Journal of Educational Psychology
2007, Vol. •, No. •, 000–000

0022-0663/07/$12.00 DOI: 10.1037/0022-0663.•.•.000

Growth in Reading Skills of Children From Diverse Linguistic Backgrounds: Findings From a 5-Year Longitudinal Study

Nonie K. Lesaux
Harvard Graduate School of Education;

Andre A. Rupp
Humboldt State University

Linda S. Siegel
University of British Columbia

This article reports on the results of a longitudinal investigation of the reading development of a sample of 824 children (406 girls, 418 boys). The sample included 689 native English-speaking (L1) children and 135 English-language learners (ELLs) representing 33 different native languages. In kindergarten and 4th grade, children's word reading, spelling, phonological processing, syntactic awareness, and working memory skills were assessed with standardized and experimental measures. In addition, word reading was assessed from kindergarten through 4th grade, and reading comprehension in 4th grade. Comparisons of reading skills between the ELLs and the L1 speakers demonstrated that despite slightly lower performance of the ELLs on several kindergarten tasks, differences at 4th grade were negligible. Fourth-grade word reading was predicted by the same kindergarten tasks for both language groups, and prediction of reading comprehension differed by only 1 task. Finally, the trajectory of word reading was nonlinear for both groups, although predictors of this trajectory differed between groups. The findings suggest that early identification models established through research with L1 speakers are appropriate for identifying ELLs at risk for reading difficulties.

Keywords: English-language learners, reading development, native English speakers, growth modeling

The majority of the research that has investigated children's reading development has concentrated largely on children who are native English speakers (L1 speakers; for relevant reviews, see Adams, 1990; National Institute of Child Health and Human Development, 2000; Snow, Burns, & Griffin, 1998). However, an increasing number of children, referred to as English-language learners (ELLs), are entering kindergarten with little or no exposure to English and, upon school entry, are immersed in mainstream English classrooms (Development Associates, 2003; National Center for Education Statistics, 2004). These children need to acquire oral language and literacy skills in a second language and must do so with considerable efficiency in order to acquire native levels of proficiency and, in turn, to reach grade-level proficiency in other academic areas.

For ELLs, the development of literacy skills in a second language is arguably even more challenging than for native speakers. The same array of word-level and text-level skills must be learned, though such learning may begin at a later age, perhaps without the same level of foundation in the cognitive and linguistic precursors to literacy. Therefore, it is important to establish an understanding of the development of reading for this group of learners and, more specifically, to determine the ways in which this process compares to that of their L1 peers.

The impact of ELLs' second-language learning on their reading development and academic achievement has received growing attention in recent years from researchers, practitioners, and policymakers (e.g., August & Hakuta, 1997; August & Shanahan, 2006; Crawford, 2004; Cummins, 2000; National Center for Education Statistics, 2005). However, in contrast to the research base on literacy development for L1 speakers, there is a dearth of longitudinal research that has examined the trajectories of reading development for ELLs. For example, in a review of studies on language-minority learners' literacy development conducted between 1980 and 2002 (N = 68) only 6 were longitudinal; moreover, all except 1 followed children only during the primary-grade years (for the review, see Lesaux, Koda, Siegel, & Shanahan, 2006). To our knowledge, only four longitudinal studies have been published since 2002 (e.g., Gerber et al., 2004; Manis, Lindsey, & Bailey, 2004; Stuart, 2004; Swanson, Saez, & Gerber, 2006). Of

Nonie K. Lesaux, Human Development and Psychology, Harvard Graduate School of Education, Harvard University; Andre A. Rupp, Institute for Quality Management in Education, Humboldt State University; Linda S. Siegel, University of British Columbia.

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to Linda S. Siegel, and in part by fellowships from the Michael Smith Foundation for Health Research, the Social Sciences and Humanities Council of Canada, and the Spencer Foundation granted to Nonie K. Lesaux.

We thank Kimberly McMahon for assistance with the preparation of the article as well as Robin Brayne, Jay Merilees, Larry Johnson, Cathy Molinski, Heather Calder, Ann Tarves, Audrey Hobbs-Johnson, Lorna Bennett, Fran Blake, Laurie McDonald, Pam Ottley, Linda Pearson, Mary Tennant, Brian Ward, and Rochelle Watts for their support of the project. Finally, we especially thank the principals, teachers, and students who participated in this study.

Correspondence concerning this article should be addressed to Nonie K. Lesaux, Harvard Graduate School of Education, Larsen 319, 14 Appian Way, Cambridge, MA 02138. E-mail: lesauxno@gse.harvard.edu

these studies, only the study by Swanson and colleagues followed children past second grade; in their study, the ELLs were followed from first through third grade.

Thus, the great majority of studies conducted to investigate the reading development of ELLs have been cross-sectional in nature, and most have been conducted with children in the primary grades. Furthermore, these studies have typically been conducted with a sample of ELLs but without a comparison group of L1 speakers. Although the findings of these studies have shed light on early reading skills and the factors that predict these skills at a single point in time, they have failed to provide insight into the process of reading development for ELLs over time and how it compares to that of their L1-speaking peers.

The present study was designed to add to this growing body of literature by comparing the development of reading for a group of ELLs and their L1-speaking peers, followed from kindergarten through fourth grade. The study design enabled analyses beyond straightforward comparisons of outcomes on reading and related tasks; thus, growth trajectories of word reading for ELLs and L1 speakers were investigated. Knowledge of ELLs' trajectories of reading and the skills that influence these trajectories is needed in order to establish a common understanding of the process of reading development for these learners, common expectations for their achievement, and, in turn, to inform the design of effective instructional methods to promote academic success.

Development of ELLs' Component Skills of Reading

Consistent with research conducted with L1 speakers (for a review, see Siegel, 1993), cognitive and linguistic skills, such as phonological processing, working memory, and syntactic awareness, have been identified as important for ELLs' reading development, at least in the early stages of reading acquisition (e.g., Everatt, Smythe, Adams, & Ocampo, 2000; Geva & Siegel, 2000; Limbos & Geva, 2001; Manis et al., 2004; Swanson et al., 2006). For example, the findings from research that focused on the role of phonological processing for ELLs in the primary grades suggest that it is a reliable predictor of word reading and, in some cases, a more reliable predictor than oral language proficiency (Durgunoglu, Nagy, & Hancin-Bhatt, 1993; Geva, Yaghoub-Zadeh, & Schuster, 2000; Gottardo, 2002; Limbos & Geva, 2001; Muter & Diethelm, 2001). Specifically, whereas ELLs who are typical readers have well-developed phonological processing skills, those ELLs with word reading difficulties in the primary grades demonstrate poor phonological processing skills (Gottardo, 2002; Manis et al., 2004; Swanson et al., 2006; Wade-Woolley & Siegel, 1997). This finding is important, because lack of oral language proficiency in the language of instruction is often one of the reasons for underestimating a child's reading ability in his or her second language (e.g., Limbos & Geva, 2001; Moll & Diaz, 1985). There is a general tendency within schools to overlook or delay addressing the possibility that an ELL is having word reading difficulties due to language processing skills, such as phonological processing skills, until that ELL's oral language proficiency is further developed (Limbos & Geva, 2001).

Similarly, research conducted with ELLs has suggested that working memory plays an important role for predicting individual differences in ELL reading abilities (e.g., da Fontoura & Siegel, 1995; Geva & Siegel, 2000), which is in line with findings from research conducted with L1 speakers (e.g., Chiappe, Hasher, & Siegel, 2000; McDougall, Hulme, Ellis, & Monk, 1994; Swanson & Alexander, 1997). Consequently, whereas ELLs who are typical readers have well-developed working memory skills, those ELLs classified as reading disabled show poor working memory skills (e.g., da Fontoura & Siegel, 1995; Swanson et al., 2006).

Unlike research conducted with ELLs on phonological processing and working memory, research that has investigated the role of syntactic awareness—an understanding of the grammatical structure of the language—has demonstrated that, in contrast to L1 speakers, this is an area of difficulty even for ELLs with word reading skills in the average range. Given that the studies that have demonstrated ELLs' difficulties with syntactic awareness have been conducted with children in the primary grades, the question remains whether—with increased years of English instruction and English proficiency—these learners eventually attain syntactic awareness skills that are comparable to those of their L1 peers. If this is not the case, the difficulties with these skills may eventually hinder reading comprehension, given its relationship with oral language abilities (García, 1991, 2000; RAND Reading Study Group, 2002).

Beyond those questions about the role of syntactic awareness in ELLs' reading development, research must address the overarching question as to how ELLs' early literacy skills relate to later literacy outcomes, particularly in the years following the primary grades. In order to inform researchers' understanding of the developmental process of reading for ELLs and, in turn, effective models of instruction, it is important to look longitudinally to determine reading outcomes, particularly with respect to reading comprehension.

