English) at risk for reading problems. Students were randomly assigned within schools to either supplemental English language intervention or to contrast conditions provided by the schools. Bilingual teachers were trained to provide intervention to small groups of students, 50 min a day, 5 days a week for the school year. The daily lessons included explicit instruction in five areas: phonemic awareness, letter knowledge, word recognition, connected text fluency, and comprehension strategies. Key features of instruction in both word-level and comprehension skills included: teacher modeling, feedback and scaffolding student responses, and well-paced instructional delivery to maintain student attentiveness and engagement within the small groups. At the end of first-grade, intervention students were significantly higher than contrast students on posttest measures of letter naming fluency, phonological awareness, word attack, and passage comprehension. At a 1-year follow up at the end of second grade, intervention students significantly outperformed comparison students on English measures of oral language, letter-word identification, word attack, passage comprehension, and oral reading fluency (Cirino et al., 2009). In a follow up at the end of fourth or fifth grade, the intervention students continued to outperform comparison students in letter-word identification and connected text fluency, although with small to moderate effects (+.45 and +.25, respectively) (Vaughn et al., 2008). These studies demonstrated the effectiveness, through follow up, of an intensive and multi-component reading intervention delivered by skilled bilingual teachers with a comprehensive focus on word reading, oral language, vocabulary, and comprehension.

In the current research we report findings for a supplemental intervention for first-grade English learners that emphasized foundation code-based skills. The intervention has been evaluated for use with native English-speaking students in previous randomized control trials (Vadasy, Sanders, & Peyton, 2006; Vadasy & Sanders, 2008). Although it is recognized that English learners at risk for reading problems have additional language needs that influence development of reading skills, there remains a lacuna in research on school-based interventions that provide adequate language input, practice, and support for vocabulary and language learning. With limited research on interventions tested to boost language comprehension skills of English learners, schools, particularly in parts of the U.S. serving students from diverse language groups, may turn to the phonics-based interventions they have found effective for many struggling beginning readers. For this reason, follow up of English learners who receive supplemental code-based early intervention is of special interest to determine whether long-term benefits of this instruction maintain, and whether they are moderated by language-group status or initial word-reading skills. To understand the skills that contribute to individual differences in reading comprehension and inform emphases for reading instruction, educators and researchers have used the “Simple View of Reading” (SVR) model (Gough & Tunmer, 1986; Hoover & Gough, 1990). The SVR model represents reading comprehension (RC) as the product of two primary components: competence in decoding words (D) measured with decoding or word reading tasks, and competence in linguistic comprehension (LC). Many English learners enter school with lower levels of oral language skills, including vocabulary (Verhallen & Schoonen, 1993; Verhoeven, 2000), syntactic knowledge (Gottardo, 2002; Verhoeven, 1990), and listening comprehension (Gottardo & Mueller, 2009; Proctor, Carlo, August, & Snow, 2005). Studies show that these language skills significantly influence the development of reading comprehension for English learners (Carlisle, Beeman, Davis, & Spharim, 1999; Catts, Fey, Zhang, & Tomblin, 1999; Demont & Gombert, 1996; Lesaux, Lipka, & Siegel, 2006; Verhoeven, 2011). More recently researchers have used the SVR model to describe the development of reading comprehension in L1 and in L2 learners. These studies indicate that the early development of decoding skills and SVR relationships is quite similar for L1 and L2 students (Chiappe & Siegel, 2006; Gottardo & Mueller, 2009). However, English learners' lower levels of listening comprehension and vocabulary knowledge exert an increasing influence on reading comprehension beyond the primary grades (Carlisle, Beeman, Davis, & Spharim, 1999; Catts, Hogan, & Adolf, 2005; Droop & Verhoeven, 2003; Manis, Lindsey, & Bailey, 2004; Nakamoto, Lindsey, & Manis, 2007; Proctor et al., 2005; Verhoeven, 2000). In their study of simple view relationships, Harlaar et al. (2010) found that home literacy environment accounted for significant variance in word recognition, listening comprehension, vocabulary, and reading comprehension at age nine. We expect that home literacy environments differ for EL and non-EL children in the present study. Although earlier studies describe the role of both word reading and linguistic comprehension in the reading comprehension of English learners (e.g., Hoover & Gough, 1990; Mancilla-Martinez, Kieffer, Biancarosa, Christodoulou, & Snow, 2011; Proctor et al., 2005), previous studies have not specifically examined the relationships of the SVR components across the period of early word-reading development (K-3). In this study we consider the role of early word reading skill and language status that may differentiate between EL and non-EL longer-term intervention outcomes.

Current study

The present follow-up study investigated the longer-term treatment effects of an explicit, code-oriented intervention that was previously demonstrated to improve reading skills of lower performing English learner (EL) and native speaking (non-EL) first graders at immediate posttest. We use the term English learner to describe the children in this study to be consistent with how they were identified by the participating schools, although not all students had limited proficiency in English. (Note: like many researchers, we prefer use of person-first oriented language such as "children identified as English Learners"; however, for brevity and clarity throughout the remainder of this article, we employ simple "EL" and "non-EL" labels for group comparisons.) Research questions guiding the current study were as follows.

1. Are first-grade treatment effects (across both EL and non-EL students) on literacy outcomes maintained 2 years after intervention?

2. Do either language status or first-grade pretest word reading moderate longer-term treatment effects?
3. Are treatment effects on longer-term outcomes dependent on a joint interaction with EL status and pretest word reading?

Original intervention study findings

Previously, Vadasy and Sanders (2011) examined the efficacy of a supplemental phonics-based intervention with a sample of lower-skilled EL and non-EL first graders screened at relative risk for reading difficulties, and randomly assigned within classroom either to individual supplemental instruction or to classroom instruction control. Students in the treatment condition received 20 weeks of individual tutoring from trained paraeducator tutors using scripted lessons (30 min/day for 4 days/week). (Additional details about the original study procedures are provided in the "Methods" section.) There were no significant pretest differences between treatment and control students on any measure; however, EL students were over one standard deviation lower than non-EL students on receptive language (measured with the *Peabody Picture Vocabulary Test-IIIA*), a difference similar to other previous research (Bowers & Vasilyeva, 2011; Oller & Eilers, 2002) (no other differences between EL and non-EL students were detected at pretest). At first-grade posttest, treatment students significantly outperformed controls on measures of alphabetic knowledge, word reading, spelling, passage reading fluency, and reading comprehension. Nevertheless, EL students were significantly lower than non-EL students on passage reading fluency, and effect sizes for EL students were consistently lower than for non-ELs (effect sizes were approximately one-third the effect size for non-EL students) (Vadasy & Sanders, 2011).

