Teacher Characteristics, Classroom Instruction, and Student Literacy and Language Outcomes in Bilingual Kindergartners

Paul T. Cirino
University of Houston

Sharolyn D. Pollard-Durodola
Texas A & M University

Barbara R. Foorman
Florida State University

Coleen D. Carlson
David J. Francis
University of Houston

The Elementary School Journal
Volume 107, Number 4

0013-5984/2007/10704-0002$05.00

Abstract

This study investigated the relation of teacher characteristics, including ratings of teacher quality, to classroom instructional variables and to bilingual students' literacy and oral language outcomes at the end of the kindergarten year. Teacher characteristics included observational measures of oral language proficiency, quality, and classroom activity structure, as well as surveys of knowledge of reading-related skills. Student outcomes in both languages included letter naming, word reading, and phonological awareness and oral language composites. The study involved 141 teachers from a multisite project who were observed up to 3 times at the beginning, middle, and end of the year during their reading/language arts block while teaching English language learners to read in their primary language (Spanish) and/or in English. Teacher quality, but not teacher knowledge, was related positively to student engagement and negatively to time spent in noninstructional activities. Initial student and classroom performance, language of instruction and of outcomes, and teacher oral language proficiency in both Spanish and English predicted outcomes, whereas teacher quality was less related, and teacher content knowledge was consistently not related to student outcomes.

This study is about the effect of teacher quality and associated teacher characteristics on student language and literacy achievement in bilingual kindergarten classrooms. Under the No Child Left Behind Act (NCLB, 2001), states must ensure that school districts provide every child with a qualified teacher by the end of the 2005–2006 academic year. Conceptions of teacher quality vary across studies but typically include content knowledge as well as pedagogical skill (Shulman, 1986). We contend that discussions of teacher quality in bilingual classrooms must also address language of instruction and the oral-language proficiency of the teacher. The primary objective of this study was to examine how all of these characteristics affect kindergarten students' language and literacy end-of-year outcomes in bilingual classrooms, controlling for beginning-of-the-year abilities.

Teacher Quality and Related Characteristics

In 1966, Coleman et al. concluded that schools brought little to a child's achievement that was independent of the child's home background and social context. Yet, in subsequent years, studies have consistently reported positive associations between teacher characteristics and student achievement. These teacher characteristics include academic skills such as performance on the Scholastic Aptitude Test or National Teacher Examination (Ballou, 1996; Ehrenberg & Brewer, 1995; Ferguson, 1991; Ferguson & Ladd, 1996); years of teaching experience (Darling-Hammond, 2000; Hanushek, Kain, O'Brien, & Rivkin, 2005; Murnane & Phillips, 1981; Rivkin, Hanushek, & Kain, 2005; U.S. Department of Education, 2001); content knowledge, often measured by number of courses or major (Darling-Hammond, 2000; Goldhaber & Brewer, 2000; Monk & King, 1994); and participation in professional development activities (Cohen & Hill, 2000; Wenglinsky, 2002; Wiley & Yoon, 1995). Studies of the relation between teaching credentials and student achievement have reported mixed results (Darling-Hammond, Berry, & Thorenson, 2001; Goldhaber & Brewer, 2000), although NCLB defines teacher quality in terms of certification or licensure, in addition to having a bachelor's degree and content knowledge.

The inconsistency between the conclusion of Coleman et al. (1966) and findings of associations between teacher characteristics and student achievement has been addressed in recent multilevel analyses. Using data from 1,450 Virginia schools, Tuerk (2005) found that student poverty accounted for a relatively large percentage of the variance in pass rates in rural and urban schools (38% in eighth-grade writing and 15% of the variance in high school Standards of Learning [SOLs]), whereas teacher quality accounted for a relatively small percentage (7% in eighth-grade writing and 3% in high school SOLs). Foorman and colleagues found that teacher quality moderated relations between students' initial abilities and achievement outcomes in 17 high-poverty schools. In grades 1 and 2, students' initial reading ability and interactions of ratings of teacher quality with time allocation to instructional strategies predicted reading and spelling outcomes (Foorman, Schatschneider, et al., 2006). As in Tuerk (2005), effect sizes were small (1%–4%) for teacher quality and large (31%–50%) for initial reading ability. Teacher quality also predicted students' writing performance in grades 2 through 4 (Mehta, Foorman, Branum-Martin, & Taylor, 2005; Moats, Foorman, & Taylor, 2006). In a recent study by Borman and Kimball (2005), teacher quality affected grade 4 reading and grade 5 math performance with effect sizes of .21 and .11, respectively, but did not affect these academic skills in other grades. Teacher quality was measured in the Mehta et al. and Moats et al. studies by ratings of teaching effectiveness and the quality of instructional strategies, and in Borman and Kimball (2005) by standards-based evaluation ratings.

Primary-grade teachers' literacy knowledge is also a facet of teacher quality that is related to student achievement. Significant links between gains in teacher knowledge, ratings of teaching quality, and student achievement gains were found by McCutchen et al. (2002) and Foorman and Moats (2004). Other factors related to student achievement are time on task and student engagement (Fisher et al., 1980; Stallings, Robbins, Presbrey, & Scott, 1986; Wharton-McDonald, Pressley, & Mistretta Hampston, 1998) and percentage of time in small-group instruction (Taylor, Pearson, Clark, & Walpole, 2000).

In summary, associations between teacher characteristics (including quality broadly defined, and content knowledge) and student outcomes seem to contradict the Coleman et al. (1966) conclusion that schools contribute little to achievement beyond a student's home background. Although recent literacy studies that controlled for poverty or its proxy of initial reading ability have shown that (*a*) effect sizes are large for poverty or initial ability and small for teacher quality, and (*b*) access to qualified teachers appears to be limited by geography and by poverty of the school population, small effects for teacher quality do not minimize the cumulative effects of poor teaching (Sanders & Horn, 1998).

Teacher Characteristics and English Language Learners

Similar research to that described above, linking instructional quality and/or teacher characteristics with student achievement in bilingual settings, is needed. In such settings, teacher characteristics previously associated with English-only classrooms (and reviewed above) are likely to have an effect. However, language must also be considered directly in a variety of contexts, such as teacher knowledge of reading-related skills in both English and Spanish, teacher oral language proficiency in English and Spanish, the language of instruction, and the language of the outcome. We address the content of instruction in these bilingual classrooms in another article (Saunders, Foorman, & Carlson, in press). Our goal here is to bring the current framework on teacher quality to the study of kindergarten classrooms instructed in English and Spanish.

August and Hakuta (1998) noted that few studies have examined the relation of classroom and teacher characteristics to student outcomes for English language learners (ELLs) because most research on this population has focused on the language of instruction (English vs. Spanish). However, some more recent classroom observational studies have examined relations between teaching effectiveness (defined by those who implemented standards of effective pedagogy for bilingual classrooms) and student learning.

Padrón and Waxman (1999) found that effective teachers had more engaged third- and fourth-grade students who saw themselves as being more capable of completing class work and being able to read and who spent more time reading relative to their peers in non-standards-based classrooms. Estrada (2004) noted that students in 27 first- and fourth-grade classrooms were more successful when teachers changed their instructional behaviors to incorporate more effective instructional and organizational practices. In a third study, Hilberg, Doherty, Epaloose, and Tharp (2004) found that Latino students' gains in comprehension, spelling, reading, and vocabulary were related to the effectiveness of their teachers in implementing standards-based instruction. Finally, Graves, Gersten, and Haager (2004) rated classroom instruction in six areas (e.g., explicit teaching, sheltered English techniques, vocabulary development) and found that teachers with high scores had English learners who made relatively greater gains in oral reading fluency in grade 1. These studies, although small in number, suggest that effective teachers enhance student learning. However, it is less clear how these instructional characteristics affect student outcomes in the context of other factors, including teacher knowledge, language of instruction, teacher language proficiency, and student outcome language.

Research Questions

In general, we assumed that instructional variables found to be significant predictors in English-only classrooms would be relevant to kindergarten bilingual classrooms as well. Therefore, we hypothesized that instructional quality and teacher knowledge would be related to time on task, student engagement, and use of a small-group format. However, our primary hypothesis was that a variety of teacher characteristics, including quality ratings of teachers by trained observers across several measures, teacher knowledge, oral language proficiency, and language of instruction, would predict student outcomes (end-of-year performance) in academic areas especially important for bilingual kindergartners, including letter naming, phonological awareness, word reading, and oral language. We also expected that these relations might vary according to the language of the outcome measure, and some might moderate the effect of initial status on outcomes. We hypothesized that these relations would be evident even after accounting for beginning-of-the-year performances and the clustering of scores that results from students being nested within classrooms to receive instruction.

Method

This study was conducted as part of a large, multistate, multisite, longitudinal project focusing on language and literacy development in young bilingual (Spanish/English) English language learners from kindergarten through second grade. Data were collected during the 2002–2003 academic year.

Participants

Sites and schools. Sites were selected from areas where large numbers of Spanish-speaking bilingual students go to school—Los Angeles, and Houston, Austin, and Brownsville, Texas. Within these sites, schools were selected where (*a*) 40% of the students were Latino, (*b*) 30% of the kindergarten students were considered limited English proficient, and where (*c*) the schools were performing adequately or better on their respective state assessments. To match the California and Texas assessments, we used data from schools in a large urban district in Texas that administers the Stanford Achievement Test, which forms the basis of California's state assessment. We limited our sample to schools that scored above this minimum threshold because this ensured that we would select schools that were functional and that did not seriously lack instruction and/or leadership. It is important to note that, although many Spanish-speaking students find themselves in more poorly performing schools, the schools in our sample were not anomalies, or atypically performing schools. Schools across sites were not characterized by a uniform curriculum, although most reading curricula emphasized foundational skills (e.g., phonological awareness, alphabetic knowledge, word study, and oral language development) in either the primary or secondary language.