Comparisons of L1 Speakers' and ELLs' Reading Development

The findings from studies investigating ELLs' reading development, which have included a comparison group of L1-speaking peers, have indicated that the two groups typically perform at similar levels on measures of phonological processing, word reading, and spelling (for a review, see Lesaux et al., 2006). Specifically, a meta-analysis based on data from studies that compared language-minority learners' and L1 speakers' word reading skills indicated that these groups were equivalent in word reading accuracy (effect size = –.09, *ns*). These studies were conducted with children in the elementary school years from varying home-language backgrounds in different linguistic and demographic contexts in Canada, England, the Netherlands, and the United States. Together, the language-minority learners in these studies included native speakers of Punjabi, Urdu, Arabic, Italian, Portuguese, and various Asian languages; some studies described the sample of ELLs as one with individuals from a myriad of language backgrounds.

The findings of comparative studies on reading comprehension have painted a very different picture from those focused on word reading and spelling. The few available studies that have compared language-minority learners with their native-speaking peers—conducted primarily in the Netherlands—yielded highly consistent results, indicating that the reading comprehension performance of language-minority learners is an area of weakness. Where there has been a comparison group of L1 speakers, language-minority

tapraid5/zcz-edu/zcz-edu/zcz00307/zcz2073d07g | sangreyj | S=4 | 6/28/07 | 21:40 | Art: 2004-0215-RRRRR

learners' performance levels have been well below those of their L1 speaking peers (e.g., Carlisle, Beeman, Davis, & Spharim, 1999; Hutchinson, Whiteley, Smith, & Connors, 2003; Lindsey, AQ: 3 Manis, & Bailey, 2003; Reese, Garnier, Gallimore, & Goldenberg, 2000; Verhoeven, 1990, 2000). This finding is not surprising given that reading comprehension is compromised when skills such as oral language proficiency and relevant prior knowledge are insufficiently developed to support understanding of a text. Although such factors can also compromise reading comprehension among native speakers, language-minority learners reading in a second language are more likely to have underdeveloped skills in these areas.

A shortcoming of this area of research, however, is that few studies have empirically examined those skills that have an influence on reading comprehension for this group of learners. Establishing the early predictors of ELLs' reading comprehension—and whether they differ from those predictors for L1 speakers—deserves significant attention from researchers.

Published Reports on ELLs' Reading Achievement

It is important to note that, in addition to the developmental research referred to previously, there are a number of published reports based on district-, state-, and federal-level data on the reading achievement of ELLs (e.g., Kindler, 2002; Thomas & Collier, 1997, 2002) Although often relevant for informing policymakers' decisions, data collected for progress monitoring and accountability are of limited value in advancing researchers' understanding of the reading development of ELLs.

First, these reports always reflect data from only one subset of the ELL population, namely those learners who meet the district criteria for the limited English proficiency (LEP) classification. These ELLs are considered to have the lowest levels of English proficiency and thus receive additional support for language development. Yet other ELLs whose academic achievement is not represented in these data sets include learners who, although initially classified as LEP, progressed in language proficiency to a point where they lost their LEP designation. Also not included are learners whose English proficiency was sufficiently developed such that, upon school entry, they were not classified as LEP. It is crucial that developmental research is not limited to one specific subset of the population of language-minority learners.

Another difficulty that is related to using school-based data sets is that many of these reports, despite presenting longitudinal data, are retrospective in nature and include only those learners for whom there are complete data. Given the relatively transient nature of this population, it is especially important to be able to analyze the effects of attrition on the results.

Finally, most of these reports do not feature standardized assessments of language skills and reading development that are norm referenced and instead report on results from standards-based large-scale assessments (e.g., California English Language Development Test, National Assessment of Educational Progress). Therefore, although proficiency profiles of ELLs in these reports are compared to those of L1-speaking peers, they do not allow for comparisons with the findings from many other studies, particularly those developmental studies that use standardized norm-referenced measures of language skills and reading development.

Present Study

For the many aforementioned reasons—substantive and methodological—there has been a press for research on ELLs' reading development, particularly for longitudinal research that focuses on establishing normative trajectories in this area. To provide insight into developmental trajectories of ELLs in comparison to their L1-speaking peers and to identify early predictors of later reading outcomes, researchers must conduct studies with samples of ELLs with a range of achievement levels and follow them over time using standardized measures (National Institute of Child Health and Human Development & U.S. Department of Education, 2003). Thus, the present study is a longitudinal investigation of the reading development of a sample of ELLs and their L1-speaking peers—instructed only in English—followed from kindergarten through fourth grade. To our knowledge, there are no other prospective longitudinal studies conducted with ELLs that capture data from the very initial stages of reading acquisition through to a stage when text is central to the curriculum and when the demands of reading, and the language of text, are increasingly complex.

The study was guided by three overarching research questions:

1. Are there differences between ELLs and L1 speakers on component measures of reading at kindergarten and fourth grade?
2. Which kindergarten skills predict fourth-grade word reading and reading comprehension performance? Do these predictors differ for ELLs and their L1-speaking peers?
3. Do differential developmental trajectories of word reading between kindergarten and Grade 4 exist for ELLs and their L1-speaking peers? What are the kindergarten skills that predict these trajectories?

Method

Participants

The participants in this study included children from each of the 30 schools in a school district in a port city on the west coast of Canada; all children who entered kindergarten in the year the study began were invited to participate. The study began with 1,238 children who entered kindergarten and by Grade 4, due to attrition, 824 children remained in the sample. Of the remaining children, 689 (83.6%) were classified as L1 speakers, and 135 (16.4%) were classified as ELLs. Children were classified as ELLs in kindergarten if a language other than English was the primary language of the home; this information was obtained through school records. The ELL sample included children from 33 different language backgrounds; the predominant native languages in the group were Cantonese, Mandarin, Korean, Spanish, Polish, and Farsi.

Of the full sample, 406 (49.3%) were girls and 418 (50.7%) were boys. There were no significant gender differences between the two (ELL, L1) language groups, $\chi^2(1, N = 824) = .712$. The mean age of the sample in kindergarten was 64.33 months (i.e., about 5.3 years), with a standard deviation of 3.44 months; there

were no significant age differences between the two (ELL, L1) language groups, $t(809) = 1.96$.

The district's school-level data indicated that the study's ELLs were dispersed across all schools and communities and found not only in low- or high-income school communities. The distribution of ELLs across the district and therefore across neighborhoods of varying socioeconomic status levels differed from many of the ELL samples in U.S. research, in which the ELLs are often clustered in low-income neighborhoods and (often underperforming) schools in which the great majority of students are from low-income backgrounds and are eligible for free or reduced-fee lunch (e.g., Gandara, Rumberger, Maxwell-Jolly, & Callahan, 2003; Gerber et al., 2004; Gottardo, 2002; Swanson et al., 2006). In the present sample of 824 children studied, ELLs were enrolled in 26 of the 30 schools in the district, ranging from 4% to 53% of the student population in each of these schools. Analyses of school-level effects (presented in the Results section) revealed that this range in composition of the study population (ELLs, L1s) did not have significant effects on the research questions of interest and that the great majority of differences by language group were at the student level.

Attrition

The attrition rate of approximately 33% was not surprising given that this study was conducted over 5 years in a port city on the west coast of Canada, where there are high rates of immigration and subsequent mobility within the country. For example, in 2001, 18.8% of the Canadian population was foreign born (Migration Information, 2005), but this number does not even include the growing population of children of immigrants. Within the immigrant population, there is an increased propensity to relocate often due to employment opportunities as well as upward mobility following initial arrival in the country (Portes & Rumbaut, 1990; Suarez-Orozco, 2004). In order to examine whether those children who remained in the sample differed from those who were no longer in the study, we conducted a multivariate analysis of variance with all kindergarten measures. The results of the multivariate analysis of variance indicated that there were no significant differences on any of the variables, $F(11, 1226) = 0.735$, and thus no significant effect of attrition.

Risk Status and Reader Group Classification

In kindergarten, all children were classified as at risk or not at risk for reading difficulties based on their performance on the reading subtest of the Wide Range Achievement Test-3 (WRAT-3; Wilkinson, 1993). Children were classified as at risk for reading difficulties if their score on the WRAT-3: Reading subtest was at or below the 29th percentile and not at risk for reading difficulties if their performance was at or above the 30th percentile. In Grade 4, children were classified as poor readers or average readers using the same classification criteria and measure of word reading. In kindergarten, 28.2% of ELLs and 25.6% of L1 speakers were identified as at risk. However, by Grade 4, only 4.3% of ELLs and 3.8% of L1 speakers were classified as poor readers.