Methods

Original intervention study sample

Recruitment and screening

In October of 2007–2008, all students ($n = 903$) in first-grade classrooms at 13 U.S. urban public elementary schools (45 first-grade classrooms) in the Pacific Northwest known for relatively large proportions of English learner (EL) student enrollment were invited to participate in the research study. About 40 % of students in the district were eligible for free or reduced-price lunch, and about 20 % of students had a non-English speaking background. Over 100 languages are spoken by students enrolled in the district. During recruitment, a student was designated as EL if the student's parent had reported the primary home language as other than English on the student's school registration record. Consents were sent home via students' classrooms; for children identified as English learners, consents were provided in both English and the child's home language. After a 1-month period, we had a 61 %

active consent rate, with 553 students (282 EL) available for screening. However, due to insufficient student sample sizes for random assignment to experimental conditions (within EL/non-EL subgroups within classrooms), 15 classrooms (and thus two schools) were removed from study participation prior to screening. Additionally, some students had moved from their schools prior to, or were absent during, screening. As such, 399 students (214 EL) were screened using measures of alphabetic knowledge (number of letter sounds and letter names produced out of 52 randomly ordered upper-case English letters), and phonological awareness (Sound Matching subtest from the Comprehensive Test of Phonological Processing (Wagner, Torgesen, & Rashotte, 1999)). These short measures that strongly predict reading outcomes were used to efficiently identify the lower performers in each classroom (in order to begin full the pretesting as quickly as possible).

Eligibility and random assignment

We randomly assigned children who were relatively low at screening to treatment and control groups within classrooms, within EL/Non-EL group. To accomplish this we first separated students by classroom; then, we separated EL and non-EL students within each classroom. Next, we computed a composite z-score for each student, based on the mean of the three screening measure z-scores, within EL/non-EL group, within classroom. Students in the lower half of their classroom's EL/non-EL group were then randomly assigned to treatment (supplemental phonics-based tutoring) or control (regular classroom instruction, no tutoring) conditions. We employed this relative-risk sampling strategy primarily to assure roughly balanced subgroup sizes for testing whether treatment effects were moderated by EL status. However, this sampling method resulted in the inclusion of students who would not typically be identified for early intervention. We note that we examined the issue of risk status by explicitly testing whether pretest word reading moderated treatment response and found no evidence for this in the original intervention analyses; in the present paper we again test this issue with longer-term outcomes.

Attrition and final sample

During the intervention year, fifteen students moved (seven treatment and eight controls) and six treatment students (five EL and one non-EL) were randomly selected out of the study (within classroom) due to insufficient tutoring resources at their school. Hence, the final original study sample included $N = 187$ students: 93 treatment (48 of whom were EL) and 94 controls (50 EL students) from 29 classrooms across 11 schools. The final set of 11 schools (all of which were designated Title I) had student enrollments averaging 84 % minority, 74 % free/reduced lunch, 35 % transitional bilingual, and 16 % special education during the intervention year. Finally, the original EL sample represented 11 home languages: Spanish (64 % of ELs), Vietnamese (9 %), Chinese (6 %), Somali (5 %), Amharic (2 %), Arabic (2 %), French (2 %), Russian (2 %), Samoan (2 %), Tagalog (2 %), Cambodian (1 %), Oromo (1 %), Punjabi (1 %), and Tigrigna (1 %). Chi-square tests showed no evidence that treatment and control conditions differed on student

Table 1 Demographic characteristics

Characteristic	Original sample (grade 1)				Follow-up sample (grade 3[a])			
	EL, $n = 98$		Non-EL, $n = 89$		EL, $n = 95$		Non-EL, $n = 85$	
	N	%	*N*	%	*N*	%	*N*	%
Male	63	64	48	54	60	63	45	53
Bilingual	88	90	0	0	85	89	0	0
FRL	88	90	58	65	86	91	55	65
SPED	5	5	4	4	5	5	4	5
Minority	96	98	70	79	93	98	66	78
Asian	20	20	17	19	19	20	17	20
Black	11	11	37	42	11	12	35	41
Hispanic	62	63	9	10	60	63	9	11
Mixed/other	3	3	7	8	3	3	5	6

EL = English learner; bilingual = receives bilingual services; FRL = eligible for free or reduced lunch; SPED = receives special education services

[a] Five of the students were retained and in grade 2 at the end of the study

demographic characteristics (other than EL services), both within and across EL and non-EL groups. Table 1 reports sample characteristics at first (original study) and third (current study) grades, showing that the majority of both EL and non-EL students were low income.

Original study intervention procedures

Original study intervention procedures are briefly reviewed below, and described in detail in Vadasy and Sanders (2011).

Instruction lessons

Students assigned to treatment received individual phonics instruction in English, which included letter-sound correspondences, phonemic decoding and segmenting, spelling, reading irregular words, and assisted oral reading practice in decodable texts matched for consistency with the phonics elements previously taught (for further details see Vadasy, Sanders, & Peyton, 2006; Vadasy & Sanders, 2011). Each trained paraeducator tutor ($n = 25$) was provided with a set of 108 scripted lessons (with 4–8 activities per lesson). All tutoring was conducted during the school day. In a typical session, paraeducators spent 20 min on phonics activities and 10 min scaffolding students' oral reading practice. Instruction occurred 30 min/day, 4 days/week, over 20 weeks (Nov–May). Treatment students received on average 66.30 tutoring sessions ($M = 33.15$ h of tutoring, SD = 2.90), and completed 66.03 lessons (SD = 15.76). There were no significant differences between EL and non-EL students in lesson completion, tutoring attendance, or lesson coverage rate.

Instruction fidelity

Fidelity observations were conducted by trained research staff using a fidelity checklist involving a 5-point rating scale ranging from 1 (*never implements correctly*) to 5 (*always*) for each of the instructional components. Prior to field observations, researchers viewed six videotaped tutoring sessions of paraeducators implementing instruction with students. Interrater reliability was calculated based on the internal consistency of the observers' mean implementation ratings for the videotaped sessions (using observers' ratings as items and each videotape as subjects): Cronbach's alpha was .97. After establishing reliability, researchers conducted a total of 240 observations during the intervention, averaging 9.60 observations per tutor. Fidelity ratings across the 25 tutors' means averaged at 4.49 (SD = 0.36).