Across sites, 35 schools were represented (12 from Los Angeles, 6 in Austin, 11 in Brownsville, and 6 in Houston). The average percentage of Hispanics across these schools was 85% (SD = 16%). The average population of Spanish-speaking ELLs in kindergarten and first grade was 66% (SD = 22%) and ranged from 31% to 100%. All schools participated in the free or reduced-price lunch program, and the average proportion of students who qualified was 89% (SD = 13%) and ranged from 59% to 100%.

Teachers. A total of 141 teachers contributed data to the study; all were from the schools noted above. Of these, 50 were from Los Angeles schools, 41 from Brownsville, and 25 each were from Houston and Austin schools. Each school was represented by between two and nine teachers. The sample was predominantly Hispanic (76%) and female (91%). The average number of years taught overall was 9.7 (SD = 7.5) years. In terms of teaching credentials, 21% were certified in both bilingual and ESL areas, 39% were certified only as bilingual teachers, and 6% were ESL certified only; 31% of the remainder had some other credential, and 3% were uncredentialed. One hundred five (75%) were kindergarten teachers, with the remainder teaching grade 1. Eighty-three of the kindergarten teachers were linked to students (see below). These were the primary teachers of interest, and they had taught kindergarten an average of 6.3 (SD

= 5.4) years. Kindergarten teachers may not have been linked to students if they taught subjects other than reading/language arts (RLA), taught students who were not available at both testing dates, or entered or left the school in the middle of the year. In general, these linked teachers were similar to the full sample of teachers in terms of experience, credentialing, and performance on measures of quality, oral language proficiency, and knowledge. Given these similarities, nonlinked kindergarten teachers and grade 1 teachers were retained for their contribution to descriptive data on the teacher measures, as well as for comparisons among teacher-only measures, thereby increasing power for these analyses.

The stated instructional program models of the 83 teachers who were linked to students included dual language (15), immersion (30), maintenance (5), and transitional (33). Based on observations from the Timed Observations of Student Engagement (TOSE, see below), the following pattern was observed: 28 teachers taught only English language development (ELD) or reading/language arts (RLA) with an emphasis on English skills; 25 teachers taught only RLA with an emphasis on Spanish skills; and 30 teachers taught both types of classes. As determined jointly by the language model of the classroom and observations from the TOSE, 32 teachers taught predominantly in English (30 immersion, 1 dual language, and 1 transitional), and 51 taught predominantly in Spanish (5 maintenance, 14 dual language, and 32 transitional).

Students. A total of 1,451 Spanish-speaking English language learning (ELL) kindergarten students (50% female, all were Hispanic) from the 35 schools were assessed (range = 24 to 73 students per school). Students were linked to their RLA teacher rather than their classroom teacher, although these were often the same person. Of these students, 1,296 were matched to one of the kindergarten teachers described above and did not change throughout the year. In the remaining cases, teachers left or were replaced, or students changed RLA teachers. Of these 1,296 students, 1,156 provided data in at least one language at both testing waves. In general, these 1,156 students did not differ at pretest from those who left during the year, or whose teachers changed over the course of the year. A total of 83 kindergarten RLA teachers taught between one and 43 of these children (median = 15 students). The students' mean age was 5.5 years (SD = 0.3) in the fall of kindergarten, and 6.1 years (SD = 0.3) in spring. Of these students who were linked to teachers, 379 were linked to the 32 teachers who taught predominantly in English, and 777 were linked to the 51 teachers who taught primarily in Spanish.

Measures

Three groups of measures were employed, two for teachers and one for children. For teachers, the two groups were observational measures and teacher questionnaires. For children, measures consisted of language (e.g., phonological awareness, oral language skill) and achievement (e.g., alphabet knowledge, decoding) outcomes relevant in kindergarten, in both Spanish and English.

Teacher observational measures. Each measure was completed by a trained bilingual observer who had a bachelor's degree or higher and had used the instrument during the study's pilot year. Training consisted of 2 days at a central site, with an additional week of training at the observer's site. Two additional days at the beginning of each wave were used for "booster" training, and observers met with coordinators weekly throughout the year to address any difficulties. Observers were not allowed to conduct field observations until their interrater reliability on all instruments was 80% or better. In most cases where multiple observations were taken, the same observer assessed the same teacher at each observation. The average total minutes of observation per teacher was 378.27 (SD = 160.3). Eighty-nine percent of teachers were observed three times.

We used the Texas Teacher Appraisal System (TTAS; Texas Education Agency, 1984) to measure teacher quality in four primary domains (instructional strategies, classroom management and organization, presentation of subject matter, and learning environment) at the end of the school year. Within each domain, subheadings are provided (e.g., "Provides Opportunities for Students to Participate Actively and Successfully"), within which items ask whether several activities were present (e.g., "teacher solicits student participation," "implements at appropriate level"). Activities are summed into a total score, and a quality ranking is also solicited in which the observer describes effectiveness within each primary domain (e.g., classroom management) with a five-point Likert scale ranging from (1) "bottom 10%" to (5) "top 10%" (similar in scale to the oral language proficiency quality totals described below).

Coefficient alphas for all scales from this and other measures were computed for internal consistency reliability using SAS (SAS Institute, Inc., 2001). Coefficient alphas of the four primary domains were $\alpha = .70$ (instructional strategies), $\alpha = .83$ (classroom management and organization), $\alpha = .78$ (presentation of subject matter), $\alpha = .73$ (learning environment), and of the total item composite with $\alpha = .89$. Only the composite was retained, given the strong correlation of the four scales with the total item composite score (median $r = .78$). Quality rankings averaged within each of the four primary domains also correlated highly with one another ($r = .68$ to $r = .80$) and so were combined into a single quality score with a coefficient alpha of $\alpha = .93$ (quality total). Total item composite and quality total scores correlated $r = .78$ with one another, and, given the focus on teacher quality in this study, only the quality total dependent variable was retained for future analyses.

The Checklist of Teacher Competencies (CTC) assesses teaching competencies across several items in each of five domains (planning, management, instruction, mentoring of students, and personal characteristics). The CTC was completed by trained observers once during the middle of the school year. Each rater responded to each item using a five-point Likert scale ranging from (1) "never" to (5) "all the time." In addition, there was an optional choice for "no opportunity to observe." The median internal consistency value of the five domains was $\alpha = .85$, although a total score (total CTC) correlated $r = .81$ to .91 with the domain scores and showed $\alpha = .95$, and so we retained only the total CTC score as a dependent variable in future analyses.

The Timed Observations of Student Engagement (TOSE) measures classroom literacy instruction using an observation schema developed by Foorman and colleagues (Foorman, Goldenberg, Carlson, Saunders, & Pollard-Durodola, 2004; Foorman & Schatschneider, 2003). Trained observers used the TOSE three times during the year (fall, winter, and spring) to record within-the-minute the instructional format and content of reading/language arts instruction, whether students were on task, and the language the teacher used. The 28 teachers who taught only ELD or RLA with an emphasis on English skills used English 97% of the codable time, Spanish 1%, and mixed 2%. The 25 teachers who taught only RLA with an emphasis on Spanish skills used Spanish 72% of the codable time, English 17%, and mixed 11%. Proportions of language use varied among the 30 teachers who taught both types of classes, with four teachers having distributions where total English usage exceeded total Spanish plus mixed usage and the remaining 26 showing the reverse pattern. Therefore, we grouped the former teachers ($n = 4$) with the 28 who taught only ELD or RLA with an emphasis on English, and the latter teachers ($n = 26$) with the 25 who taught only RLA with an emphasis on Spanish. These results highlight the need to consider both the language model employed as well as actual language usage in the classroom (Foorman, Carlson, & Santi, in press).

For this study, we derived three observational measures from the TOSE—student engagement, small-group instructional format (where academic content was taught), and noninstructional time (e.g., intercom announcements, lining up, washing hands, chaos). Each measure was computed as a proportion by dividing the number of times that an identified student was on task, that a small group was observed, and that noninstructional time was observed, within a given minute (scored 1 or 0), by the total number of minutes of observations summed across RLA blocks and across observation waves. Codes are not mutually exclusive. Each index reflects the proportion of total instructional time observed for the academic year that was spent (*a*) in small-group instruction, (*b*) in noninstructional activities, and (*c*) with students actively engaged.

In addition to these minute-by-minute ratings, trained observers also rated the quality of instruction in reading, spelling, language, and writing, as well as overall instruction. Raters responded to each of these items via a five-point Likert scale ranging from (1) "bottom 10%" to (5) "top 10%" (similar to the scales of the TTAS), with an additional choice for "cannot determine." Two measures were derived—the average of quality rankings across the four content domains assessed (reading, spelling, language, and writing, referred to as the average TOSE score, $\alpha = .88$), and an overall quality ranking. The overall quality ranking correlated $r = .89$, $p < .01$ with the average TOSE, which contained more information and so was the only measure retained for further analyses.