Instructional Context

The school district serving the children has made a commitment to a balanced reading acquisition program that includes phonological awareness instruction (Bennett & Ottley, 2000). The district has its own published reading curriculum for elementary grades (i.e., Reading 44, 1999) and has also published a kindergarten early literacy curriculum (i.e., Firm Foundations, 2001). In this district's elementary schools, ELLs receive the same early classroom instruction in English as their L1 peers. Thus, ELLs begin the same schooling in mainstream English classrooms at the same time as their L1 peers despite having generally more limited oral proficiency. AQ: 4

Following the kindergarten assessment, each school received feedback on the performance of the children who participated in the study. Children classified as at risk for reading difficulties then received specialized phonological awareness training—three to four times a week for 20 min—that continued until the end of kindergarten. The training was provided by the classroom and resource teachers and consisted of a mix of activities with an explicit emphasis on sound–symbol relationships and independent activities, such as cooperative story writing and journal writing, using invented spelling.

Measures

The children in the study were administered standardized and experimental measures of reading at kindergarten and fourth grade, which formed the basis of our analyses for addressing the first two research questions. Together, the measures assessed several component skills of reading, including working memory, phonological processing, and syntactic awareness, as well as other literacy skills including word and nonword reading, spelling, and reading comprehension. The measure of word reading was administered annually from kindergarten through fourth grade and formed the basis of our growth modeling, the focus of the third research question. Many of the measures selected have a long-standing history of use in reading research; in this study, some evidence of validity was provided for each measure, though the study was not necessarily designed to include multiple measures of every construct assessed. More extensive reporting on the technical properties of many of these measures appears in published volumes (e.g., Murphy & Impara, 1994; Spies & Plake, 2005).

Longitudinal Measure of Word Reading: WRAT-3: Reading subtest (blue form; Wilkinson, 1993) AQ: 5

Children were asked to read as many words as possible from a list of real words of increasing difficulty (e.g., "in," "cat," "stretch," "triumph"). The task administration was discontinued when 10 consecutive words were read incorrectly. The raw score was calculated as the number of words read correctly ($\alpha = .93$ for kindergarten; item-level data were not available for subsequent years). In kindergarten, this measure correlated highly ($r = .83$) with the letter identification measure. In fourth grade, this measure was highly correlated with other measures of word reading (word identification, $r = .76$; word attack, $r = .70$; speed of word reading, $r = .74$).

Kindergarten Measures

Letter identification (Gottardo, Chiappe, & Siegel, 2003). AQ: 6 Children were asked to name lowercase letters from a page of 26 letters presented in a random order. The raw score was calculated

as the number of letters read correctly out of 26 (α = .94). This measure correlated highly (r = .83) with the WRAT-3 measure of word reading.

Stanford-Binet Working Memory for Sentences (Thorndike, Hagen, & Sattler, 1986). Children were asked to repeat sentences ranging in complexity from simple two-word sentences (e.g., "Drink milk") to more complex sentences (e.g., "Ruth fell in a puddle and got her clothes all muddy"). The raw score was calculated as the number of sentences read correctly out of 42 (α = .86). This measure was moderately correlated (r = .41) with sound mimicry, which tapped children's ability to perceive and repeat sounds in pseudowords and was the kindergarten task most closely related to this one with respect to skills assessed.

Sound mimicry (Goldman, Fristoe, & Woodcock, 1974). Children were asked to repeat pseudowords of increasing difficulty that had been read to them by the examiner (e.g., "ab," "dod," "baf-motbem"). The raw score was calculated as the number of pseudowords read correctly out of 55 (α = .88). This measure was moderately correlated (r = .41) with working memory for sentences, the kindergarten task most closely related to this one with respect to skills assessed.

Rhyme detection (Muter, Hulme, & Snowling, 1997). For each task, children were shown four pictures in which a picture of a single target word appeared above three pictures. Children were asked which of the three words associated with the three pictures rhymed with the target word (e.g., "What rhymes with cat? Fish, sun or hat?"). The raw score was calculated as the number of words identified correctly out of eight (α = .87). This measure was moderately correlated (r = .49) with syllable identification.

Syllable identification (Muter et al., 1997). For each task, children were shown a picture (e.g., a rabbit). The examiner said the first part of the word (e.g., "ra"), and the children were asked to finish the word (e.g., "bit"). The raw score was calculated as the number of syllables pronounced correctly out of eight (α = .85). This measure was moderately correlated (r = .51) with phoneme identification.

Phoneme identification (Muter et al., 1997). For each task, children were shown a picture (e.g., a watch). The examiner said the first part of the word (e.g., "wa"), and the children were then asked to finish the word (e.g., "tch"). The raw score was calculated as the number of phonemes pronounced correctly out of eight (α = .91). This measure was moderately correlated (r = .51) with syllable identification.

Phoneme deletion (Muter et al., 1997). For each task, children were shown a picture of the target word (e.g., a bus) and were asked to delete either the initial or the final phoneme from the target word (e.g., "Bus without /b/ says ____"). The test consisted of 16 tasks, such that 8 tasks required initial phoneme deletion, and 8 tasks required final phoneme deletion. The raw score was calculated as the number of phonemes deleted correctly out of 16 (α = .94). This measure was moderately correlated with syllable identification (r = .50) and phoneme identification (r = .48).

Oral cloze (Siegel & Ryan, 1989; Willows & Ryan, 1986). For each task, children were read a sentence with a missing word (e.g., "The moon shines bright in the ____") and were asked to provide the missing word. The raw score was calculated as the number of words provided correctly out of 12 (α = .84). This measure was only somewhat correlated with other kindergarten measures in the study, with the strongest correlation being with kindergarten working memory (r = .379). Both of these tasks demanded children's verbal skills but were designed to tap different aspects of oral language. Although this task has been used in previous research, its construct validity has not been established via empirical study.

Simple spelling (Lesaux & Siegel, 2003). Children were asked to print their names along with five simple words (i.e., "mom," "no," "I," "cat," "dad"). The raw score was calculated as the number of words spelled correctly out of six (α = .78). This task correlated relatively strongly with reading (r = .63) and letter identification (r = .55).

Fourth-Grade Measures

AQ: 7

Stanford Diagnostic Reading Test (Karlsen & Gardner, 1994). Children were asked to read short passages and to answer multiple-choice questions related to each passage in a booklet within a fixed time frame of 40 min. The raw score was calculated as the number of questions answered correctly (α = .93). This measure was most strongly correlated with measures of word reading and word reading efficiency administered in fourth grade (WRAT-3, r = .48; word identification, r = 53; speed of word reading, r = .46).

Word identification (Woodcock Reading Mastery Tests–Revised; Woodcock, 1987). Children were asked to read as many words as possible from a list of real words of increasing difficulty (e.g., "is," "find," "mathematician"). The task administration was discontinued when six consecutive words were read incorrectly. The raw score was calculated as the number of words read correctly (α = .90). This measure was strongly correlated with other measures of word reading (WRAT-3, r = .76; word attack, r = .75; speed of word reading, r = .78).

Word attack (Woodcock Reading Mastery Tests–Revised; Woodcock, 1987). Children were asked to read as many words as possible from a list of pseudowords of increasing difficulty (e.g., "dee," "ap," "straced"). The task administration was discontinued when six consecutive words were read incorrectly. The raw score was calculated as the number of words read correctly (α = .91). This measure was strongly correlated with other measures of word reading (WRAT-3, r = .70; word identification, r = .75; speed of word reading, r = .72).

1-min word reading. Children were asked to read as many words as possible from a list of real words of increasing difficulty (e.g., "as," "because") within a 1-min time period; the tan form of the WRAT-3: Reading subtest was used to develop the word list. The raw score was computed as the number of words read correctly (α = .89). This measure was strongly correlated with other measures of word reading (WRAT-3, r = .75; word identification, r = .79; speed of nonword reading, r = .78).

1-min pseudoword reading. Children were asked to read as many pseudowords as possible from a list of pseudowords of increasing difficulty (e.g., "yee," "dreek") within a 1-min time period; Form H of the Woodcock Reading Mastery Tests–Revised Word Attack subtest (Woodcock, 1987) was used to develop the word list. The raw score was computed as the number of words read correctly (α = .90). This measure was strongly correlated with other measures of word reading (WRAT-3, r = .74; word identification, r = .76; speed of word reading, r = .76).

WRAT-3: Spelling (tan form; Wilkinson, 1993). Children were read real words of increasing difficulty (e.g., "must," "enter") and

were asked to spell them. The raw score was calculated as the number of words spelled correctly (α = .89). This measure was strongly correlated with measures of word reading (WRAT-3, r = .68; word identification, r = .71; speed of word reading, r = .74).