Follow-up study sample

We attempted to follow-up all students from the original study sample in the fall and spring of each subsequent year for 2 years after first grade. Attrition was minimized by testing any student who moved within a 30-mile radius of their original school. By the end of the study, students had dispersed from the original 11 schools to 57 different schools (ELs represented in 35 schools and non-ELs represented in 37 schools). At the end of the follow-up period, data from 96 % ($n = 180$) of the original sample (95 ELs and 85 non-ELs) were available for analyses. Again, Table 1 reports sample characteristics for the original and follow-up study samples. Without exception, t tests showed no significant differences on first-grade pretests or posttests between those who were included in the follow-up sample and those who were not, for both EL and non-EL students. Also, we note that, although we refer to the follow-up years in this study as second and third grades for brevity, five of the students we followed had been held back 1 year, ending the study in the second grade (3 ELs, 2 non-ELs). All retainees had been assigned to the control condition.

Follow-up study student assessments

All participating students were tested in English in the fall and spring of first grade, and each fall and spring for two consecutive years. All tests were individually administered by trained testers unaware of student group assignment. Norm-referenced standard scores for each of these measures were used in analyses since our research questions focus on predicting final outcomes (i.e., maintenance and moderation of treatment effects), rather than growth over time. For the purpose of the current follow-up study, we used pretest grade one word reading (a composite measure that captured children's emerging phonetic decoding and whole word reading strategies) to test for potential interactions with treatment and EL status on end of grade three outcomes.

Word reading was measured at pretest and each follow-up test using the mean standard score of the Word Attack and Word Identification subtests from the *Woodcock Reading Mastery Test-Revised/Normative Update* (WRMT-R/NU;

Woodcock, 1998). Word Attack includes 50 increasingly difficult, isolated nonsense words students must decode using standard English orthography. For this sample, internal consistencies were .88, .95, .94, .98, and .94 for ELs at pretest and each respective test wave. For non-ELs, sample reliabilities were .90, .94, .94, .98, and .96. Word Identification comprises 106 increasingly difficult, isolated sight words that students are asked to read aloud. Sample, internal consistencies were .95, .97, .97, .99, and .98 for ELs at pretest and each test wave. Non-EL sample reliabilities were .97, .97, .97, .99, and .96.

Spelling was assessed at each follow-up test using the standard score from the *Wide Range Achievement Test-IV* (WRAT- IV; Wilkinson & Robertson, 2006) Spelling subtest, which requires students to spell dictated words. Words increase in linguistic difficulty. Sample, internal consistencies were .75, .82, .94, and .85 for ELs at each test wave. For non-ELs, sample reliabilities were .79, .87, .94, and .90.

Reading comprehension was assessed at each follow-up test with the Gray Oral Reading Test-IV Form B (GORT; Wiederholt & Bryant, 2001) Comprehension subtest. Students read passages aloud, beginning at their grade level, and were asked a series of multiple-choice comprehension questions. Performance on one passage determined subsequent passage reading. Because there has been some evidence in prior research on the GORT that children with poor decoding skills can correctly answer some test items on the GORT without actually reading the passages (using their general knowledge; see Keenan & Betjemann, 2006; Keenan, Betjemann, & Olson, 2008), we employed an adapted scoring method to reflect only responses to items pertaining to passages that students read with some degree of fluency. Sample internal consistencies were .84, .84, .93, and .82 for ELs at each test wave. For non-ELs, sample reliabilities were .90, .87, .93, and .86.

Analytic strategy

We adopted multilevel modeling as our primary analytic tool for testing our research questions. Compared with traditional unilevel methods (e.g., analysis of variance and multiple regression), this more complex analysis method accounts for dependencies among measurement occasions within students as well as within schools. Multilevel models were estimated in *HLM 6.08* (Raudenbush, Bryk, & Congdon, 2004), and basic statistics were computed in *SPSS 13.0* (SPSS 1989–2004).

Establishing general growth trajectories

Prior to testing our research questions (which focus on the final follow-up test scores), we established the general growth patterns during the 1-year follow-up period. We specified straightforward three-level hierarchical growth models for each outcome in which the four measurement occasions (Level 1) were nested within students (Level 2), who were in turn were nested within their first-known schools (Level 3). We combined EL and non-EL samples for each model as one population in order to explicitly test the treatment-by-language status interaction in our final models; however, we note that the growth patterns did not differ when we examined the subgroups separately. We also note that we used the first-known school (i.e., with 11 schools) as our Level 3 identifier (rather than last-known school, with 57 schools) since (a) our final models include pretest as a predictor of follow-up outcomes, and pretests were collected while students were at their first-known school and were more likely to have greater dependencies among their school at that time, and (b) there was a slightly higher average intraclass correlation across our three follow-up outcomes for first known school compared with last known school (discussed in Results). Finally, we note that classroom membership was ignored because (a) students changed classrooms at least once (for many, it was multiple times) during the 2-year follow-up period, (b) with the small number of students relative to the large number of classroom combinations spread amongst many schools, our data do not support cross-classified modeling of classrooms, and (c) we reasoned that classroom membership would have less impact on final student outcomes (i.e., dependencies) than school membership (especially for children who moved classrooms but stayed within their school).

Again, because our emphasis in this paper is on students' end-of-study outcomes (2 years post-intervention), we fixed the intercept at the final measurement occasion (end of third grade for the majority of students), modeled standard scores (rather than raw scores, which would be appropriate if our research questions were growth-focused), and coded time for each test wave to reflect the average length between assessments (Biesanz, Deeb-Sossa, Papadakis, Bollen, & Curran, 2004). Specifically, linear change over time was coded in months as -19 = second grade fall follow-up, -12 = second grade spring, -7 = third grade fall, 0 = third grade spring. To test for quadratic change over time, the linear values were squared (the quadratic term allows for testing acceleration or deceleration in the growth pattern). We purposefully tested both a linear-only model (simpler) and linear-quadratic model (more complex) so that we could compare model fit using Chi-square likelihood ratio tests (Raudenbush & Bryk, 2002). We note that we did not expect our growth terms to necessarily capture all nuances in each trajectory; rather, our ultimate purpose was to obtain a robust estimation of final (grade three) outcomes. Finally, in all models we allowed students and schools to vary at random on the intercept and slope(s).

Final models

After establishing the best-fitting growth models, we tested our research questions testing the following predictors on the final follow-up (grade three) score (i.e., the intercept in the growth model): condition ($+1$ = treatment, -1 = control), language status ($+1$ = EL and -1 = non-EL), pretest grade one word reading (grand-mean centered and standardized using z-scores for ease of interpretation), and corresponding interactions.