The Total Teacher Quality Composite (Total Teacher Quality) was derived from the average of the TTAS quality total, total CTC, and the average TOSE. Total Teacher Quality had a scale of (1) "bottom 10%" to (5) "top 10%," which was the scale of all three of the constituent measures. The median correlation of these three scores was $r = .54$, $p < .01$, and their median correlation with the composite was $r = .84$, $p < .01$. We used this composite in most of the analyses that followed.

The Oral Language Proficiency (OLP) scale is a brief questionnaire completed by a trained observer that assesses the language quality, fluency, and facility of teachers during their reading instruction. The OLP was administered once in the middle of the year. Teachers are rated on a five-point Likert scale within each of four domains (fluency, vocabulary, pronunciation, and grammar), with varying descriptors depending on the domain. Separate ratings are provided for both Spanish and English, where relevant. Scales were created within each language across the four domains, and the coefficient alphas were both $\alpha = .92$ (OLP: domain totals). In addition, each teacher was rated on an overall five-point Likert scale ranging from (1) "bottom 10%" to (5) "top 10%" for each language (OLP: quality totals). The within-language correlations of domain totals with quality totals were $r = .76$, $p < .01$ (Spanish) and $r = .74$, $p < .01$ (English). Given that a primary focus in this study was the attainment of overall teacher quality rankings, we focused analyses on the quality totals. The correlation of these totals across languages was low ($r = -.02$, $p > .05$), suggesting the need to measure oral language proficiency separately in each language.

Teacher questionnaires. The Background Information Survey (BIS) collects information (gender, ethnicity, degree) on participating teachers at the beginning of the school year. Information such as years teaching each grade, credentialing, certification, and specialized training was also collected. We used it in the current study to provide descriptive demographic data.

The Beginning of Year Survey (BOYS) is a questionnaire that assesses teacher knowledge across 70 questions (35 each in Spanish and English) in a variety of domains related to reading. Previously used with third- and fourth-grade teachers in poor urban English-only schools in Houston and Washington, DC, this instrument measures phonological, orthographic, and morphographic knowledge (Foorman & Moats, 2004) and was modified to assess this knowledge in both Spanish and English. Teachers record the number of phonemes in a given word (phoneme counting), the number of syllables in a given word (syllable counting), match a specified sound from one word to one of four words in a multiple-choice format (phoneme matching), break a written word into its constituent phonemes and determine if the word is phonetically irregular or not (recognition of sound symbol correspondences), and answer multiple-choice questions in which they are asked to analyze student errors in spelling, oral reading, comprehension, and writing in both Spanish and English using hypothetical student work samples. Items were scored as correct or incorrect and summed to generate a percentage correct total score.

Coefficient alpha for the 35 English items was $\alpha = .72$, and for Spanish, $\alpha = .79$. Two English and three Spanish items were found to be poor items, and the coefficient alpha for the remaining 33 English items was $\alpha = .77$, and for the remaining 32 Spanish items, $\alpha = .84$. The correlation between the Spanish and English scales was significant ($r = .29$, $p < .01$) but low considering the scale structural similarity. This pattern suggested that knowledge of reading-related skills was relatively specific to the language in which it was assessed, and therefore we used separate scores for Spanish and English for analyses.

To summarize, the teacher observational variables used in further analyses included (*a*) Total Teacher Quality Composite (range = 1–5), (*b*) OLP quality totals in Spanish and English (range = 1–5), and (*c*) proportions of student engagement, small-group instructional format, and noninstructional time from the TOSE (score range = 0–1). For the questionnaires, total percentage scores (range 0 to 1) of the teacher knowledge questionnaire (BOYS) in both Spanish and English were used.

Student achievement. To assess letter name identification, we asked students to identify each of the 26 letters of the English alphabet and each of the 30 letters of the Spanish alphabet. This is a key skill that kindergartners are expected to master and on which they should demonstrate substantial growth. Internal consistency reliabilities (coefficient alpha) across languages and testing waves ranged from 0.94 to 0.97. Dependent measures were the raw score totals.

The Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999) has nine subtests measuring phonological awareness (PA), rapid naming (RN), and phonological memory (PM). Phonological awareness has consistently been identified as an important predictor of early literacy, particularly before and as individual word reading skills are developed. The normative base consisted of 1,656 individuals from ages 5 to 24, similar to the 1997 U.S. Census statistics. Coefficient alphas for all three composites in the entire normative sample ranged from .83 to .95, and .83 to .92 in the age range of this sample; test-retest estimates in a small sample ($n = 32$) of children ages 5 to 7 ranged from .70 to .92 for the three composites. Additional content, concurrent, predictive, and construct validity data are provided in the CTOPP manual (Wagner et al., 1999).

For this study, we used five subtests, including elision, blending words, blending nonwords, segmenting words, and sound matching (first sound and last sound). Raw percentage correct scores were used to compare performance with a Spanish language version of this instrument (see below). We created a phonological awareness (PA) composite from the subtest scores, and branching rules enabled us to shorten testing time, based on empirical modeling of performance on this test using item-response methods (Schatschneider, Francis, Foorman, Fletcher, & Mehta, 1999), as well as work in the parent project examining the properties of this assessment in a larger sample (n = 1,600) of bilingual students. Further details on the computation and use of this PA composite are available in Vaughn et al. (2006).

The Test of Phonological Processing—Spanish (TOPP-S) was developed to align with the English CTOPP in terms of the skills being addressed and the linguistic complexity of the items within each subtest, as well as the number of items, while still being appropriate for the Spanish language. With the exception of sound matching, all subtests were built entirely of production-based items. Reliability estimates for the TOPP-S were determined on a sample of approximately 1,500 students, and the coefficient alphas were very high, ranging from 0.93 to 0.97. Branching rules similar to those used for the CTOPP were also employed for the TOPP-S, and a PA composite in Spanish was created using the analogous subtests and the same rules as the English PA composite.

The Experimental Word List was developed by first generating words from kindergarten to grade 3 instructional cumulative vocabulary materials. The generated word list was then matched to the LEXESP database of printed word frequencies in Spanish (Sebastián, Cuetos, Martí, & Carreiras, 2000), similar to the Zeno, Ivens, Millard, and Duvvuri (1995) database of word frequencies in English. We selected 40 words from the kindergarten to grade 3 instructional corpus with varying probability depending on grade and printed word frequency, with the final list consisting of 40 words representing a diversity of linguistic features, ordered by difficulty to span kindergarten to third grade. This measure has very high reliability in grades kindergarten through 3 (internal consistency was over .90 in the parent project sample with a total sample size of 4,003 students). This is the only measure of actual word reading we administered, and we chose it because of its ability to measure the skill as it is emerging for most kindergarten children.

We used the Woodcock Language Proficiency Battery—Revised: English and Spanish Forms (WLPB-R; Woodcock, 1991; Woodcock & Munoz-Sandoval, 1995) to assess oral language. Oral language is clearly an important skill for kindergartners, particularly because many early literacy predictors are assessed orally, before children develop literacy skills. The WLPB-R is a well-standardized instrument whose normative sample, concordant with 1980 U.S. Census statistics, consisted of 6,359 subjects (3,245 in K to 12) and was the same as that of the Woodcock-Johnson Psychoeducational Battery—Revised (Woodcock & Johnson, 1989). Median coefficient alphas range from .81 to .92 across all age ranges (and from 0.77 to 0.96 at ages 6 to 9) for the subtests used; test-retest measures for selected subtests in a sample of 504 ranged from 0.75 to 0.95. In addition, content, concurrent, and construct validity data are also available in the WLPB-R manual (Woodcock, 1991). All measures were available for Spanish and English, and the scaling process allows scores on the English and Spanish language assessments to be directly compared in the sense that it places the Spanish language norms on the same scale as the English language norms. Details are provided in the test manuals (Woodcock, 1991; Woodcock & Munoz-Sandoval, 1995). Subtests of the WLPB-R employed in this study were measures of oral language, including the subtests of listening comprehension, picture vocabulary, and verbal analogies. Also, memory for sentences was administered in the fall only. We computed an oral language composite standard score based on the average W score of measures available at each testing time; it was the dependent measure used in analysis.

Analysis Approach

We used descriptive analyses to detail initial and final performance on the student variables and to describe performance on the teacher variables as well. Correlational analyses addressed the hypothesis regarding the relation of teacher quality and knowledge to classroom observations. Data related to the hypothesis concerning the relation of teacher quality to student outcomes were analyzed in a multilevel framework using the mixed procedure of SAS (Littell, Milliken, Stroup, & Wolfinger, 1996; SAS Institute, Inc., 2003), which assesses not only the degree to which performance varies among students within the same classroom but also among students in different classrooms. These analyses proceeded in several steps. Most important, analyses took into consideration both the language of instruction and the language of the outcome. In all, 32 teachers of 379 students taught predominantly in English, and 51 teachers of 777 students taught predominantly in Spanish. We measured each outcome both in English and in Spanish.