Pseudoword spelling. Children were read different pseudowords (e.g., "tave") and were asked to generate a plausible spelling of each pseudoword. The raw score was calculated as the number of pseudowords spelled correctly out of 15 (α = .85). This measure was moderately correlated with WRAT-3: Spelling (r = .47) and measures of word-level reading including word attack (r = .41), and word identification (r = .43).

Working memory for words (Siegel & Ryan, 1989). For each task, children were read sets of sentences for which the final word was missing (e.g., "Snow is white, grass is _____") and were asked to provide the missing word of each sentence. They were then asked to repeat all missing words from each set of sentences for a total of three trials within each set of sentences. The task administration was discontinued when all words were incorrect for one set of sentences. The raw score was computed as the number of correct responses (α = .84). This measure was moderately correlated with working memory for numbers (r = .43). Although this task has been used in previous research, its construct validity has not been empirically studied.

Working memory for numbers (Siegel & Ryan, 1989). For each task, children were asked to count yellow dots within a field of blue and yellow dots arranged in a randomly determined irregular pattern on a 5- × 8-in index card with an increasing number of cards in each set. For each set, children were then asked to recall the number of yellow dots on each card and the order in which they were presented with three trials within each set of cards. The task administration was discontinued when all answers at one level were incorrect. The raw score was computed as the number of correct responses (α = .73). This measure was moderately correlated with working memory for words (r = .43). Although this task has been used in previous research, its construct validity has not been empirically studied.

Rosner Auditory Analysis Test (Rosner & Simon, 1971). Children were asked to say a real word and then to say the word again without one of its sounds (e.g., "Say *smell*. Now say *smell* without the /m/ sound"). Children were asked to delete syllables, single phonemes from both the initial and final positions in each word, and single phonemes from blends. The raw score was calculated as the number of sounds deleted correctly out of 40 (α = .88). This measure was moderately correlated with word attack (r = .54), WRAT-3: Spelling (r = .52), and speed of nonword reading (r = .51).

Oral cloze (Siegel & Ryan, 1989; Willows & Ryan, 1986). For each task, children were read a sentence with a missing word (e.g., "The moon shines bright in the _____") and were asked to provide the missing word. The raw score was calculated as the number of words provided correctly out of 12 (α = .84). Although this task has been used in previous research, its construct validity has not been established via empirical study. Given the relationship of syntactic awareness and vocabulary to reading comprehension, this measure was most strongly related to the Stanford Diagnostic Reading Test comprehension measure (r = .44).

Procedure

Children were assessed late in the fall semester of kindergarten and early in the spring of all subsequent grades. In kindergarten and first grade, children were assessed individually in a quiet room for all tasks. In Grades 2 through 4, children were individually assessed for all tasks except for reading comprehension and spelling, which were administered in a group setting in the children's classrooms. All individual tests were administered in one session that lasted between 25 and 45 min, depending on the year of administration and the abilities of the individual child. Generally, testing time increased as the children became older, given that they were able to respond to more items on each of the tests administered. Some children were not administered every task due to absence from the classroom on the day of testing, though we made every effort to ensure that complete data were obtained for all participants.

Informed consent was obtained on an annual basis from parents or guardians of the participating children. In some cases, the consent forms were translated into a language other than English in order to ensure that all parents had an understanding of the study and were able to make an informed decision about their child's participation.

Results

The purpose of this article was to investigate differences in the reading achievement of ELLs and their L1-speaking peers at kindergarten and Grade 4 (Research Question 1), to investigate kindergarten predictors of ELLs' and L1 speakers' fourth-grade word reading and reading comprehension (Research Question 2), and to model the developmental trajectories of word reading for these two groups of learners (Research Question 3). Three statistical techniques—multiple analysis of variance, hierarchical linear regression, and growth modeling—were used to address the three research questions, in turn. For each analysis, a level of statistical significance of α = .05 was used.

Differences Between ELLs and L1 Speakers on Kindergarten and Fourth-Grade Skills Assessed

The ELLs and L1 speakers were compared on component skills and reading outcomes in kindergarten and Grade 4 using analysis of variance models. Due to the nested structure of the data with children sampled within schools, hierarchical linear models were constructed using the HLM 6.0 software package (Raudenbush & Bryk, 2002, 2005). Missing data was imputed five times under a multivariate normal assumption, and results from these analyses were combined to provide the correct model estimates. Raw scores were used for missing data analyses, following which all variables were standardized using a z transformation to eliminate scale differences. AQ: 8

Three models were fitted for each component measure of reading assessed. In the first model, the mean proficiency level was modeled without predictors and was allowed to vary across schools. In the second model, language status (i.e., ELL, L1) was added as a predictor at Level 1 to detect differences between language groups; this effect was held constant across schools. In the third model, the percentage of ELLs for each school was added

tapraid5/zcz-edu/zcz-edu/zcz00307/zcz2073d07g | sangreyj | S=4 | 6/28/07 | 21:40 | Art: 2004-0215-RRRRR

as a predictor at the school level to detect whether performance differences on the component measures existed between the language groups.

The descriptive statistics for ELLs' and L1 speakers' performance on the kindergarten component measures are shown in Table 1, whereas those for the fourth-grade component measures are shown in Table 2. In each table, the skills that were significant at Level 1 or Level 2 of the model are marked with an asterisk. As shown in Table 1, in kindergarten ELLs performed more poorly than their L1-speaking peers on the following measures: working memory, $t(822) = -5.54$, $p < .05$; sound mimicry, $t(822) = -2.03$, $p < .05$; rhyme detection, $t(822) = -5.52$, $p < .05$; and oral cloze, $t(822) = -3.36$, $p < .05$. Specifically, the ELLs performed approximately two thirds of a standard deviation lower than L1 speakers on working memory ($d = -.62$), approximately one fifth of a standard deviation lower on sound mimicry ($d = -.22$), approximately half a standard deviation lower on rhyme detection ($d = -.54$), and approximately one third of a standard deviation lower on oral cloze ($d = -.34$). Although there were no differences between the two groups on the phoneme deletion measure, there was a significant school-level effect on this measure, $t(29) = -2.88$, $p < .05$, which indicates that ELL and L1 children in schools with a high proportion of ELLs performed more poorly on this task.

As shown in Table 2, by fourth grade, the groups differed in performance on only two of the tasks—both spelling tasks. These differences were detected at different levels in the model. At the individual level (Level 1), the ELLs had significantly higher scores than the L1 speakers on WRAT-3: Spelling measure, $t(822) = 3.05$, $p < .05$; the ELLs' performance was about a quarter of a standard deviation higher than that of the L1 speakers ($d = .27$). In contrast, a school-level effect in performance was detected for pseudoword spelling, which indicates that, if one compares schools with the same proportion of ELLs, the L1 speakers performed about a quarter of a standard deviation ($d = -.22$) higher in those schools than did ELLs, $t(821) = -2.01$, $p < .05$. These analyses demonstrate that, despite higher performance by the L1 speakers on several of the kindergarten tasks, the two groups performed very similarly on the battery of fourth-grade measures. The analyses further demonstrate that the language differences at the student level were essentially independent of the language composition of the school except in two instances (kindergarten phoneme deletion and fourth-grade pseudoword spelling). Similarly, the intraclass correlation coefficients demonstrate that between 90% and 99% of the variation in performance was due to individual differences between children, not to differences between schools.

Kindergarten Predictors of Fourth-Grade Word Reading and Reading Comprehension

The second research question focused on kindergarten predictors of fourth-grade performance on the WRAT-3: Reading measure of word reading and the Stanford Diagnostic Reading Test measure of reading comprehension. A hierarchical regression model was fitted to the *z*-standardized measures for each of the outcome variables. In these models, the measures of kindergarten component skills were entered as Level 1 predictors along with a main effect for language (ELL, L1) and interaction variables for language status and the kindergarten component measures. In both models, the intercept was allowed to vary randomly across schools to allow for mean differences across schools, but the slope parameters were fixed across schools to model the same relationship across all schools. The results of these analyses are presented in Table 3.