Results

Observed assessment means and standard deviations at each test wave for EL and non-EL students are presented in Table 2 by condition and outcome (we display these in

Table 2 Assessment descriptive statistics for follow-up sample

Measure	EL, $n = 95$						Non-EL, $n = 85$					
	Treatment, $n = 48$			Control, $n = 47$			Treatment, $n = 43$			Control, $n = 42$		
	n	*M*	SD	*n*	*M*	SD	*n*	*M*	SD	*n*	*M*	SD
Fall grade 1 (pretest)												
Word reading	48	99.29	13.07	47	101.04	10.80	45	103.98	12.33	44	104.01	10.80
Word reading												
Fall grade 2	47	102.60	12.59	43	99.71	12.36	43	105.64	12.13	43	100.94	11.98
Spring grade 2	48	103.36	11.43	45	100.57	11.64	43	107.23	13.33	42	102.26	11.02
Fall grade 3	47	100.59	11.42	47	99.65	12.93	41	105.02	13.60	37	101.30	11.92
Spring grade 3	47	100.80	11.20	47	100.53	11.92	41	106.76	11.98	39	102.37	11.96
Spelling												
Fall grade 2	47	90.28	12.87	43	90.84	11.54	43	95.21	12.76	43	89.81	13.23
Spring grade 2	48	98.21	14.76	45	95.36	14.39	43	98.74	12.88	42	95.50	14.65
Fall grade 3	47	94.36	14.51	47	93.15	13.20	41	97.02	13.19	37	92.76	12.86
Spring grade 3	47	96.06	14.19	47	95.83	12.85	41	100.90	12.96	39	93.54	12.85
Reading compreh												
Fall grade 2	47	87.55	11.41	43	83.95	9.91	43	92.21	14.36	43	90.47	13.13
Spring grade 2	48	91.77	12.09	45	89.67	13.29	43	97.67	17.37	42	94.40	13.08
Fall grade 3	47	92.02	12.10	47	92.34	12.59	41	97.56	15.54	37	98.38	11.18
Spring grade 3	47	94.15	11.48	47	92.23	12.50	41	99.76	14.27	39	98.97	10.65

EL = English learner; word reading = mean of Woodcock Reading Mastery Test-Revised/Normative Update Word Identification and Word Attack subtest standard scores; Spelling = Wide Range Achievement Test-4 Spelling subtest; reading compreh = standard score from fluency-based adapted scoring of Gray Oral Reading Test-4 Comprehension subtest

disaggregate form for interested readers). The mean standard scores (when translated to approximate norm-referenced percentile ranks) show that, across conditions, EL children who participated in the original study ended the study near the 50th percentile on word reading, 40th percentile on spelling, and 30th percentile on reading comprehension. Non-EL children ended the study approximately 10 percentile points higher compared with their EL peers. Importantly, there was little loss over time in their word reading scores after first grade; moreover, small gains were observed on spelling and reading comprehension. Simple bivariate correlations among each predictor and grade three outcomes are reported in the Appendix; as can be seen in the correlations, pretest grade 1 word reading shows a strong positive relationship with all of the outcomes.

Intraclass correlations for school nesting

We estimated the intraclass correlations (ICCs) present in our data, which is the percent of variance in scores accounted for by levels (or factors) of a hierarchical nesting structure, using unconditional models (also known as "random intercept models") in which the total variance in scores is decomposed into respective variance components. In our case, there are three sources of variation: between-schools, between-students, and within-students (i.e., residual). Because first-known school was estimated to account for 1 % more variation in outcomes (approximately 11 % of the variance on average across measures, ranging from an ICC of .10–.12, compared with last-known school, which averaged 10 % and ranged from an ICC of .05–.13), and because our research questions involve pretest scores that were assessed when students were located at their first-known school, all of our subsequent statistical models use first-known school as the Level 3 identifier. (Nevertheless, we note that using last-known school as the Level 3 identifier made no substantive difference in model results.)

Establishing growth patterns across grades 2 and 3

Using likelihood ratio tests to compare linear growth models (additive change over time) with quadratic growth models (change that accelerates or decelerates over time), we found that a linear fit was better suited to students' word reading trajectories, whereas a quadratic model was a better fit for spelling and reading comprehension. The predicted and observed means for each outcome are displayed in Fig. 1. We note that preliminary analyses showed that both EL and non-EL children had similar trajectory pattern, although EL children began and started the follow-up period lower than non-EL children.

Longer-term effects on end of grade 3 outcomes

The focus of our model results (presented in Table 3), which are the fixed effects tests on the intercept (the end of grade three outcomes), show significant positive longer-term effects for treatment as well as pretest word reading on all end of grade three outcomes (our first research question). One significant negative effect for language status and two significant interactions with language status (i.e., EL vs. non-EL) were also present (our second research question). No three-way interactions were detected (our third research question). We discuss each set of effects in turn below.

Treatment effects

Positive grade one treatment effects were found to be maintained 2 years after intervention. Controlling for language status and pretest, treatment students were estimated to average 4.44 points higher than controls on word reading, 3.66 points higher on spelling, and 2.60 points higher on reading comprehension. We computed an approximate treatment effect size, d^*, by dividing the difference in treatment and control predicted values by the square root of the sum of the model-estimated variance components, and found that these three effects were small to modest in size, with $d^* = 0.45$, 0.36, and 0.24, for word reading, spelling, and reading comprehension, respectively.

EL effects

Holding condition and pretest constant, by the end of grade three, EL students were only significantly lower than non-EL students on reading comprehension (by 3.22

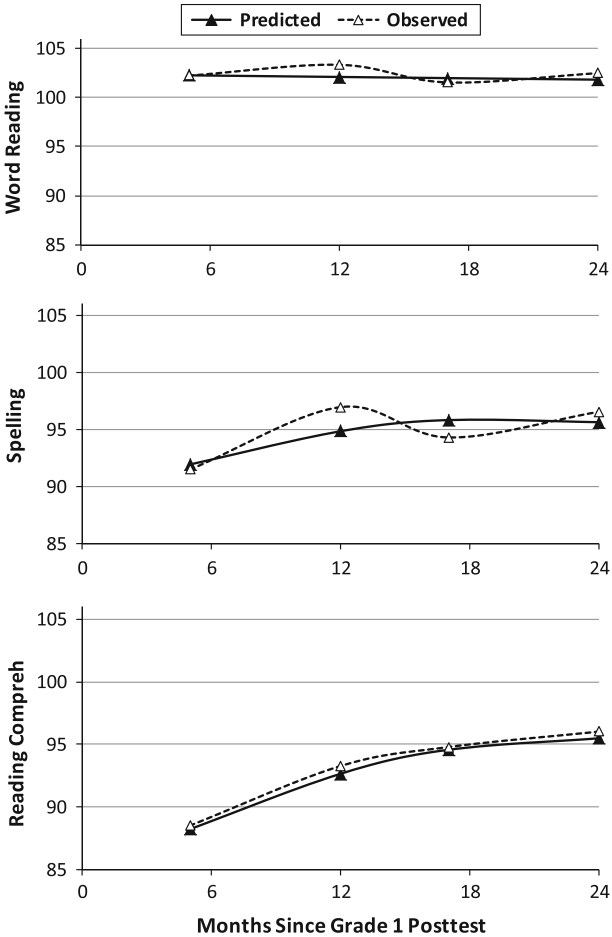

Fig. 1 Growth model-predicted (fitted) and observed change/month since grade 1 posttest

points, $d^* = -0.29$) (although there was a trend for EL students to underperform on the other two measures).