We built models of the effect of teacher quality on student outcome in four steps. In Step I, we computed the unconditional intraclass correlation coefficient (ICC) as a measure of the effect of nesting students within classrooms (Singer, 1998). All models used a full maximum-likelihood approach to deriving model fit and parameter estimates. Step II added a measure of fall performance as a student-level covariate with both fixed and random effects at the classroom level. The *fixed* effect of the covariate represents the average relation of initial student performance to student outcomes *within* classrooms, whereas the *random* effect represents the degree to which this relation varies *across* classrooms. Where this random effect (in addition to the intercept) was significant, it suggests that classrooms differed in the relation of the covariate to the dependent measure. When the random effect was not significant, it suggests that the relation of the covariate to the outcome did not differ across classrooms. Step III added covariates of fall performance representing the degree to which *classroom* means differed on initial performance (i.e., between-classroom variance standardized across classrooms) as a fixed effect. Steps I, II, and III together describe the degree of clustering in the data and the effect of the initial performance covariate on outcomes.

Finally, in Step IV, we added teacher-specific fixed effects. These were the primary models of interest insofar as they test the effect of teacher-level characteristics (quality and language of instruction) in explaining student performance over and above the effects of student- and classroom-level covariates representing the effect of students' initial status on outcomes. Interactions of each fixed effect with both outcome language and the student-level, within-classroom covariate of initial status were also considered. We considered fixed effects significant at an alpha level of .01, in an effort to balance Type I and Type II errors given the number of terms in Step IV models. Where fixed effects or interactions that involved classificatory factors were significant, we conducted post-hoc pairwise comparisons employing a Tukey-Kramer family-wise correction at .05. Higher-level, nonsignificant interaction effects were trimmed, to yield the final models for four outcomes: letter name identification, phonological awareness (PA), experimental word list, and oral language composite.

To make more explicit the models being fit in Steps I–IV, we provide equations that describe the four steps for the phonological awareness outcome. Models are presented in hierarchical form. Equations are numbered to reflect the step in the model-fitting process, which is given to the left of the decimal, and to reflect the level in the hierarchy, which is given as the number to the right of the decimal, with X.1 being the student level and X.2 being the classroom level.

Step I:

$$PA_{2ijk} = \beta_{0jk} + \varepsilon_{ijk}, \quad \text{(I.1)}$$

$$\beta_{0jk} = \gamma_{00k} + \delta_{0jk}, \quad \text{(I.2)}$$

where PA_{2ijk} is the phonological awareness score at time 2 for student i in classroom j

measured in language k, and β_{0jk} is the average phonological awareness score at time 2 in classroom j in language k, and γ_{00k} is the grand mean of phonological awareness in language k taken across all classrooms. The terms ε_{ijk} and δ_{0jk} are random effects and indicate the extent to which individual subjects deviate from the mean of their classroom, and the extent to which a given classroom mean deviates from the grand mean for all classrooms, respectively. They are assumed to be distributed normally, to have mean 0, and their variances are estimated as part of the model-estimation process. In this study, we estimated these variance components in such a way that there was one estimate of the level 1 residual variance ($\text{Var}(\varepsilon_{ijk})$) but separate estimates of the variance in random effects at level 2 ($\text{Var}(\delta_{0jk})$) for each language k (k = English, Spanish).

To ascertain the effect of clustering, we computed the ICC for language k from the variance components computed in Step I. Specifically, the ICC for language k is given by the following formula:

$$\text{ICC}_k = \text{Var}(\delta_{0jk}) \,/\, [\text{Var}(\delta_{0jk}) + \text{Var}\varepsilon_{ijk}],$$

where the terms are as described in the preceding paragraph. It is important to note that the second term in the denominator is the same for both languages, whereas the variance in classroom means (i.e., the numerator term and the first term in the denominator) is specific to each language. This ratio of variance in the classroom means relative to total variance in the outcome gives an index of the degree to which students in the same classroom are more alike in their performance than students in different classrooms. It is the variability in classroom means that can be explained by teacher- or classroom-level predictors such as teacher quality, teacher knowledge, and teacher language. In contrast, the residual or within-classroom variance is explained by introducing student-level predictors such as initial status on the outcome, which is accomplished in Step II.

Step II:

$$\text{PA}_{2ijk} = \beta_{0jk} + \beta_{1jk}\,(\text{PA}-\text{W}_{1ijk}) + \varepsilon_{ijk,} \quad \text{(II.1)}$$
$$\beta_{0jk} = \gamma_{00k} + \delta_{0jk}, \quad \text{(II.2a)}$$
$$\beta_{1jk} = \gamma_{10k} + \delta_{1jk}, \quad \text{(II.2b)}$$

where PA_{2ijk} is, as before, the phonological awareness score at time 2 for student i in classroom j measured in language k, and PA-W_{1ijk} is the corresponding score for the same student at time 1. Because PA-W_{1ijk} is centered at the classroom mean for language k, it has an overall mean of 0 for each language, and classrooms do not differ on average on this variable in either language. We designate this within-classroom centering by the W following the name of the measure. Because of the within-classroom centering of PA-W_{1ijk}, the coefficients have the following interpretations: β_{0jk} remains the average phonological awareness score at time 2 in classroom j in language k, and β_{1jk} is the effect of PA-W_1 in classroom j in language k. Both β_{0jk} and β_{1jk} are allowed to vary across classrooms as evidenced by the random effects in the level 2 equations for these terms, δ_{0jk} and δ_{1jk}, in equations II.2a and II.2b, respectively. The coefficient γ_{00k} in equation II.2a gives the average value in the outcome across the set of classrooms, whereas the coefficient γ_{10k} in equation II.2b gives the average effect of student-level initial status across the set of classrooms, respectively. If the variance in the random effects of equation II.2b is 0 (i.e., $\text{Var}(\delta_{1jk}) = 0$), then the effect of student initial status is the same across classrooms, and the random effects for initial status would be dropped from the model at this step. Step II models improve the prediction of student performance because they predict end-of-year performance from both the classroom average as well as from students' initial status in the fall. As a result, Step II models will produce a reduction in $\text{Var}(\varepsilon_{ijk})$ to the extent that initial status predicts final status.

Step III: Unlike the models in Step II, which build up the level 1 (i.e., student level) equation, Step III models build up the level 2 equation for the classroom mean by adding initial status in language k at the classroom level as a predictor of end-of-year performance in that language.

$$PA_{2ijk} = \beta_{0jk} + \beta_{1jk}(PA - W_{1ijk}) + \varepsilon_{ijk}, \quad (III.1)$$
$$\beta_{0jk} = \gamma_{00k} + \gamma_{01k}(PA - B_{1jk}) + \delta_{0jk}, \quad (III.2a)$$
$$\beta_{1jk} = \gamma_{10k} + \delta_{1jk}, \quad (III.2b)$$

where all terms from prior steps retain the same meaning, with the exception of $\gamma_{00k,}$ which now gives the conditional grand mean for end-of-year performance, or, put another way, the mean at the end of the year for classrooms performing at the mean at the of beginning of the year. The only new terms, PA-B_{1jk} and γ_{01k}, represent the mean initial status in classroom j in language k, and its effect on end-of-year performance in language k, respectively. Thus, Step III models are expected to reduce the variability in classroom means in language k at the end of the year. That is, Step III models are predicted to reduce Var(δ_{0jk}) relative to Step II. It is the Var(δ_{0jk}) at Step III that is available to be explained by classroom-level predictors to be entered in Step IV.

Step IV: Like Step III, the models at Step IV build up the level 2 equation for the classroom mean (eq. IV.2a). However, in addition, interaction effects between teacher predictors and the student covariate were also considered in Step IV. Thus, unlike Step III, Step IV also built up the level 2 equation for the effect of the student covariate (eq. IV.2b). We present here the general model under consideration at Step IV in terms of the variables used to explain the classroom mean in language k (β_{0jk}) and the effect of the student covariate of initial status (β_{1jk}).

$$PA_{2ijk} = \beta_{0jk} + \beta_{1jk}(PA - W_{1ijk}) + \varepsilon_{ijk}, \quad (IV.1)$$
$$\beta_{0jk} = \gamma_{00k} + \gamma_{01k}(PA - B_{1jk}) + \gamma_{02k}BOY_{jk} + \gamma_{03k}BOY_{jk'} + \gamma_{05k}OLP_{jk'} + \gamma_{04k}OLP_{jk} + \gamma_{06k}TQC_{j} + \gamma_{06k}TL_{j} + \gamma_{07k}OL_{jk} + \delta_{0jk}, \quad (IV.2a)$$
$$\beta_{1jk} = \gamma_{10k} + \gamma_{11k}BOYS_{jk} + \gamma_{12k}BOYS_{jk'} + \delta_{1jk'}, \quad (IV.2b)$$

where $BOYS_{jk}$ is teacher knowledge for teacher j in language k and $BOYS_{jk'}$ is teacher knowledge for teacher j in the language other than the outcome language (i.e., if k = English then k' = Spanish, and vice versa), OLP_{jk} is oral language proficiency for teacher j in language k, and $OLP_{jk'}$ is oral language proficiency for teacher j in the language other than the outcome language (i.e., similar to coding for BOYS), TQC_j is the total quality composite rating for teacher j, TL_j is language of instruction used by teacher j, and OL_{jk} is the language of the outcome and has the same value for all teachers for any particular outcome. The coefficients γ_{0pk} give the effect of variable p on the classroom means in language k. In addition to the terms presented in equation IV.2a, models in Step IV included interactions of OL_{jk} with variables in equation IV.2a as additional terms in equation IV.2a, and interactions of variables in equation IV.2a with student initial status. These latter interactions appear as effects of the variable on the coefficient to the left of the equals sign in equation IV.2b. We show these "interactions" in equation IV.2b above for $BOYS_{jk}$ and $BOYS_{jk'}$.