Letter identification, working memory, rhyme detection, and oral cloze, administered in kindergarten, were identified as significant predictors of fourth-grade reading comprehension. Very similarly, letter identification, working memory, rhyme detection, and phoneme deletion were the kindergarten predictors of fourth-grade word reading. With respect to the effect of language status on predicting fourth-grade word reading and reading comprehension, simple spelling was the only variable that had a significant interaction with language status (ELL, L1) to influence reading com-

Table 1
Kindergarten Measures: Descriptive Statistics and Hierarchical Analysis of Variance Results

	L1s (n = 689)		ELLs (n = 135)		Level 1	Level 2
Measure (raw score)	*M*	*SD*	*M*	*SD*	Sig	Sig
WRAT-3: Reading	11.57	5.17	11.58	6.14		
Letter identification	15.25	7.34	16.22	8.81		
Working memory	16.76	3.62	14.43	3.97	*	
Sound mimicry	48.28	6.19	46.62	7.98	*	
Rhyme detection	7.00	3.02	5.34	3.22	*	
Syllable identification	4.58	2.63	4.23	2.38		
Phoneme identification	2.70	2.92	2.52	2.79		
Phoneme deletion	3.40	4.50	3.23	4.53		*
Oral cloze	2.32	2.63	1.45	2.13	*	
Simple spelling	2.59	1.80	2.24	1.86		

Note. Variables for which coefficients in the hierarchical analyses of variance were significant at $\alpha = .05$ are marked with an asterisk in the respective column. L1 = native English speaker; ELL = English-language learner; Sig = significance; WRAT-3 = Wide Range Achievement Test-3.
* $p < .05$.

Table 2
Fourth-Grade Measures: Descriptive Statistics and Hierarchical Analysis of Variance Results

	L1s (n = 689)		ELLs (n = 135)		Level 1	Level 2
Measure (raw score)	*M*	*SD*	*M*	*SD*	Sig.	Sig.
Stanford Diagnostic Reading Test	41.26	7.51	40.36	7.78		
WRAT-3: Reading	37.58	4.47	38.20	4.63		
Word identification	6.91	2.61	6.89	2.18		
Word attack	21.85	5.08	22.36	4.65		
1-min word reading	21.48	5.06	21.92	4.75		
1-min pseudoword reading	31.33	7.07	31.56	7.07		
Working memory (words correct)	4.62	1.97	4.18	1.95		
Working memory (numbers correct)	7.70	2.35	7.39	2.49		
Rosner Auditory Analysis Test	31.50	6.57	31.42	6.22		
Oral cloze	8.61	1.38	8.35	1.65	*	
WRAT-3: Spelling	66.72	24.79	74.65	23.03		*
Pseudoword spelling	9.92	2.75	9.64	2.62		

Note. Variables for which coefficients in the hierarchical analyses of variance were significant at $\alpha = .05$ are marked with an asterisk in the respective column. L1 = native English speaker; ELL = English-language learner; Sig = significance; WRAT-3 = Wide Range Achievement Test-3.
* $p < .05$.

prehension scores. In contrast, in the model with word reading as the outcome there were no interaction effects but there was a main effect of language. The main effect indicates that among children with identical profiles on the kindergarten component skills, ELLs performed, on average, about a quarter of a standard deviation (d = .26) higher on word reading than did L1 speakers. Similar to the hierarchical analysis of variance presented above, the results of the regression analyses show that between 94% and 99% of the variation in performance was due to individual differences between children, not to differences between schools.

Table 3
Hierarchical Regression Models for Predicting Fourth-Grade Outcome Measures From Kindergarten Skills

Variable	Stanford Diagnostic Reading Test	Sig	WRAT-3: Reading	Sig.
Intercept	−.04 (.05)		−.04 (.05)	
WRAT-3: Reading	.00 (.06)		.05 (.06)	
Letter identification	.15 (.06)	*	.13 (.06)	*
Working memory	.18 (.03)	*	.14 (.03)	*
Sound mimicry	.05 (.04)		.05 (.05)	
Rhyme detection	.13 (.03)	*	.09 (.04)	*
Syllable identification	.03 (.05)		.04 (.04)	
Phoneme identification	−.01 (.04)		.03 (.04)	
Phoneme deletion	.05 (.03)		.12 (.03)	*
Oral cloze	.12 (.03)	*	.07 (.04)	
Simple spelling	.08 (.05)		−.01 (.04)	
Language	.06 (.10)		.26 (.11)	*
Language × WRAT-3	−.04 (.11)		.22 (.18)	
Language × Letter Identification	.15 (.14)		−.19 (.18)	
Language × Working Memory	.09 (.11)		.01 (.08)	
Language × Sound Mimicry	−.10 (.08)		.04 (.07)	
Language × Rhyme Detection	−.10 (.07)		−.06 (.12)	
Language × Syllable Identification	.06 (.12)		−.01 (.14)	
Language × Phoneme Identification	.04 (.09)		−.11 (.12)	
Language × Phoneme Deletion	.03 (.10)		−.12 (.10)	
Language × Oral Cloze	−.05 (.09)		−.05 (.11)	
Language × Simple Spelling	−.28 (.09)	*	.04 (.11)	

Note. All coefficients that are statistically significant at $\alpha = .05$ are marked with an asterisk in the Sig column. Standard errors are in parentheses. Sig. = significance. WRAT-3 = Wide Range Achievement Test-3.
* $p < .05$.

Table 4
Model Fit Statistics for Sequential Growth Models

Growth model	−2LL	χ^2	*df*	CFI	TLI	RMSEA	$SRMR_W$	$SRMR_B$	BIC
Trend models (2 level)									
Linear growth					Noconvergence				
Grade 4 free	−11173.41	1085.49	18	.41	.34	.38	.27		22424.68
Grades 3 and 4 free	−10776.47	291.61	16	.85	.81	.20	.17		21637.87
Grades 2, 3, and 4 free	−10659.13	56.91	14	.98	.97	.09	.15		21410.25
Invariance models (3 level)									
Equal slope coefficients	−10649.80	216.29	41	.92	.92	.10	.31	.04	21402.22
Equal residual variances	−10650.73	197.27	46	.93	.94	.09	.31	.04	21386.39
Equal latent variances	−10653.74	202.93	50	.93	.94	.09	.33	.04	21378.25
Equal latent correlations	−10654.92	204.97	52	.93	.94	.08	.33	.05	21373.53
Predictor models (3 level)									
With kindergarten predictors	−9430.89	300.93	106	.93	.91	.07	.21	.03	19049.14
Equal regression coefficients	−9440.04	321.70	124	.93	.93	.06	.21	.04	19004.98

Note. All chi-square difference tests in the Trend models block are statistically significant at $\alpha = .05$. However, none of the chi-square difference tests in the Invariance models or Predictor models blocks are statistically significant at $\alpha = .05$. All models estimated with a maximum likelihood estimator with robust standard errors. CFI = confirmatory fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; $SRMR_W$ = standardized root-mean-square residual at the within level (Level 2, individuals); $SRMR_B$ = standardized root-mean-square residual at the between level (Level 3, schools); BIC = sample-size-adjusted Bayesian information criterion.

Modeling and Predicting Growth in Word Reading

The third research question focused on the developmental trajectories of word reading for ELLs and their L1-speaking peers from kindergarten through fourth grade, and the kindergarten predictors of these trajectories. Using Mplus 3.11 (Muthén & Muthén, 2005), we conducted a hierarchical latent growth curve analysis. WRAT-3: Reading raw scores[1] were used in this analysis, and missing data were imputed using the full-information maximum likelihood routine. A sequence of 10 models was fitted to the data, and each model was concurrently estimated across the two language groups. The 10 models were arranged by their level of complexity and their degree of restriction in three blocks; the fit statistics for all models are presented in Table 4.

In the first step, the nature of the growth trend was estimated using two-level models that allowed for a free estimation of the slope parameters[2] to be used as input for subsequent models; in Table 4 these models are referred to as *trend models*. The models ranged from a model for linear growth to a model in which the slope coefficients for Grades 2, 3, and 4 were estimated freely. As the fit statistics demonstrate, the model postulating a linear trend did not converge, whereas the model in which all slope coefficients were estimated freely best fit the data. The pattern of the estimated coefficients indicated a nonlinear growth trajectory for both groups with about half-step increments after a growth of about two thirds from Grade 1 to Grade 2. This result is supported descriptively in Figure 1, which shows a clear nonlinear trend in the mean WRAT-3: Reading raw scores and percentiles across the 5 years with only a small, and practically negligible, decrease in mean percentile scores in fourth grade. For all subsequent models, the slope coefficients were, thus, fixed to 0.00, 1.00, 1.67, 2.00, and 2.16 for all models to capture this nonlinear trajectory.

In a second step, in order to test for increasing degrees of measurement and structural invariance of the parameters across the two language groups (e.g., Meredith, 1993), we fitted a sequence of another four models, referred to as *invariance models* in Table 4. These were models that imposed an increasing number of restrictions on selected parameters. The chi-square difference tests[3] between the first and second models, $\chi^2(5) = 6.13$; the second and third models, $\chi^2(4) = 8.82$; as well as the third and fourth models, $\chi^2(2) = 3.37$, were nonsignificant, indicating that the most restrictive model was supported by these data.