Grade 1 pretest word reading effects

Irrespective of treatment or language status, grade one pretest word reading was a significant predictor of end of grade three literacy outcomes. The model results estimate that students who were relatively high in word reading skills at the beginning of first grade (i.e., one standard deviation higher than average) were estimated to be 8.42 points higher than average performing peers at the end of third grade on word reading, 9.21 points higher on spelling, and 7.73 points higher on reading comprehension. This finding is also reflected in the simple bivariate correlations that show a strong relationship between beginning of grade one word reading and each of the end of grade three outcomes.

Table 3 Final model fixed effects results

Fixed effects	Word reading			Spelling			Reading compreh		
	Coeff	SE	*df*	Coeff	SE	*df*	Coeff	SE	*df*
Intercept: end of grade 3 mean	101.88	0.84	10***	95.61	0.92	10***	95.48	0.76	10***
Linear change/month since fall grade 1	−0.03	0.05	10	−0.17	0.10	10	−0.03	0.13	10
Quadratic change/month since fall grade 1	–			−0.02	0.00	10**	−0.02	0.01	10*
Predictors of intercept (end of grade 3)									
Condition (1 = treatment, −1 = control)	2.22	0.64	179***	1.83	0.61	179**	1.30	0.66	179*
Language status (1 = EL, −1 = non-EL)	−0.20	1.02	179	0.66	0.74	179	−1.61	0.78	179*
Pretest grade 1 word reading (z-score)	8.42	0.50	179***	9.21	0.60	179***	7.73	0.75	179***
Condition * language	−0.65	0.50	179	−1.40	0.60	179*	0.32	0.44	179
Condition * pretest	−0.60	0.57	179	−0.96	0.51	179	−0.27	0.59	179
Language * pretest	−0.66	0.60	179	−0.95	0.65	179	−1.93	0.84	179*
Condition * language * pretest	−0.09	0.50	179	0.53	0.46	179	−0.62	0.56	179

$N = 180$; Pretest = standardized word reading pretest in fall grade 1; word reading = mean of Woodcock Reading Mastery Test-Revised/Normative Update Word Identification and Word Attack subtest standard scores; Spelling = Wide Range Achievement Test-4 Spelling subtest; reading compreh = standard score from fluency-based adapted scoring of Gray Oral Reading Test-4 Comprehension subtest

* $p \le .05$, ** $p \le .01$, *** $p \le .001$

*EL*treatment and EL*pretest interactions*

Language status negatively interacted with condition on spelling, and negatively interacted with pretest on reading comprehension. Using the model-estimated values, we graphed predicted end of grade three means by condition, language status, and pretest level (in standard deviations from the mean) to better understand the nature of these two interactions.

With regard to the language*condition interaction detected on spelling (middle bar graph in Fig. 2): the longer-term treatment effect was clearly lower in magnitude for EL students compared to non-ELs. Holding pretest constant, the estimated difference between treatment and control EL students was only +0.86 points ($d^* = 0.09$) on spelling whereas the predicted treatment–control difference was +6.47 ($d^* = 0.64$) points for non-EL students. Another way of interpreting this finding is that the early treatment effects found on spelling at grade one posttest were not maintained for EL students by grade three.

With regard to the language*pretest interaction found for reading comprehension (last bar graph in Fig. 2): the grade three advantage from having higher grade one