The models of Step IV reduce variability in the classroom means based on the predictors entered into equations IV.2a and IV.2b. By comparing the variance of the residual term in equation IV.2a (Var(δ_{0jk})) with that from equation 1.2a, we can determine the variance in classroom means in each language accounted for by the complete set of predictors in equations IV.1, IV.2a, and IV.2b. By comparing the variance of the residual terms between the models of Steps IV and III, we can determine the percentage of variance in the classroom means in each language explained by measures of teacher quality and instruction over and above the contributions of initial status at the student and classroom levels. We can also use the variance components from the Step IV equations to recompute the ICC and thereby determine the effects of clustering when teacher and student characteristics are accounted for in the model. This latter number is important for effective design of treatment studies that might consider using the predictors of Step IV as covariates in a model to assess the effect of a classroom-level intervention.

Results

Descriptive Analyses

Many teacher measures used in this study were new, and performances on these predictor measures have not been reported in a sample this large of this age and population. We examined distributions through box and stem-and-leaf plots, as well as normative statistics (e.g., skewness, kurtosis). Performances are presented in Table 1. The beginning-of-year knowledge survey (BOYS) in Spanish was negatively skewed (an accumulation of scores toward the high end of the score distributions), which indicated that most teachers showed good knowledge of these skills. Rankings of oral language proficiency (OLP) had a limited range of responses, were not available in both languages for all teachers (e.g., Spanish OLP was not ranked for a teacher who used only English in the classroom), and so were considered categorically (high, low, unobserved). All other teacher measures were normally distributed across nearly every index examined. Quality rankings in general corresponded to categories of between (3) "average" and (4) "above average." Overall, students were engaged during much of the time that they were observed (M = 96%, SD = 4%). Small-group instruction was observed infrequently but with considerable variability (M = 20%, SD = 21%), and nonreading instruction was observed approximately 10% of the time (M = 12%, SD = 6%).

Student performances are presented in Table 2. In the fall, scores on Spanish and English versions of the same measure appeared similar in their distributions (e.g., letter name identification in Spanish and English). Raw score measures (letter name identification, experimental word list, PA composite) were positively skewed toward the low end of the performance distribution, as expected. Oral language composite standard scores were normally distributed, albeit with mean performance well below the normative population mean values. By spring testing, letter name identification was negatively skewed (i.e., most students learned this skill), whereas the PA composites were much more normally distributed relative to the fall testing. Distributions for experimental word lists remained positively skewed but with greater variability relative to the fall performance distribution. Oral language composites were again normally distributed in the spring, although little improvement was apparent for this measure in either language.

Primary Analyses

Teacher characteristics and classroom observational data. First, analyses focused on the relations of teacher quality, knowledge, and oral language proficiency to observable classroom characteristics (e.g., time on task, student engagement, small-group format). We expected that student engagement and small-group instruction would be positively related to measures of quality and knowledge, such that higher-quality teachers would have students who were more frequently on task or who were instructed in small groups. In contrast, we expected that noninstructional time would be negatively related to teacher quality, with more effective teachers spending more time in instruction.

Results were similar whether all teachers (N = 141), only kindergarten teachers (N = 105), or only kindergarten teachers with students "linked" to them (N = 83) were considered (see Table 3). Analyses by subgroups according to teacher's language of instruction (e.g., examining student engagement during RLA in Spanish for teachers who taught only, or predominantly, in Spanish) were considered, although these results were similar to those in Table 3, so only the global analyses are presented.

Among teachers in this study, overall quality was positively related to overall student engagement ($p < .01$), such that more

Table 1. Descriptive Data for Teacher Measures ($N = 141$)

Measure	N	M	SD
Observations:			
Checklist of Teacher Competency (CTC)	133	3.58	.55
Texas Teacher Appraisal System (TTAS)	137	3.42	.54
Timed Record/Student Engagement (TOSE)	141	3.22	.39
Total teacher quality (TTAS, CTC, TOSE)	141	3.39	.41
Oral language proficiency:			
Spanish	93	3.33	.68
English	121	3.44	.58
Questionnaires:			
Beginning of Year Survey:			
Spanish	124	.82	.14
English	136	.71	.15

Note.—Range of possible scores for observation measures = 1–5; range for questionnaire = 0–1. See text for derivation of variables from questionnaires and observation measures.

Table 2. Descriptive Data for Achievement Tests ($N = 1,156$ Students)

		Fall		Spring	
Language/Measure	N	M	SD	M	SD
Spanish:					
Letter name identification	1,143	8.49	7.96	20.29	10.11
Phonemic awareness composite	1,148	21.27	16.83	44.90	20.28
Experimental word list	1,114	1.07	4.69	12.74	15.02
Oral language composite	1,124	75.47	22.34	81.33	22.71
English:					
Letter name identification	1,137	9.43	9.30	16.34	9.35
Phonemic awareness composite	1,142	21.12	15.65	40.64	16.70
Experimental word list	1,105	.99	3.08	6.67	7.33
Oral language composite	1,120	55.56	19.69	59.66	20.51

Note.—Means and standard deviations are for students with data at both time points. Numbers for phonemic awareness composites are percentage correct (range = 0–100), from the subtests of the Test of Phonological Processing (TOPP-S), or the Comprehensive Test of Phonological Processing (CTOPP). Letter name identification scores are raw scores (range = 0–26 in English, 0–30 in Spanish). Experimental word list scores are also raw scores (range = 0–40). Oral language composite scores, from the Woodcock Language Proficiency Battery—Revised (WLPB-R), are standard scores ($M = 100$, $SD = 15$).

Table 3. Correlations of Teacher Quality and Knowledge with Minute-by-Minute Observations

Observation/Measure	Quality	BOYS English	BOYS Spanish
All teachers ($N = 141$):			
Student engagement	.36**	.11	.14
Small-group instruction	−.08	.11	−.05
Noninstructional time	−.19*	−.19*	−.02
Teachers linked to students ($N = 83$):			
Student engagement	.19	.02	.26*
Small-group instruction	−.05	.03	−.05
Noninstructional time	−.28**	−.17	−.07

Note.—Teacher quality = average composite score from Checklist of Teacher Competencies, Texas Teacher Appraisal System, and timed observations of student engagement. BOYS = Beginning of Year Survey.

* $p < .05$.

** $p < .01$ (other correlations are nonsignificant).

effective teachers had more engaged students. However, quality was unrelated to time of content instruction in a small-group format ($p > .05$). Time spent in noninstructional activities was significantly though modestly negatively related to quality ($p < .05$), with higher-quality teachers spending more time in content-based instruction. We noted the same pattern whether all teachers, only kindergarten teachers, or only teachers linked to students were considered. Although quality was related to these observational data, in general, teacher knowledge of reading-related information in either language was unrelated to observational data.

Teacher and language variables and student outcomes. The second set of primary analyses examined the relation of teacher quality, knowledge, oral language proficiency, and language of instruction to student outcomes (which may vary by language of the outcome), considering clustering effects and beginning-of-the-year student performance. Results of initial and final models for each outcome are presented in Tables 4 and 5. As shown under the Step I models in Table 4, intraclass correlations ranged from .09 (for experimental word list in English) to .62 (for letter name identification in Spanish), with a median value of .44 (.28 in English and .52 in Spanish). Such values highlight the benefit of a multilevel approach to these data and that roughly half of the variance in Spanish outcomes (less in English) was due to variability between classrooms, and about half the variance in Spanish outcomes (more in English) was due to variability within classrooms.

However, each Step II model showed an improvement in fit over Step I models, and the within-classroom covariate was highly significant in each case (all $p < .01$), as expected. In general, between-classroom variance estimated in Step II models did not change from Step I models, with the median across languages near zero (as expected), whereas the within-classroom variance decreased substantially (median = 24%, also as expected). Put another way, Step II models did not explain any of the variability in classroom means but accounted for approximately one-quarter of the variability between students in the same classroom (though only about 12% of the total variability in outcomes in either language was explained by within-classroom measures of initial status). In addition, each Step III model showed an improvement in fit over Step II models, and the between-classroom covariate was highly significant in each case (all $p < .01$), as expected. In general, within-classroom variance did not change from Step II models, with the median across languages near zero (as expected), whereas between-classroom variance decreased (as expected) in both Spanish (median = 38%) and English (median = 64%). Put another way, Step III models did not explain any of the explainable individual differences between students in the same classroom but did explain one-third to two-thirds of the explainable differences in classroom means, that is, differences between students in different classrooms (or an additional 19% of the overall variance in Spanish outcomes, and an additional 24% of the overall variance in English outcomes). Therefore, inclusion of a child's, and a child's classroom's, initial performance predicted students' end-of-year performance, highlighting its importance for assessing student outcomes at either the student or classroom level.