In a third step, two models were fitted that included all component measures from kindergarten as predictors for the latent intercept and slope at the level of the individual student. These models are referred to as *predictor models* in Table 4. In the first model, the regression coefficients between the kindergarten predictors and the latent variables were estimated freely across both language groups, whereas in the second model they were constrained to be equal across the ELL and L1 groups. The chi-square difference test for the two models, $\chi^2(18) = 21.75$, *ns*, showed that they could be considered as statistically equivalent models, which is why the most complex model was retained.

The parameter estimates of this final model, presented in Table 5, showed that for ELLs and their L1-speaking peers, letter identification was most predictive of the initial status and growth in reading. As expected, the model indicated that children with higher levels on the statistically significant component skills in kindergarten showed higher mean scores on the WRAT-3: Reading measure in kindergarten as well as higher amounts of growth.

In sum, the kindergarten measures accounted for a large proportion of the variation in the intercept parameter ($R^2 = .87$)—mean ability in kindergarten—and a moderate amount of variation

[1] The same sequence of models as described in this section was also fitted for the WRAT-3: Reading percentiles and produced almost identical fit statistics. To economize on space, we present here only the results for the raw scores.

[2] The coefficients at kindergarten and Grade 1 were fixed to 0 and 1, respectively, for model identification.

[3] Because a maximum likelihood estimator with robust standard errors was used, a corrected chi-square test statistic was used (see Technical Appendices at www.statmodel.com).

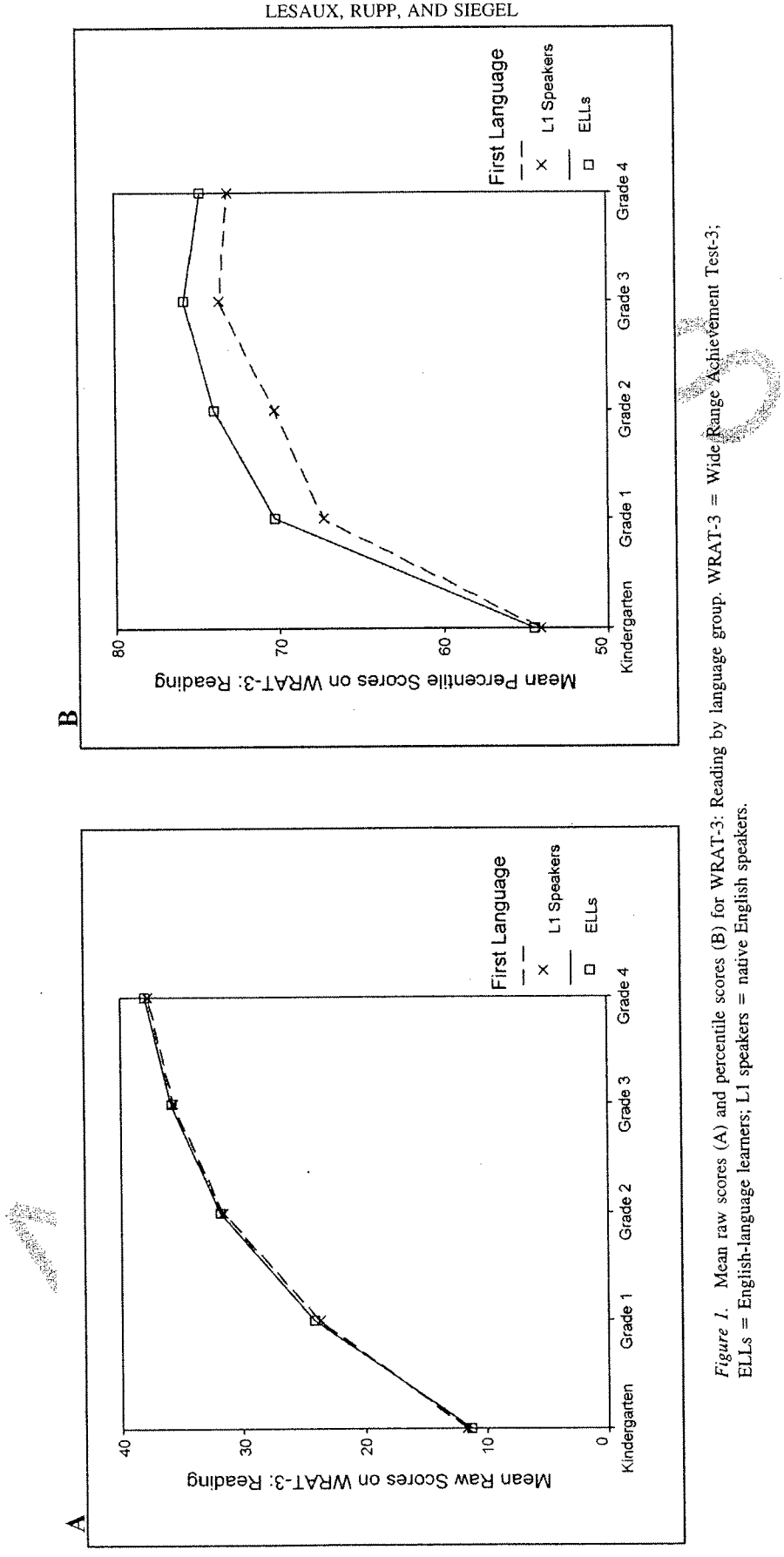

Figure 1. Mean raw scores (A) and percentile scores (B) for WRAT-3: Reading by language group. WRAT-3 = Wide Range Achievement Test-3; ELLs = English-language learners; L1 speakers = native English speakers.

Table 5
Estimated Model Parameters for Nonlinear Growth Model With Kindergarten Component Variables

Variable	Intercept	Slope
Outcome variables		
WRAT-3: Reading (kindergarten)	1	0
WRAT-3: Reading (Grade 1)	1	1
WRAT-3: Reading (Grade 2)	1	1.67
WRAT-3: Reading (Grade 3)	1	2.00
WRAT-3: Reading (Grade 4)	1	2.16
Kindergarten predictor variables		
Letter identification	**.77**	**−0.64**
Working memory for sentences	−.01	**0.12**
Sound mimicry	**.06**	0.03
Rhyme detection	**.05**	0.03
Syllable identification	−.05	0.08
Phoneme identification	**.08**	−0.03
Phoneme deletion	.04	**0.09**
Oral cloze	.10	0.00
Simple spelling	**.11**	**−0.13**
Latent variable information		
R^2 for latent variables (ELL)	.91	.53
R^2 for latent variables (L1)	.88	.45
Latent variable correlation (within, L1)	$r = -.06$	
Latent variable correlation (between, L1)	$r = -.30$	
Latent variable correlation (within, ELL)	$r = -.05$	
Latent variable correlation (between, ELL)	$r = -.30$	

Note. Statistically significant standardized regression coefficients are shown in boldface. WRAT-3 = Wide Range Achievement Test-3; ELL = English-language learner; L1 = native English speaker.

in the slope parameter ($R^2 = .42$)—the amount of the individual nonlinear growth in word reading over time. Moreover, as demonstrated by the nonsignificant latent correlations between the intercept and slope for the L1 speakers ($r = -.05$) and the ELLs ($r = -.06$), children's initial status in reading was not predictive of their growth in reading.

Discussion

Whereas prior research conducted with ELLs has typically been cross-sectional in nature and has included relatively small samples, this study investigated the reading development for a large sample of ELLs and L1 speakers from kindergarten through fourth grade. Overall, the findings suggest that differences in the reading development for ELLs and L1 speakers are negligible. Although in kindergarten the ELLs performed more poorly than the L1 speakers on several tasks of early literacy, by fourth grade these differences had generally disappeared. On all but one fourth-grade task, the ELLs performed similarly to—or better than—the L1 speakers. Similarly, the kindergarten predictors of fourth-grade word reading and reading comprehension performance were almost identical for each of the two groups; the only exception was that the kindergarten measure of spelling showed a differential effect across the two language groups for predicting reading comprehension. Finally, the nonlinear developmental trend in word reading from kindergarten through fourth grade was strikingly similar for ELLs and L1 speakers. For both language groups, letter identification skills in kindergarten were predictive of initial mean differences and growth over time in word reading. Also for both groups, initial word reading ability in kindergarten was not predictive of subsequent growth in word reading.

These findings converge and diverge in several ways with respect to the design and findings from the extant research focused on the development of reading for ELLs. Previous research has shown that, compared to their L1-speaking peers, ELLs enter schools at greater risk for oral language difficulties and phonological and print-related difficulties (e.g., Foorman & Torgesen, 2001; Gerber et al., 2004; Gottardo, 2002; Manis et al., 2004). The kindergarten performance of the ELL sample in this study is consistent with these findings in that they had significantly lower scores than their L1-speaking peers on 4 of the 10 measures administered.