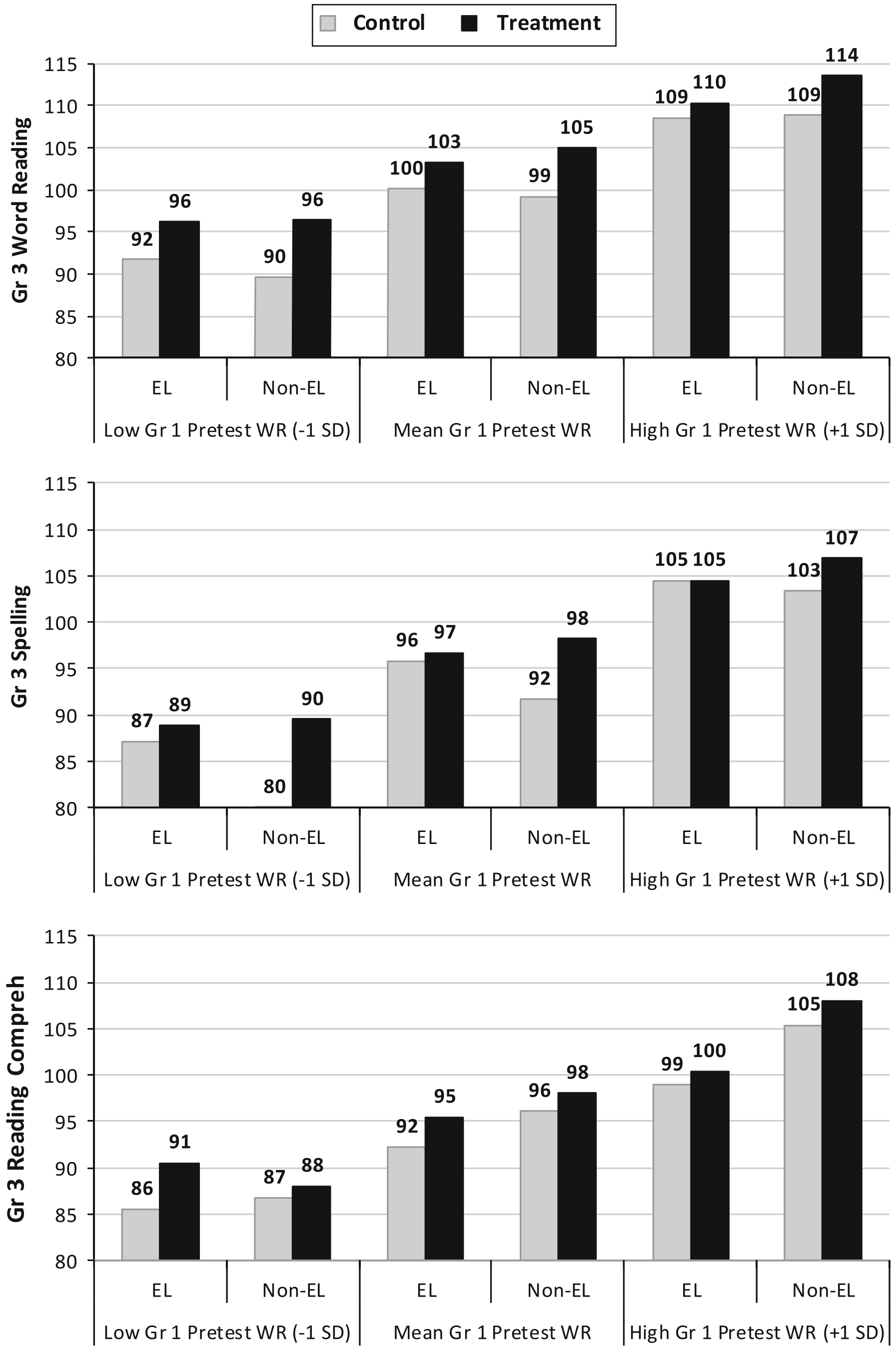

Fig. 2 Model-predicted (fitted) mean end of grade 3 outcomes by condition, language status, and levels of pretest grade 1 word reading. *Note*: Significant two-way interactions between condition and language status on spelling, and between language status and pretest on comprehension. No other significant interactions detected

pretest skills is much more pronounced for non-ELs compared with ELs. Specifically, EL students with mean first grade pretest scores (within one standard deviation of average, $n = 64$ in the sample) were predicted to average 93.87 points on grade three reading comprehension whereas ELs with relatively high pretest scores ($n = 12$ in the sample) were predicted to average 99.67 points, an advantage of 5.80 points. Comparatively, non-ELs who were higher-skilled in grade one ($n = 17$ in the sample) were estimated to average 106.75 points, which is an advantage of 9.66 points over peers with typical grade one pretest levels ($n = 60$ in the sample). Another way of interpreting this is that EL students' grade one word reading skills are less predictive of grade three comprehension compared with non-EL students.

*Lack of condition*EL*treatment interactions*

Finally, the lack of a significant three-way interaction among condition, language status, and pretest word reading shows that treatment effects, for either EL or non-EL students, were not dependent on a particular level of word reading skills in the fall of grade one.

Discussion

In our original study, we found large, significant treatment benefits of a code-oriented intervention for first-grade English learner (EL) and native English speaking (non-EL) students on a host of literacy outcomes, including word reading, spelling, and reading comprehension. The current study, which follows the original sample participants into grades two and three, shows that treatment benefits were maintained across all three outcomes for non-ELs, and for two of the three measures (all but spelling) for EL children. Further, pretest word reading scores did not moderate treatment effectiveness for either EL or non-EL students: this indicates that the intervention had benefits for students who might typically be identified as the lower-performers in their classroom and served in a similar supplemental intervention (i.e., not only those students in the absolute lowest quartile of reading skill, but inclusive of those in the lower half of their classrooms). Nevertheless, it is interesting to note that early grade one word reading skills are significantly less predictive of grade three reading comprehension for ELs in particular compared to non-ELs. Effect sizes were modest in size (across EL and non-EL students: $d^* = .45$, .36, and .24 for word reading, spelling, and reading comprehension, respectively) and similar to earlier reports on longer term outcomes of early intervention for English learners.

The students in this study were in first grade when they received intervention. Across word reading, spelling, and comprehension outcomes there were modest treatment effects at 2-year follow up. In a parallel study, a cohort of EL and non-EL kindergarten students participated in a similar intervention, also with a 2-year follow up. When intervention was provided at kindergarten, there were significant treatment effects for both EL and non-EL students at posttest and at second-grade follow up. Those results have been reported elsewhere (Vadasy & Sanders, 2010; 2012) and considered together with the findings for the significant role of first-grade word reading we report here, support interventions that develop strong code-oriented foundations for kindergarten students.

We examined the influence of first-grade pretest word reading skill, using a covariate that reflects student skills in phonological and orthographic skills at grade one entry, and that incorporates alphabet and phonological awareness skills that are robust predictors of reading outcomes. Pretest word reading predicted all grade three outcomes for both English learners and native English speaking students. Treatment effects in spelling were only present for non-EL students. The findings for pretest word reading effects underscore the importance of early literacy skills in predicting reading outcomes for native English speaking and English language learners (e.g., Juel, 1988; Lindsey, Manis, & Bailey, 2003). Pretest word reading has less influence on grade three comprehension for English learners, perhaps reflecting the decrement in comprehension observed for English learners compared to native-English speaking children that begins about third grade (e.g., Nakamoto et al., 2007). The significant influence of pretest word reading on third-grade spelling outcomes may reflect the cumulative phonological and orthographic knowledge that becomes necessary for accurate English spelling (e.g., Burgess & Lonigan, 1998; Apel & Masterson, 2001).

Results show that native English speaking children who received individual code-oriented tutoring in first grade continued to demonstrate treatment benefits 2 years later on all three outcomes, with an average predicted effect size of 0.47. Comparatively, English learners were only found to have longer term effects for word reading and reading comprehension, but not spelling (average effect size was 0.23 across all three measures). However, it is important to note that our sample selection procedures using relative risk rather than a cut-score risk for study entry allowed for students with less risk than most intervention studies to participate (sample pretest grade one word reading levels were near 60th percentile). Nevertheless, grade one word reading did not moderate treatment effects for grade three word reading.

Overall, findings provide support for longer-term benefits of first-grade intervention, although to a somewhat lesser degree for EL students. Explicit instruction in code-oriented (alphabetic and phonics) skills offers students a foundation in word recognition skills (decoding and word identification) and understanding of how to represent spoken words in accurate spellings. The instruction used in the treatment condition covered all single-letter and many two-letter spelling patterns, as well as spellings of irregular sight words. By the end of third grade, it appears that native English speaking students had an understanding of basic systematic spelling correspondences, including flexible spelling rules for irregular words, and possible greater exposure to printed English for self-teaching effects (Share,) that may lead to remembering accurate word spellings. The strong code emphasis in the intervention may have strengthened1995 the reciprocal relationship between decoding and encoding skills for the native English speaking children in the study.