Step IV (final) models, which added teacher-specific characteristics of quality and language of instruction (that varied across but not within classrooms), appear in Table 4. Step IV models for all outcomes showed an improvement in fit over Step III models. Also, the within-classroom variance did not change from Step III models, with the median across languages near zero (as expected), but between-classroom variance decreased by approximately one-half in English (median = 49%) and three-quarters (median = 75%) in Spanish (or an additional 6% of the total variance in English outcomes and an additional 23% of the total

TABLE 4. Random Effects across Step I through IV Models, by Spanish and English Measures at End of Kindergarten

	Letter Name Identification			Phonological Awareness			Oral Language Composite			Experimental Word Reading		
Model	Effect	PE (*SE*)	ICC	Effect	PE (*SE*)	ICC	Effect	PE (*SE*)	ICC	Effect	PE (*SE*)	ICC
Step I	BE	44.72 (8.02)	.51	BE	.006 (.001)	.18	BE	189.30 (48.87)	.37	BE	9.96 (2.82)	.09
	BS	70.99 (12.49)	.62	BS	.01 (.002)	.28	BS	330.96 (81.86)	.50	BS	114.73 (20.36)	.54
	Res.	43.00 (1.36)		Res.	.026 (.001)		Res.	327.28 (10.40)		Res.	96.34 (3.08)	
Step IV	BE	5.31 (1.23)	.16	BE	.002 (.001)	.09	BE	10.82 (3.13)	.09	BE	1.36 (1.08)	.02
	BS	9.46 (1.95)	.25	BS	.005 (.001)	.21	BS	12.31 (3.44)	.10	BS	29.38 (5.77)	.28
	Res.	28.23 (.93)		Res.	.019 (.001)		Res.	112.83 (3.74)		Res.	77.27 (2.50)	

Explainable Variance Explained

Effect	S4S1	%	Effect	S4S1	%	Effect	S4S1	%	Effect	S4S1	%
BE	.88	.62	BE	.44	30	BE	.94	77	BE	.86	26
BS	.87	.67	BS	.55	34	BS	.96	81	BS	.74	49
Res.	.34		Res.	.27		Res.	.66		Res.	.20	

NOTE.—Effect = parameter term (BE = between classrooms, English language; BS = between classrooms, Spanish language; Res. = residual); PE/*SE* = parameter estimate and standard error of variability at the classroom level, or of the residual (at the student level); ICC = intraclass correlation coefficient, a measure of the degree to which students' scores within a classroom are related (estimate/estimate + residual) in unconditional means models and in final models; S4S1 = comparison of the models, indicating the size of the reduction in variability (parameter estimates) in Step IV relative to Step I (estimate Step I − estimate Step IV/estimate Step I); % = proportion of the explainable variance in Step I explained by effects in Step IV at the between- and within-classroom levels for each language.

variance in Spanish outcomes). These results suggest that teacher characteristics, language of instruction, and language of the outcome measure affect student outcomes, over and above initial performance. Altogether, models accounted for approximately 42% of the variance in English outcomes (12% + 24% + 6%) and 54% of the variance in Spanish outcomes (12% + 19% + 23%). Results by outcome measure are presented below, and these fixed effects are noted in Table 5.

Letter name identification: Main effects were noted for both of the initial-status covariates (both $p < .01$), oral language proficiency (OLP) in the language of the outcome ($p < .01$) and predominant teaching language ($p < .01$), and for outcome language (both $p < .01$). In addition, the interaction of the within-classroom covariate with OLP in the language other than the language in which the outcome was measured was statistically significant ($p < .01$), as were interactions of outcome language with the within- ($p < .01$) and between-classroom initial-status covariate ($p < .01$), the predominant language of the teacher ($p < .01$), and with OLP in the language of the outcome ($p < .01$).

Estimates for the effects of the initial-status covariates were positive, indicating that beginning-of-the-year letter-naming performances (both at the student and classroom levels) were positively related to letter-naming outcomes at the end of the year. For OLP of the teacher in the language of the outcome, estimates and Tukey-Kramer corrected *p*-values showed that students' letter naming in a given language was higher when their teacher received an

Table 5. Fixed Effects in Final Models

Effect	Letter Name Identification		Phonological Awareness		Oral Language Composite		Experimental Word Reading	
	df	*F*	*df*	*F*	*df*	*F*	*df*	*F*
Initial (within)	1,144	332.72**	1,153	381.16**	1,144	932.27**	1,100	157.02**
Initial (between)	1,1859	135.54**	1,1869	104.09**	1,1823	356.37**	1,1843	64.78**
OLP (same)	2,1859	8.86**	2,1869	5.54**				
OLP (different)							2,1843	9.37**
Total quality							1,1843	17.78**
Teacher language	1,1859	10.51**					1,1843	12.58**
Outcome language	1,134	9.45**	1,143	15.62**	1,134	298.01**	1,134	11.86**
Initial (within) * OLP (different)	2,1859	8.88**			2,1823	9.50**	2,1843	6.43**
Initial (within) * outcome language	1,144	7.01**					1,100	8.95**
Initial (between) * outcome language	1,1859	19.88**						
OLP (same) * outcome language	2,1859	8.88**						
Teacher language * outcome language	1,1859	7.29**			1,1823	10.06**	1,1843	16.12**

Note.—Effect = Type III term in final model (not all 22 terms are listed, including those not significant for any dependent measure). Main effect terms consisted of initial (within) and initial (between), representing the within-classroom covariate of initial status, and the between-classroom covariate of initial status, respectively; oral language proficiency (OLP) in the language of the outcome (same), and the other language (different); Beginning of Year Survey in the language of the outcome (same), and the other language (different); total quality (see Method for derivation); predominant teaching language of teacher (English or Spanish); and outcome language of the dependent measure (English or Spanish). Interactions of the within-classroom covariate, and of outcome language, with each of the other terms were also included in final models. Where values are not shown, they were not significant for that dependent variable. All interactions were trimmed from the Step IV phonological awareness model because none was significant. An asterisk between terms denotes an interaction, and asterisks after values indicate significance.

** $p < .01$.

OLP rating of either "high" ($p < .01$) or "low" ($p < .01$) relative to when the students' teacher was not observed to teach in that language. The estimate for teachers whose teaching language was predominantly English was lower than that for teachers whose primary teaching language was Spanish, and Tukey-Kramer corrected *p*-values indicated that this difference was statistically significant ($p < .01$). In other words, students' letter-naming outcomes, on average (i.e., across languages with respect to the outcome measures), were higher when the predominant teaching language was Spanish, relative to when it was English. Also, the estimate for English outcomes was higher than that for Spanish, and Tukey-Kramer corrected *p*-values indicated that this difference was significant ($p < .01$). Thus, students' English letter-naming outcomes were higher than they were in Spanish.

Regarding the interaction of the within-classroom initial status covariate and OLP in the language other than the one in which the outcome was measured, estimates were positive when this OLP was either "high" or "low" relative to when it was unobserved, indicating a stronger relation of this covariate to outcomes when students' teachers were rated. Regarding the interactions of language of the outcome and both the within- and between-classroom covariates, the estimates were positive when the outcome language was English relative to when it was Spanish, suggesting that the relation of these covariates to letter-naming outcomes was stronger in English than Spanish. For the interaction of outcome language and predominant teaching language, outcomes were highest for students of teachers who taught predominantly in Spanish, with outcomes measured in Spanish. These outcomes were higher than Spanish outcomes for students of teachers who taught predominantly in English and for outcomes measured in English in general, regardless of the predominant teaching language (all Tukey-Kramer adjusted $p < .01$). Finally, Tukey-Kramer corrected p-values of the interaction of outcome language with OLP in the language of the outcome suggested that outcomes were highest when they were measured in Spanish and when OLP was either "high" or "low" relative to all other combinations of outcome language and OLP (all $p < .05$).

Phonological awareness (PA): Outcome language and within-classroom initial-status covariates did not interact with other fixed effects in the final model. Main effects were noted for both initial-status covariates ($p < .01$), for oral language proficiency (OLP) in the language of the outcome ($p < .01$), and for outcome language ($p < .01$). Estimates for initial-status covariates were positive, indicating that beginning-of-the-year PA performances were positively related to PA outcomes at the end of the year. For teacher OLP in the language of the outcome, estimates indicated that students' PA in a given language was significantly higher when their teacher received an OLP rating of either "high" ($p < .05$) or "low" ($p < .01$) relative to when students' teacher was not observed to teach in that language. Finally, the estimate for English outcomes was significantly lower than that for Spanish ($p < .01$). In other words, students' English PA outcomes were lower than they were in Spanish PA at the end of kindergarten.

Oral language composite: Main effects were noted for both initial-status covariates and for outcome language (all $p < .01$). In addition, we also found interactions of the within-classroom covariate with teachers' OLP in the language other than the language in which the outcome was measured ($p < .01$), and of outcome language with the predominant language of the teacher ($p < .01$). Estimates for covariates were positive, indicating that oral language performances at the beginning of the year were positively related to oral language outcomes at the end of the year. The estimate for English outcomes was significantly lower than that for Spanish ($p < .01$), indicating that students' English oral language outcomes were lower than those of Spanish.

Regarding the interaction of the within-classroom covariate and teachers' OLP in the language other than the one in which the outcome was measured, estimates were positive when teachers' OLP was either "high" or "low" relative to when it was unobserved, indicating a stronger relation of this covariate to outcomes when teachers were rated. Regarding the interaction of outcome language and predominant teaching language, Spanish outcomes were higher than English outcomes regardless of whether predominant teaching language was English or Spanish (both $p < .01$, and as indicated by the main effect of outcome language), but the difference between Spanish and English outcomes was larger when the predominant teaching language was Spanish (approximately 21 standard score points) relative to when the predominant teaching language was English (approximately 12 standard score points).