It is important to note that each of these four tasks (i.e., working memory, sound mimicry, rhyme detection, and oral cloze) are mediated very strongly by language proficiency. For three of these tasks (i.e., working memory, sound mimicry, and oral cloze), remembering long sentences and subsequently producing language were central to a child's performance on the task. Although producing language was not central to performance on the rhyme detection measure, for this task each child had to retain the name for each of the four pictures of common objects and subsequently decide on the one that rhymed with the target picture. Degree of familiarity with these words and names for the objects is likely to have affected the ELLs' performance. For each of the other phonological awareness tasks, only one word, and in turn picture, was used.

An additional point of convergence with previous research is that, at fourth grade, ELLs' performance on word-level tasks, including word reading, nonword reading, and spelling, was commensurate with that of their L1-speaking peers. This is consistent with previous cross-sectional research conducted with ELLs mainly in the primary grades (e.g., Chiappe & Siegel, 1999; D'Angiulli, Siegel, & Serra, 2001; da Fontoura & Siegel, 1995; Geva et al., 2000; Limbos & Geva, 2001; for a review, see Lesaux et al., 2006).

Where the design of this study and, in turn, the findings diverge markedly from previous research with ELLs is that the cohort of ELLs and their L1-speaking peers was followed from kindergarten through fourth grade. To our knowledge, there are no other prospective longitudinal studies conducted with ELLs that capture data from the very initial stages of reading acquisition through to a stage when text is central to the curriculum and when the demands of reading, and the language of text, are increasingly complex. Thus, in addition to assessing cognitive and linguistic skills related to reading and word-level skills, such as word reading and spelling, we administered a measure of reading comprehension to each of the participants. The study design allowed for an examination of the kindergarten predictors of fourth-grade word reading and reading comprehension skills.

On the fourth-grade measure of reading comprehension—the goal of reading instruction and a central predictor of school success—the ELLs performed at comparable levels to the L1 speakers, and there were no differences in kindergarten predictors by language group. The similarity in reading comprehension performance between the groups is in contrast to previous research, primarily conducted in the Netherlands, that has shown that reading comprehension is a significant weakness for language-minority learners (e.g., Aarts & Verhoeven, 1999; Droop & Verhoeven,

1998; Hacquebord, 1994; Verhoeven, 1990, 2000). Although it is not possible to draw conclusions about the reason for this divergence in findings, it raises very important questions for researchers.

Some of these questions relate to the contextual differences between this study and other studies with language-minority learners that have included a comparison group of L1 speakers. In the present study, the ELLs were distributed across all schools and neighborhoods in the district, whereas in most other studies with a comparison group of native speakers, the language-minority learners—as a group—had lower levels of socioeconomic status and were concentrated in a small number of (often underperforming) schools. Given the established relationship among socioeconomic status, vocabulary, and reading comprehension based on research conducted with L1 speakers (e.g., RAND Reading Study Group, 2002; Serpell, Baker, & Sonnenschein, 2005; Snow et al., 1998), this hypothesis warrants investigation in research with ELLs.

Equally plausible is that there were differences in the instructional context of this study and previous studies; most developmental research in reading has not systematically accounted or controlled for instructional practices. For the present study, due to a lack of available resources, it was not feasible to collect and analyze information on instruction. However, this information is likely to have shed further light on the findings. Developmental research that includes a focus on context—instructional and/or demographic—is needed in order to investigate these differences in findings as they relate to the reading comprehension of language-minority learners.

In addition to studying ELLs' development of reading comprehension skills, this longitudinal design afforded the opportunity to move beyond traditional methods of analysis using group means. Thus, growth modeling was used to examine the trajectories of word reading from kindergarten through fourth grade, and the kindergarten predictors of these trajectories. Although several longitudinal studies conducted with monolingual English speakers have modeled the trajectories of word reading development for these learners (e.g., Compton, 2000; Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996), to our knowledge this is the first study that conducted similar analyses with a sample of ELLs and simultaneously allowed for a comparison to a group of L1 speakers. The nonlinear developmental trajectory of word reading was strikingly similar for the two groups. Furthermore, whereas letter identification skills in kindergarten were predictive of mean differences in initial word reading ability and growth in word reading ability, initial word reading ability in kindergarten was not predictive of subsequent growth in word reading ability; this finding raises important questions.

Most obviously, this finding is further cause to again consider the effects of the instructional context in which the present study was conducted. As noted, the study was conducted in a school district with a very strong emphasis on literacy achievement and a balanced approach to early literacy instruction that included systematic instruction in phonological awareness in kindergarten and phonics in the primary grades. Findings from intervention studies suggest that ELLs' early reading skills can be significantly improved with targeted phonics instruction in the primary grades (e.g., Gerber et al., 2004; Linan-Thompson, Vaughn, Hickman-Davis, & Kouzekanani, 2003; Stuart, 1999, 2004). In the absence of other longitudinal studies with similar design, these findings suggest that the reading development of ELLs and L1 speakers is very similar and that if ELLs experience early difficulties similar to those of native speakers, this is not necessarily indicative of later difficulties. However, these findings need to be replicated with different samples and in different contexts in order to arrive at consensus on patterns of ELLs' reading development; thus, the developmental trajectory of ELLs and their L1 peers must be further examined in other longitudinal studies.

In conclusion, in addition to undertaking studies designed to replicate these findings, an obvious next step to continue promoting an understanding of ELLs' reading development is to focus on their achievement through middle and high school, when the demands on reading comprehension abilities continue to increase and when reading comprehension is central to academic success. Many ELLs—even those who enroll in U.S. schools in kindergarten—perform on par with peers in the primary grades but struggle with the demands of the middle and high school curricula (e.g., De George, 1988; de Jong, 2004; Gandara & Rumberger, 2003; Lesaux, Kieffer, & Crosson, 2006). Beyond fourth grade, heavier demands are placed on the reader, not just for obtaining a deeper understanding of text but also for dealing with more challenging aspects of foundational processes, given that words to be read and spelled become more phonologically and morphologically complex. Even for L1 speakers, progress in reading can falter when they are faced with these new challenges (e.g., Juel, 1991; Leach, Scarborough, & Rescorla, 2003). Longitudinal research conducted with ELLs through the middle and high school years would shed light on whether their gains made in previous grades are retained and progress is maintained in order to promote and support later academic achievement. AQ: 12

References

Aarts, R., & Verhoeven, L. (1999). Literacy attainment in a second language submersion context. *Applied Psycholinguistics, 20,* 377–393.

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

August, D., & Hakuta, K. (Eds.). (1997). *Improving schooling for language-minority children: A research agenda.* Washington, DC: National Academy Press.

August, D. L., & Shanahan, T. (2006). Introduction and methodology. In D. L. August & T. Shanahan (Eds.), *Developing literacy in a second language: Report of the National Literacy Panel.* Mahwah, NJ: Erlbaum. AQ: 13

Bennett, L., & Ottley, P. (1996). *Launch into reading success.* Austin, TX: PRO-ED. AQ: 14

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

Chiappe, P., Hasher, L., & Siegel, L. S. (2000). Working memory, inhibitory control, and reading disability. *Memory & Cognition, 28,* 8–17.

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

Chiappe, P., Siegel, L. S., & Gottardo, A. (2002). Reading-related skills of kindergartners from diverse linguistic backgrounds. *Applied Psycholinguistics, 23,* 95–116. AQ: 15

Compton, D. L. (2000). Modeling the growth of decoding skills in first-grade children. *Scientific Studies of Reading, 4,* 219–285.

Crawford, J. (2004). *Educational English learners: Language diversity in the classroom.* Los Angeles: Bilingual Education Services.

Cummins, J. (2000). *Language, power, and pedagogy: Bilingual children in the crossfire.* Clevedon, England: Multilingual Matters.

da Fontoura, H. A., & Siegel, L. S. (1995). Reading, syntactic, and working memory skills of bilingual Portuguese-English Canadian children. *Reading and Writing: An Interdisciplinary Journal, 7,* 139–153.

D'Angiulli, A., Siegel, L. S., & Serra, E. (2001). The development of reading in English and Italian in bilingual children. *Applied Psycholinguistics, 22,* 479–507.

De George, G. P. (1988). *Assessment and placement of language minority students: Procedures for mainstreaming. Focus, 3.* Retrieved October 16, 2004, from http://www.ncela.gwu.edu/pubs/ classics/focus/03mainstream.htm

AQ: 16

de Jong, E. J. (2004). After exit: Academic achievement patterns of former English language learners. *Educational Policy Analysis Archives, 12.* Retrieved September 22, 2005, from http://epaa.asu.edu/epaa/v12n50/v12n50.pdf

Development Associates. (2003). *Descriptive study of services to LEP students and LEP students with disabilities* (Vol. 1A). Retrieved October 15, 2005, from http://www.devassoc.com/pdfs/vol_1_text.pdf

Droop, M. V., & Verhoeven, L. (1998). Background knowledge, linguistic complexity, and second-language reading comprehension. *Journal of Literacy Research, 30,* 253–271.