Study limitations

This study carries with it the limitations outlined in the original intervention year study (Vadasy & Sanders, 2011). The largest caveat is that the sample selected for participation involved a relative-performance (bottom half of classroom), rather than cut-score, criterion at first-grade screening to allow adequate sample sizes for testing treatment effects for both English learner and native English-speaking children. Pretest word reading means (near 60th percentile for native speaking children and 50th percentile for children who are English learners) show that our sample included false positives—children who were not necessarily in need of a code-based-level intervention. Inclusion of these students limits the generalizability of our findings to typical intervention populations; however, we note that we found no two-way treatment interactions with word reading levels on outcomes for children who were either English learners or native English speakers. We also note that, in an extended series of earlier studies, the efficacy of the intervention used in the current study was demonstrated with primarily native English-speaking students who performed in the lower quartile in word reading at the beginning of kindergarten or first grade (Vadasy & Sanders, 2008; Vadasy, Sanders, & Peyton, 2006).

A second major limitation for this study is the lack of information on English learners' L1 and English language proficiency. English learners English (L2) literacy development has been shown previously to be mediated by their L1 early literacy skills (e.g., L1 phonological awareness and vocabulary) (see reviews by Dressler & Kamil, 2006; Geva & Genesee, 2006). The large number of home languages represented in this sample, however, prevented collection of data on children's L1 language and literacy skills and home language and literacy influences.

Finally, we note the original study design involved a pull-out treatment compared with business-as-usual classroom instruction. This type of design is always susceptible to Hawthorne effects as it is impossible to disentangle receipt of tutoring with tutoring instruction. Nevertheless, we observed longer-term treatment effects that are unlikely to be the result of simply being pulled out for tutoring.

Conclusion

An important finding is that across experimental conditions, word reading skills at entry to grade one strongly predicted word reading, spelling, and reading comprehension outcomes at end of grade three for both English learner and native English speaking students. The powerful relationship between initial word reading skills (prior to classroom or intervention instruction) and later literacy skills supports the benefits of early literacy interventions and experiences provided prior to first grade. Additional support comes from a similar follow-up study with kindergarten children by Vadasy and Sanders (2012): in that study, lower-skilled EL and non-EL children with relatively greater alphabetic skills in the fall of kindergarten had significantly higher performance on word reading, spelling, and reading comprehension outcomes at end of grade two. The findings from these parallel studies, taken together, suggest that pre-requisite literacy skills (alphabetic, phonological, and oral language skills) that support acquisition of word reading at school entry allow relatively lower-skilled children to benefit from both regular classroom instruction and supplemental intervention in reading skills.

Although there was a treatment advantage on reading comprehension at the end of grade three for English learners, they remained below grade level in comprehension. It is well established that English learners enter school with initial differences in vocabulary, reported to be up to a 2-year developmental lag in receptive and expressive vocabulary (Hutchinson, Whiteley, Smith, & Connors, 2003), that increases over time (Biemiller & Slonim, 2001). English learners experience greater difficulties than native English speaking peers in reading comprehension beginning in the primary grades and increasing and extending through high school (Lesaux, Koda, Siegel, & Shanahan, 2006; Nakamoto et al., 2007). The treatment advantage observed for comprehension may reflect the effect of 33 h of individual tutoring time with an English-speaking adult. We suspect the intensity of this language interaction was an unusual opportunity for many of the young English learners in this study. Others have reported that school is the primary source of English linguistic input for young English learners (Bowers & Vasilyeva, 2011). The treatment advantage may also reflect carefully scaffolded interactions that occurred during the phonics instruction as well as the storybook reading activity when tutors offered incidental but individualized support in reading comprehension. Closing the vocabulary and reading comprehension gap for English learners may require dedicated language and vocabulary interventions (Dickinson, Golinkoff, & Hirsh-Pasek, 2010; Hutchinson et al., 2003; Manyak & Bauer, 2009; Verhoeven, 2011).

Overall the findings replicate research on the word-level benefits of code-based interventions for native English speaking students, and the findings of others for English learners (Vaughn et al., 2008). The present study adds data on a sample of linguistically diverse EL students to complement most U.S. studies of EL students conducted in communities with predominantly Spanish-speaking student populations. Code-oriented interventions may be most beneficial for English learners, as well as at-risk native English speaking students, when they explicitly develop and build connections between language comprehension and word reading skills. Studies of longer-term effects are needed to examine students' response to instruction over time and as language comprehension skills that are interdependent with early word reading assume a central role in skilled reading comprehension.

Acknowledgments The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education through Grant #R305A070324. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Appendix

See Table 4.

Table 4 Observed bivariate correlations among follow-up study variables

Variable	1	2	3	4	5	6
Grade 1 predictors						
1. Condition (1 = treatment, −1 = control)	–					
2. Language status (1 = EL, −1 = non-EL)	−.02	–				
3. Pretest grade 1 word reading	−.03	**−.17**	–			
End of grade 3 outcomes						
4. Word reading	.09	**−.17**	**.58**	–		
5. Spelling	.13	−.05	**.63**	**.80**	–	
6. Reading compreh	.06	**−.25**	**.47**	**.63**	**.55**	–

$N = 174$ observed scores at grade 3 posttest due to missingness. Word reading = mean of Woodcock Reading Mastery Test Revised/Normative Update Word Identification and Word Attack subtest standard scores; Spelling = Wide Range Achievement Test-4 Spelling subtest; reading compreh = adapted Gray Oral Reading Test-4 Comprehension subtest. Significant correlations ($p < .05$) are boldfaced

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Apel, K., & Masterson, J. J. (2001). Theory-guided spelling assessment and intervention: A case study. *Language, Speech, and Hearing Services in the Schools, 32*, 182–195.

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

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods, 9*(1), 30–52.

Bowers, E. P., & Vasilyeva, M. (2011). The relation between teacher input and lexical growth of preschoolers. *Applied Psycholinguistics, 32*, 221–241.

Burgess, S. R., & Lonigan, C. J. (1998). Bidirectional relations of phonological sensitivity and prereading abilities: Evidence from a preschool sample. *Journal of Experimental Child Psychology, 70*, 117–141.

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

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

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

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities. *Scientific Studies of Reading, 3*, 331–361.

Catts, H. W., Hogan, T. P., & Adolf, S. M. (2005). Developmental changes in reading and reading disabilities. In H. Catts & A. Kamhi (Eds.), *The connections between language and reading disabilities* (pp. 25–40). Mahwah: Lawrence Erlbaum Associates.

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

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

Demont, E., & Gombert, J. E. (1996). Phonological awareness as a predictor of recoding skills and syntactic awareness as a predictor of comprehension skills. *British Journal of Educational Psychology, 66*, 315–332.