Experimental word reading list: Main effects were noted for both initial-status covariates, teachers' oral language proficiency (OLP) in the language other than the language in which the outcome was measured, total teacher quality (all $p < .01$), and for predominant teaching language and outcome language (both $p < .01$). In addition, we obtained interactions of outcome language with the within-classroom initial-status covariate ($p < .01$), with predominant language of the teacher ($p < .01$), and with teachers' OLP in the language other than the one in which the outcome was measured ($p < .01$).

Regarding the main effects, estimates for covariates were positive, indicating that beginning-of-the-year word-reading performances were positively related to word-reading outcomes at the end of the year. For teachers' OLP in the language other than the language in which the outcome was measured, estimates for OLP rated "low" were positive and significantly different from estimates for OLP rated "high" ($p < .01$). In other words, outcomes for word reading in one language (e.g., English) were higher when a teacher was rated "low" in OLP in the other language (Spanish) relative to when she/he was rated "high" in Spanish. The estimate for total quality was positive and significant, suggesting that higher outcomes were associated with higher ratings of quality. The estimate for English outcomes was significantly lower than that of Spanish ($p < .01$), indicating that students' English word-reading outcomes were lower than their Spanish word-reading outcomes. The estimate for a teaching language that was predominantly English was significantly lower than where it was Spanish ($p < .01$). In other words, students' word-reading outcomes were higher when the predominant teaching language was Spanish relative to when it was English.

Regarding the interaction of language of the outcome and the within-classroom initial-status covariate, plots suggested that the relation of this covariate to word-reading outcomes was stronger in Spanish than English. Concerning the interaction of outcome language and predominant teaching language, outcomes were highest for students of teachers who taught predominantly in Spanish, with outcomes measured in Spanish, and these outcomes were significantly higher than Spanish outcomes for students of teachers who taught predominantly in English ($p < .01$), or English outcomes in general, regardless of whether the teacher taught predominantly in English ($p < .01$) or Spanish ($p < .01$). Finally, for the interaction of outcome language with teachers' OLP in the language other than that in which the outcome was measured, outcomes were highest when they were measured in Spanish and teachers' OLP was "low" in English; such outcomes were significantly higher than all other combinations of these variables (all $p < .05$).

Discussion

We investigated whether teacher quality and related characteristics were related to classroom instructional variables and whether these and other language (of instruction and of outcome) variables were related to language and literacy outcomes for ELLs who were learning to read in Spanish or English. Results are summarized and conceptualized according to hypotheses.

Instructional Time

We found significant positive relations between teacher quality and student engagement, such that higher-quality teachers had students who were more frequently judged as being on versus off task. There was not a significant relation between teacher quality and the use of small-group instruction. There were, however, significant negative relations between teacher quality and the use of noninstructional time: teachers rated high in quality did not lose instructional time in lengthy transitions that were unrelated to reading (e.g., disciplining students, making announcements, having students line up and go to the restroom, being out of the classroom, and dispelling chaotic disruptions) but focused their energies on academic activities such as oral language development, phonemic awareness, and letter-sound instruction. Teachers with high-quality ratings, therefore, focused on instructional tasks as reported in other literature (Fisher et al., 1980; Stallings et al., 1986; Wharton-McDonald et al., 1998). These findings provide some evidence for the validity of the teacher-quality composite used, although fewer relations were noted between instructional variables and teacher knowledge.

Literacy and Language Outcomes

For literacy and language outcomes examined, teacher instructional language, language of the outcomes, oral language proficiency, and teacher quality were all related to end-of-kindergarten performance. These findings held even considering initial performance, which has long been known to influence end-of-year outcomes in native English (e.g., Coleman et al., 1966) and bilingual (e.g., August & Hakuta, 1997) populations in the early grades. This study extends such consideration not only to a given student's initial performance but also to the initial performance of the student's classroom, although we were most interested in teacher effects that may be evidenced over and above these context effects. Results are summarized below.

First, as noted, the nested nature of (particularly large) student datasets should be taken into account. For every outcome examined, we noted substantial nonindependence among students in a given classroom (with a given teacher), and, in most cases, clustering effects indicated that the relation of initial status to outcomes varied across classrooms. In many models, the initial status of both individual students and their classroom mean appeared similar to or outweighed teacher effects. Despite such findings, models that added teacher effects over and above these covariates were better fits to the data and highlight the importance of considering multiple predictors for outcomes.

Second, the language of the outcome was critical in predicting performance on outcomes, with Spanish performances higher than those in English for all outcomes in this bilingual sample (as shown in Table 2). Because the measures were in general designed to be equated or at least highly similar across languages, these results suggest that most bilingual students end their kindergarten year with stronger skills in Spanish than in English. Such a result begs the question as to whether the language used in the classroom has a moderating effect on this difference in language performance. For example, such a result could be evident in a classroom where the teacher teaches in Spanish for most of the kindergarten year (e.g., a transitional classroom). In fact, for all measures except phonological awareness, the predominant language of the teacher interacted with outcome language, such that outcomes were highest when both the outcome and the teaching language were Spanish. For letter naming and word reading, these "Spanish-Spanish" outcomes were higher than all other instructional language and outcome groupings; for oral language, the difference between Spanish and English outcomes was amplified when the predominant teaching language was Spanish relative to when it was English. In addition, for letter naming and word reading, there was a main effect of language of instruction such that outcomes in both languages were higher with predominantly Spanish instruction relative to predominantly English instruction, after considering all other factors.

Third, the results of this study also highlight the importance of bilingual students' teachers' oral language proficiency, both in the language of a given outcome as well as in the other language. For every outcome, one or both of these language-proficiency variables evidenced either a main or moderating effect (with outcome language or with initial status). In most cases, being rated (as either "high" or "low") in a given language was generally associated with higher outcomes either in that language or in the other language relative to when the teacher was not rated. This may be because oral language was not rated if a teacher was not observed to teach in that language. It may also be that teaching at least partially (if not predominantly) in two languages bodes well for stronger outcomes in bilingual students in both languages, at least in kindergarten. The only case where ratings of "low" produced higher outcomes than ratings of "high" was for word reading, although in this case the ratings were for oral language proficiency in the language other than the language of the outcome.

Fourth, other teacher variables did not produce strong relations to outcomes, in particular, our surveys of teacher knowledge (in both languages). We recognize that perhaps paper-and-pencil measures of teachers' content knowledge may not share a direct relation with how well teachers are able to scaffold and make quick instructional decisions while teaching children to read in a real classroom. We also recognize that the positive relations of teacher knowledge to student outcomes achieved in other studies may be due to the professional development that was provided across the school year (Foorman & Moats, 2004; McCutchen et al., 2002). Professional development was not provided in the current study, and doing so may have bolstered the connection between teacher measures and student outcomes.

With respect to the total quality composite, word reading was the only dependent measure where higher teacher quality was associated with higher outcomes. Although better measures of the teacher-quality construct may have provided stronger results, they would also need to be considered in the context of initial performance. Students within a given classroom vary in aptitude, academic history, and socioeconomic and demographic characteristics that are in one sense quantified, individually and collectively, in terms of students' own, and their classroom's initial performance. Students have these characteristics upon their entry into the classroom, before a qualified teacher may exert her/his effect. Therefore, the goal of a high-quality teacher may be to interact with initial status, such that for initially low-achieving students the relation of initial to final status is attenuated, and for initially high-performing students the relation of initial to final status would remain high. In this study, though, initial status was more likely to be moderated by language than by teacher quality or content knowledge.

Conclusion

Based on state literacy assessments, schools in our study were doing a good job in educating English language learners. Thus, it is not surprising that (*a*) teachers' mean scores on content knowledge and ratings of teacher quality and oral language proficiency were slightly above average, (*b*) students tended to be on task during reading/language arts instruction, and (*c*) scores on literacy measures increased across the kindergarten year. Within this context, then, we examined the degree to which teacher quality and literacy knowledge were related to instructional variables. We also investigated the extent to which a variety of teacher characteristics (quality, knowledge, oral language proficiency) and language (of instruction and of dependent measures) were related to students' language and literacy outcomes, considering initial status.

The most notable findings regarding classroom observations were the positive relations of teacher-quality ratings to student engagement and the negative relations of teacher quality to noninstructional time. The most notable findings regarding student outcomes were the importance of initial status (both within and between classrooms), language of instruction, and the language of measured outcomes in predicting student outcomes. Teacher quality was related to word-reading skills, but teacher knowledge in either language was unrelated to these student literacy outcomes.

Stronger or broader relations in predicting student outcomes over initial status may be evidenced in the years beyond kindergarten, as skills become more stabilized, although defining teacher quality as the ability to close within-classroom achievement gaps is an enormously high standard (see Borman & Kimball, 2005, p. 18). Doing so would require teachers to differentiate instruction based on knowledge of individual students' strengths and weaknesses, which for bilingual students include their initial literacy and language status in both Spanish and English. Barriers to enacting differentiated instruction include (*a*) teachers' lack of understanding of how to individualize instruction based on assessment data, (*b*) the adoption of curriculum materials oriented toward whole-class instruction and difficulty identifying when it is possible to form small groups, (*c*) uncertainty about how to manage and academically engage students who are not in the small group that is the focus of instruction by the teacher at a given time, and (*d*) lack of available professional development and mentoring to teachers that address these areas. Within bilingual settings, in particular, closing the achievement gap dictates that teachers of ELLs be able to respond to varied levels of student oral language proficiency and vocabulary and concept knowledge by scaffolding instruction to meet the needs of diverse learners. This translates into structuring classroom environments so that students are able to participate in lengthy discussions using complex syntactical and verbal exchanges in order to develop academic language (Gersten & Baker, 2000; Hickman, Pollard-Durodola, & Vaughn, 2004). However, such interchanges are likely mediated by teachers' capacity to respond to their own level of proficiency in both Spanish and English in order to promote active learning and provide meaningful feedback.