Durgunoglu, A., Nagy, W., & Hancin-Bhatt, B. (1993). Cross-language transfer of phonological awareness. *Journal of Educational Psychology, 85,* 453–465.

Everatt, J., Smythe, I., Adams, E., & Ocampo, D. (2000). Dyslexia screening measures and bilingualism. *Dyslexia, 6,* 42–56.

Firm foundations: Early literacy teaching and training. (2001). North Vancouver, British Columbia, Canada: School District No. 44.

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

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

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curve analysis. *Journal of Educational Psychology, 88,* 3–17.

Gandara, P., & Rumberger, R. (2002). *The inequitable treatment of English learners in California's public schools* (Paper wws-rr005–1002). Retrieved from http://repositories.cdlib.org/idea/wws/wws-rr005–1002

AQ: 17

Gandara, P., Rumberger, R. W., Maxwell-Jolly, J., & Callahan, R. (2003). English learners in California schools: Unequal resources, unequal outcomes. *Educational Policy Analysis Archives, 11.* Retrieved from http://epaa.asu.edu/epaa/v11n36/

AQ: 18

García, G. E. (1991). Factors influencing the English reading test performance of Spanish-speaking Hispanic children. *Reading Research Quarterly, 26,* 371–392.

García, G. E. (2000). Bilingual children's reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 813–834). Mahwah, NJ: Erlbaum.

Geva, E., & Siegel, L. S. (2000). Orthographic and cognitive factors in the concurrent development of basic reading skills in two languages. *Reading and Writing: An Interdisciplinary Journal, 12,* 1–30.

Geva, E., Yaghoub-Zadeh, Z., & Schuster, B. (2000). Understanding differences in word recognition skills of ESL children. *Annals of Dyslexia, 50,* 123–154.

Goldman, R., Fristoe, M., & Woodcock, R. (1974). *Goldman-Fristoe-Woodcock Auditory Skills Test Battery.* Circle Pines, MN: American Guidance Service.

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

Hacquebord, H. (1994). L2 reading in the content areas: Text comprehension in secondary education in the Netherlands. *Journal of Research in Reading, 17,* 83–98.

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. (1991). Beginning reading. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2. pp. 749–788). New York: Longman.

Karlsen, B., & Gardner, E. (1994). *Stanford Diagnostic Reading Test.* San Francisco: Harcourt, Brace.

Kindler, A. L. (2002). *Survey of the states' limited English proficient students and available educational programs and services, 2000–2001, summary report.* Washington, DC: National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs.

Leach, J., Scarborough, H., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology, 95,* 211–224.

Leafstedt, J. M., Richards, C. R., & Gerber, M. M. (2004). Effectiveness of explicit phonological-awareness instruction for at-risk English learners. *Learning Disabilities Research & Practice, 19,* 252–261.

AQ: 19

Lesaux, N. K., Kieffer, M. J., & Crosson, A. C. (2006, July). *Spanish-speakers' reading comprehension in English.* Paper presented at the annual meeting of the Society for the Scientific Study of Reading, Vancouver, British Columbia, Canada.

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

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

Limbos, L. M., & Geva, E. (2001). Accuracy of teacher assessments of second-language students at risk for reading disability. *Journal of Learning Disabilities, 34,* 136–151.

Linan-Thompson, S., Vaughn, S., Hickman-Davis, P., & Kouzekanani, K. (2003). Effectiveness of supplemental reading instruction for second-grade English language learners with reading difficulties. *Elementary School Journal, 103,* 221–238.

McDougall, S., Hulme, C., Ellis, A., & Monk, A. (1994). Learning to read: The role of short-term memory and phonological processing skills. *Journal of Experimental Child Psychology, 58,* 112–133.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika, 58,* 525–543.

Migration Information. (2005). *Global data.* Retrieved March 19, 2005, from http://www.migrationinformation.org/Global Data/countrydata/data.cfm

Moll, L. C., & Diaz, S. (1985). Ethnographic pedagogy: Promoting effective bilingual instruction. In E. E. Garcia & R. V. Padilla (Eds.), *Advances in bilingual education research* (pp. 127–149). Tucson: University of Arizona Press.

Murphy, L. L., & Impara, J. C. (Eds.). (1994). *Buros desk reference: Psychological assessment in the schools.* Lincoln: University of Nebraska Press.

Muter, V., & Diethelm, K. (2001). The contribution of phonological skills and letter knowledge to early reading development in a multilingual population. *Language Learning, 51,* 187–219.

Muter, V., Hulme, C., & Snowling, M. (1997). *The Phonological Abilities Test.* London: Psychological Corporation.

Muthén, L. K., & Muthén, B. O. (2005). *Mplus user's guide* (2nd ed.). Los Angeles: Author.

National Center for Education Statistics. (2004). *The condition of educa-*

'tion. Retrieved September 21, 2005, from http://nces.ed.gov/pubs2004/2004009.pdf

National Center for Education Statistics. (2005). *National assessment of educational progress, 2005, reading assessments.* Washington, DC: U.S. Department of Education, Institute of Education Sciences.

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

National Institute of Child Health and Human Development & U.S. Department of Education. (2003). *National symposium on learning disabilities and English language learners (2003). Symposium summary.* Washington, DC: Author.

Portes, A., & Rumbaut, R. G. (1990). *Immigrant America: A portrait.* Berkeley: University of California Press.

RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension.* Washington, DC: RAND Education.

Reading 44: A core reading framework. (1999). North Vancouver, British Columbia, Canada: School District No. 44.

Reese, L., Garnier, H., Gallimore, R., & Goldenberg, C. (2000). Longitudinal analysis of the antecedents of emergent Spanish literacy and middle-school English reading achievement of Spanish-speaking students. *American Educational Research Journal, 37,* 633–662.

Rosner, J., & Simon, D. (1971). The auditory analysis test: An initial report. *Journal of Learning Disabilities, 4,* 384–392.

Serpell, R., Baker, L., & Sonnenschein, S. (2005). *Becoming literate in the city: The Baltimore Early Childhood Project.* New York: Cambridge University Press.

Siegel, L. S. (1993). The development of reading. In H. W. Reese (Ed.), *Advances in child development and behavior* (pp. 63–97). San Diego: Academic Press.

Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development, 60,* 973–980.

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

Spies, R. A., & Plake, B. S. (Eds.). (2005). *The sixteenth mental measurements yearbook.* Lincoln: University of Nebraska 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.

Stuart, M. (2004). Getting ready for reading: A follow-up study of inner city second language learners at the end of Key Stage 1. *British Journal of Educational Psychology, 74,* 15–36.

Suarez-Orozco, C. (2004). Formulating identity in a globalized world. In M. Suarez-Orozco & D. B. Qin-Hilliard (Eds.), *Globalization: Culture and education in the new millennium* (pp. 173–202). Berkeley: University of California Press.

Swanson, H. L., & Alexander, J. (1997). Cognitive processes as predictors of word recognition and reading comprehension in learning disabled and skilled readers: Revisiting the specificity hypothesis. *Journal of Educational Psychology, 89,* 128–158.

Swanson, H. L., Saez, L., & Gerber, M. (2006). Growth in literacy and cognition in bilingual children at risk or not at risk for reading disabilities. *Journal of Educational Psychology, 98,* 247–264.

Thomas, W. P., & Collier, V. (1997). *School effectiveness for language minority students.* Washington, DC: National Clearinghouse for Bilingual Education.

Thomas, W. P., & Collier, V. P. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement.* Santa Cruz: University of California, Center for Research on Education, Diversity and Excellence.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Technical manual: Stanford Binet Intelligence Scale* (4th ed.). Chicago: Riverside.

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

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

Wade-Woolley, L., & Siegel, L. S. (1997). The spelling performance of ESL and native speakers of English as a function of reading skill. *Reading and Writing: An Interdisciplinary Journal, 9,* 387–406.

Wilkinson, G. S. (1993). *The Wide Range Achievement Test-3.* Wilmington, DE: Jastak Associates.

Willows, D. M., & Ryan, E. B. (1986). The development of grammatical sensitivity and its relation to early reading achievement. *Reading Research Quarterly, 21,* 253–266.

Woodcock, R. W. (1987). *Woodcock Reading Mastery Tests–Revised.* Circle Pines, MN: American Guidance Service.

Received August 30, 2004
Revision received May 25, 2007
Accepted May 26, 2007 ■