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

Dickinson, D. K., Golinkoff, R. M., & Hirsh-Pasek, K. (2010). Speaking out for language: Why language is central to reading development. *Educational Researcher, 39*, 305–310.

Dressler, C., & Kamil, M. L. (2006). First-and second-language 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. 197–238). Mahwah, NJ: Lawrence Erlbaum.

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

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.

Fry, R., & Gonzales, F. (2008). *One-in-five and growing fast: A profile of Hispanic public school students.* Washington, DC: Pew Hispanic Center.

Gersten, R., & Baker, S. K. (2003). English language learners with learning disabilities. In H. L. Swanson, K.R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 94–109). New York: Guilford.

Geva, E., & Genesee, F. (2006). First-language oral proficiency and second-language 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. 185–196). Mahwah, NJ: Lawrence Erlbaum.

Gottardo, A. (2002). The relationship between language and reading skills in bilingual Spanish English speakers. *Topics in Language Disorders, 22*, 46–70.

Gottardo, A., & Mueller, J. (2009). Are first- and second-language factors related in predicting second-language reading comprehension? A study of Spanish-speaking children acquiring English as a second language from first to second grade. *Journal of Educational Psychology, 101*, 330–344.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6–10.

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.

Harlaar, N., Cutting, L., Deater-Deckard, K., DeThorne, L., Justice, L., Schatschnieder, C., et al. (2010). Predicting individual differences in reading comprehension: A twin study. *Annals of Dyslexia, 60*, 265–288.

Hemphill, F. C., & Vanneman, A. (2011). *Achievement gap: How Hispanic and white students in public schools perform in mathematics and reading on the National Assessment of Educational Progress (NCES 2011–459).* Washington, DC: National Center for Educational Statistics, Institute of Education Sciences, U.S. Department of Education.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2*, 127–160.

Hutchinson, J. M., Whiteley, H. E., Smith, C. D., & Connors, L. (2003). The developmental progression of comprehension-related skills in children learning EAL. *Journal of Research in Reading, 26*, 19–32.

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

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

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12*, 281–300.

Lesaux, N. K., Koda, K., Siegel, L. S., & Shanahan, T. (2006a). 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. 75–112). Mahwah, NJ: Lawrence Erlbaum.

Lesaux, N. K., Lipka, O., & Siegel, L. S. (2006b). Investigating cognitive and linguistic abilities that influence the reading comprehension skills of children from diverse linguistic backgrounds. *Reading and Writing, 19*, 99–131.

Lindsey, K. A., Manis, F. R., & Bailey, C. E. (2003). Prediction of first-grade reading in Spanish-speaking English-language learners. *Journal of Educational Psychology, 95*, 482–494.

Mancilla-Martinez, J., Kieffer, M. J., Biancarosa, G., Christodoulou, J. A., & Snow, C. E. (2011). Investigating English reading comprehension growth in adolescent language minority learners: Some insights from the simple view. *Reading and Writing: An Interdisciplinary Journal, 24*, 339–354.

Manis, F. R., Lindsey, K. A., & Bailey, C. E. (2004). Development of reading in grades K-2 in Spanish-speaking English-language learners. *Learning Disabilities Research and Practice, 19*, 214–224.

Manyak, P. C., & Bauer, E. B. (2009). English vocabulary instruction for English learners. *The Reading Teacher, 63*, 174–176.

Nakamoto, J., Lindsey, K. A., & Manis, F. R. (2007). A longitudinal analysis of English language learners' word decoding and reading comprehension. *Reading and Writing: An Interdisciplinary Journal, 20*, 691–719.

Oller, D. K., & Eliers, R. E. (2002). *Language and literacy in bilingual children*. Clevedon, UK: Multilingual Matters.

Paez, M. M., Tabors, P. O., & Lopez, L. M. (2007). Dual language and literacy development of Spanish-speaking preschool children. *Journal of Applied Developmental Psychology, 28*, 85–102.

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

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

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

Shankweiler, D., & Liberman, I. (1972). Misreading: A search for causes. In J. Kavanaugh & I. Mattingly (Eds.), *Language by ear and by eye: The relationship between speech and reading* (pp. 293–317). Cambridge, MA: MIT Press.

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

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

SPSS, Inc. (1989–2004). *Statistical Package for the Social Sciences 13.0*. New York, NY: McGraw-Hill.

Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology, 40*, 7–20.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1997). Approaches to the prevention and remediation of phonologically-based reading disabilities. In B. A. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (pp. 287–304). Mahwah, NJ: Erlbaum.

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

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

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trial with paraeducator implementers. *Journal of Educational Psychology*, *98*(3), 508–528.

Vadasy, P. F., & Sanders, E. A. (2011). Efficacy of supplemental phonics-based instruction for low-skilled first graders: How language minority status and pretest characteristics moderate treatment response. *Scientific Studies of Reading,*. doi:10.1080/10888438.2010.501091.

Vadasy, P. F., & Sanders, E. A. (2012). Two-year follow up of a kindergarten phonics intervention for English learners and native English speakers: Contextualizing treatment impacts by classroom literacy instruction. *Journal of Educational Psychology,*. doi:10.1037/10028163.

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trial with paraeducator implementers. *Journal of Educational Psychology, 93*, 508–528.

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

Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S., et al. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *The Elementary School Journal, 107*, 153–180.

Verhallen, M., & Schoonen, R. (1993). Vocabulary knowledge of monolingual and bilingual children. *Applied Psycholinguistics, 14*, 344–363.

Verhoeven, L. (1990). Acquisition of reading in a second language. *Reading Research Quarterly, 25*, 90–104.

Verhoeven, L. (2000). Components in early second language reading and spelling. *Scientific Studies of Reading, 4*, 313–330.

Verhoeven, L. (2011). Second language reading acquisition. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. P. Afflerbach (Eds.), *Handbook of reading research* (Vol. IV, pp. 661–683). New York, NY: Routledge.

Wagner, R., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing*. Austin, TX: Pro-Ed.

Wiederholt, J. L., & Bryant, B. R. (2001). *Gray oral reading tests-IV*. Austin, TX: Pro-Ed.

Wilkinson, G. S., & Robertson, G. J. (2006). *The wide range achievement test-IV*. Lutz, FL: Psychological Assessment Resources, Inc.

Woodcock, R. (1998). *Woodcock reading mastery test-revised/normative update*. Circle Pines, MN: American Guidance Service.

Yeong, S. H. M., & Liow, S. J. R. (2011). Cognitive-linguistic foundations of early spelling development in bilinguals. *Journal of Educational Psychology*. Advance online publication 21 Feb 2011. doi: 10.1037/a0022437.