Note

This research was funded by a grant from the National Institute of Child Health and Human Development and the Institute of Educational Sciences (grant award PO1 HD 39521, Oracy/Literacy Development of Spanish-Speaking Children). Address correspondence to the first author at Department of Psychology and Texas Institute for Measurement, Evaluation, & Statistics (TIMES), University of Houston–Texas Medical Center Annex, 2151 W. Holcombe Blvd., Suite 224a, Houston, TX 77204-5053. Email: pcirino@uh.edu.

References

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

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

Ballou, D. (1996). Do public schools hire the best applicants? *Journal of Economics,* **111,** 97–134.

Borman, G. D., & Kimball, S. M. (2005). Teacher quality and educational equality: Do teachers with higher standards-based evaluation ratings close student achievement gaps? *Elementary School Journal,* **106**(1), 3–20.

Cohen, D., & Hill, H. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record,* **102,** 294–343.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., et. al. (1966). *Equality of educational opportunity* (Vol. **2**). Washington, DC: U.S. Government Printing Office.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives,* **8**(1), 50.

Darling-Hammond, L., Berry, B., & Thorenson, A. (2001). Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis,* **23,** 57–77.

Ehrenberg, R. G., & Brewer, D. J. (1995). Did teachers' verbal ability and race matter in the 1960s? Coleman revisited. *Economics of Education Review,* **14,** 1–21.

Estrada, P. (2004). Patterns of language arts instructional activity and excellence in first- and fourth-grade culturally and linguistically diverse classrooms. In H. C. Waxman, R. G. Tharp, & R. S. Hilberg (Eds.), *Observational research in U.S. classrooms: New approaches for understanding cultural and linguistic diversity* (pp. 122–143). New York: Cambridge University Press.

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation,* **28,** 465–499.

Ferguson, R. F., & Ladd, H. F. (1996). How and why money matters: An analysis of Alabama schools. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 265–298). Washington, DC: Brookings Institution.

Fisher, C., Berliner, D., Filby, N., Marliave, R., Cahen, L., & Dishaw, M. (1980). Teaching behaviors, academic learning time and student achievement: An overview. In C. Denham & A. Lieberman (Eds.), *Time to learn* (pp. 7–32). Washington, DC: National Institute of Education.

Foorman, B. R., Carlson, C. D., & Santi, K. L. (in press). Lessons observed, lessons learned from classroom reading instruction in the primary grades. In D. Haager, S. Vaughn, & J. Klingner (Eds.), *Evidence-based reading practices for response to intervention.* Baltimore, MD: Brookes.

Foorman, B. R., Goldenberg, C., Carlson, C., Saunders, W., & Pollard-Durodola, S. D. (2004). How teachers allocate time during literacy instruction in primary-grade English language learner classrooms. In M. McCardle & V. Chhabra (Eds.), *The voice of evidence: Bringing research to the classroom* (pp. 289–328). Baltimore, MD: Brookes.

Foorman, B. R., & Moats, L. C. (2004). Conditions for sustaining research-based practices in early reading instruction. *Remedial and Special Education,* **25,** 51–60.

Foorman, B. R., & Schatschneider, C. (2003). Measurement of teaching practices during reading/language arts instruction and its relationship to student achievement. In S. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: Systems for observing teaching and learning* (pp. 1–30). Baltimore, MD: Brookes.

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

Gersten, R., & Baker, S. (2000). What we know about effective instructional practices for English-language learners. *Exceptional Children,* **66,** 454–472.

Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis,* **22,** 129–145.

Graves, A. W., Gersten, R., & Haager, D. (2004). Literacy instruction in multiple-language first-grade classrooms: Linking student outcomes to observed instructional practice. *Learning Disabilities Research and Practice,* **19**(4), 262–272.

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality* (Working Paper No. 11154). Cambridge, MA: National Bureau of Economic Research.

Hickman, P., Pollard-Durodola, S., & Vaughn, S. (2004). Storybook reading: Improving vocabulary and comprehension for English language learners. *Reading Teacher,* **57,** 720–730.

Hilberg, R. S., Doherty, W., Epaloose, G., & Tharp, R. G. (2004). The standards performance continuum: A performance-based measure on the standards of effective pedagogy. In H. C. Waxman, R. G. Tharp, & R. S. Hilberg (Eds.), *Observational research in U.S. classrooms: New approaches for understanding cultural and linguistic diversity* (pp. 48–71). New York: Cambridge University Press.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS system for mixed models.* Cary, NC: SAS Institute.

McCutchen, D., Abbott, R. D., Green, L. B., Beretvas, S. N., Cox, S., Potter, N. S., Quiroga, T., & Gray, A. L. (2002). Beginning literacy: Links among teacher knowledge, teacher practice, and student learning. *Journal of Learning Disabilities,* **35,** 69–86.

Mehta, P., Foorman, B. R., Branum-Martin, L., & Taylor, W. P. (2005). Literacy as a unidimensional multilevel construct: Validation, sources of influence, and implications in a longitudinal study in grades 1–4. *Scientific Studies of Reading,* **9**(2), 85–116.

Moats, L., Foorman, B., & Taylor, W. P. (2006). How quality of writing instruction impacts high-risk fourth graders' writing. *Reading and Writing,* **19,** 363–391.

Monk, D. H., & King, J. R. (1994). Multi-level teacher resource effects on pupil performance in secondary mathematics and science: The role of teacher subject matter preparation. In R. Ehrenberg (Ed.), *Contemporary policy issues: Choices and consequences in education* (pp. 29–58). Ithaca, NY: ILR Press.

Murnane, R. J., & Phillips, B. R. (1981). Learning by doing, vintage, and selection: Three pieces of the puzzle relating teaching experience and teaching performance. *Economics of Education Review,* **1,** 691–693.

No Child Left Behind Act of 2001, PL 107-110, 115 Stat. 1425, 20 U.S.C. §§ 6301 *et seq.*

Padrón, Y. N., & Waxman, H. C. (1999). Classroom observations of the five standards of effective teaching in urban classrooms with English language learners. *Teaching and Change,* **7,** 79–100.

Rivkin, S., Hanushek, E., & Kain, J. (2005). Teachers, schools, and academic achievement. *Econometrica,* **73**(2), 417–458.

Sanders, W., & Horn, S. (1998). Research findings from the Tennessee Value-Added Assessment System database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education,* **12,** 138–151.

SAS Institute, Inc. (2003). *SAS Release 8.02* [Computer program]. Cary, NC: Author.

Saunders, W., Foorman, B., & Carlson, C. (2006). Is a separate block of time for oral English language development in programs for English learners needed? *Elementary School Journal,* **107**(2), 181–198.

Schatschneider, C., Francis, D. J., Foorman, B. R., Fletcher, J. M., & Mehta, P. (1999). The dimensionality of phonological awareness: An application of item response theory. *Journal of Educational Psychology,* **91,** 439–449.

Sebastián, N., Cuetos, F., Martí, M. A., & Carreiras, M. F. (2000). *LEXESP: Léxico informatizado del español.* Edición en CD-ROM. Barcelona: Edicions de la Universitat de Barcelona.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher,* **15**(2), 4–14.

Singer, J. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics,* **24,** 323–355.

Stallings, D., Robbins, P., Presbrey, L., & Scott, L. (1986). Effects of instruction based on the Madeline Hunter model on students' achievement: Findings from a follow-through project. *Elementary School Journal,* **86,** 571–587.

Taylor, B. M., Pearson, P. D., Clark, K., & Walpole, S., (2000). Effective schools and accomplished teachers: Lessons about primary-grade reading instruction in low-income schools. *Elementary School Journal,* **101,** 121–165.

Texas Education Agency. (1984). *Texas Teacher Appraisal System.* Austin, TX: Author.

Tuerk, P. W. (2005). Research in the high-stakes era: Achievement, resources, and No Child Left Behind. *Psychological Science,* **16**(6), 419–425.

U.S. Department of Education. (2001). *Monitoring school quality: An indicators report.* Washington, DC: U.S. Government Printing Office.

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

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

Wenglinsky, H. (2002). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives* [On-line], **10**(12). Available: http://epaa.asu.edu/epaa/v10n12

Wharton-McDonald, R., Pressley, M., & Mistretta Hampston, J. (1998). Literacy instruction in nine first-grade classrooms: Teacher characteristics and student achievement. *Elementary School Journal,* **99,** 101–128.

Wiley, D., & Yoon, B. (1995). Teacher reports on opportunity to learn: Analyses of the 1993 California Learning Assessment System. *Educational Evaluation and Policy Analysis,* **17,** 355–370.

Woodcock, R. W. (1991). *Woodcock Language Proficiency Battery—Revised, English and Spanish Forms: Examiner's Manual.* Itasca, IL: Riverside.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery—Revised.* Chicago: Riverside.

Woodcock, R. W., & Munoz-Sandoval, A. F. (1995). *Woodcock Language Proficiency Battery—Revised, Spanish Form: Supplemental Manual.* Itasca, IL: Riverside.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide.* New York: Touchstone Applied Science Associates.