Article

Hammill Institute on Disabilities

Assessment for Effective Intervention
37(1) 3–16

DOI: 10.1177/1534508411402983
http://aei.sagepub.com
SAGE

Effects of Technology Enhancements and Type of Teacher Support on Assessing Spanish-Speaking Children's Oral Reading Fluency in Second Grade

Mary J. York[1], Barbara R. Foorman[2], Kristi L. Santi[3], and David J. Francis[1]

Abstract

We examined student-, classroom-, and school-level effects in predicting second-grade Spanish-speaking children's oral reading fluency in Spanish. Teachers in 67 randomly selected urban schools administered the Tejas LEE to 1,537 first- and second-grade students. Oral reading fluency was measured in the passages students read for comprehension. Covariates were mean fluency in Grade 1, variability in fluency in Grade 1, degree of grouping in the school, and the proportion of second-grade students in the classroom and/or the school taking the Tejas LEE. Treatment effects were administration format (paper, desktop, handheld) and type of teacher support (no mentoring, web mentoring, and on-site plus web mentoring). Second-grade teachers positively affected students' reading fluency when (a) they administered the Tejas LEE on paper with the associated paper reports in classrooms of bilingual students, and (b) they either received web mentoring and had relatively homogeneous classrooms or received on-site or no mentoring and had ability-grouped classes. Implications for interpreting assessment results are discussed in the context of the type of support provided to teachers and the grouping of bilingual students by language and/or by ability.

Keywords

Spanish reading assessment, oral reading fluency, mentoring, coaching, technology enhancements

If assessment data are to effectively inform instruction, the context within which those data are collected is important to consider. For example, the *purpose* of the assessment is important: Is it for schoolwide accountability or to inform an individual teacher's instruction (e.g., Abbott, 2008)? The *availability of the data to inform instruction* is important: Are teachers administering the assessment with data immediately available to them or are specialists collecting the data for the purpose of eligibility and placement decisions (e.g., Hamilton et al., 2009)? The *support for teachers* to translate the data to instruction is important: Do teachers have a coach or mentor to help with data collection and/or interpretation or are teachers on their own in these endeavors (e.g., Wayman & Cho, 2008)? When the students being assessed are English-language learners, the context within which assessment occurs can become even more crucial—in what language is the student assessed, in what language is the student instructed, and what supports are in place to interpret results for effective instruction (Francis & Rivera, 2007)?

Our goal in the current investigation is to examine some of these contextual effects in a state-supported early reading assessment administered by teachers to Spanish-speaking students. Specifically, we asked how fluent reading in Spanish text was influenced by (a) administration format (paper, desktop, handheld), (b) types of teacher support (no mentoring, web mentoring, and on-site plus web mentoring), (c) degree of instructional grouping in the school, and (d) proportion of students in the classroom and/or school taking the Spanish reading assessment.

[1]University of Houston, Houston, TX, USA
[2]Florida State University, Tallahassee, FL, USA
[3]The Santi Group, LLC

Corresponding Author:
Barbara R. Foorman, Florida Center for Reading Research, 2010 Levy Ave., Suite 100, Tallahassee, FL 32310
Email: bfoorman@fcrr.org

Assessing Early Reading in Spanish

Learning to read in a transparent orthography like Spanish is more straightforward than learning to read in English in the sense that the "grain size" (Zigler & Goswami, 2005) of the mapping of phonology to orthography is highly predictable. In fact, evidence of the ease of untimed and timed decoding in the transparent orthographies of Finnish, Italian, Spanish, German, and Greek compared to the opaque orthography of English was found in a study of beginning reading in these languages (Seymour, Aro, & Erskine, 2003). This does not mean, however, that students learning to read transparent orthographies do not have difficulties efficiently accessing the meaning of connected text (e.g., de la Colina, Parker, Hasbrouck, & Lara-Alecio, 2001; Zigler & Goswami, 2005)

Compared to research on relations of phonological awareness to reading in English, relatively little is known about these relations in Spanish (see Denton, Hasbrouck, Weaver, & Riccio, 2000). Phonological awareness skills are related to Spanish reading (Bravo-Valdivieso, 1995; Carrillo, 1994; Manrique & Signorini, 1994; Signorini, 1997), are unidimensional at both the student and classroom level, as in English (Branum-Martin et al., 2006), and are boosted by cross-linguistic effects (Anthony et al., 2009; Branum-Martin et al., 2006; Gottardo, 2002). Thus, any assessment of learning to read in Spanish typically includes untimed or timed measures of phonological awareness, letter-sound knowledge, word reading, and connected text reading (e.g., IDEL, 2007; Tejas LEE, 2004-2006).

Because accuracy in word reading is mastered in transparent orthographies such as Spanish by the end of the first school year (Seymour, 2005), oral reading fluency can be considered a dependent variable as well as an independent variable much earlier in reading development than in English. In studies of learning to read in English, oral reading fluency, as measured by the number of words read correctly per minute (WCPM) in connected text, has proven a good index of students' development of reading skills (e.g., Fuchs, Fuchs, Hosp, & Jenkins, 2001; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003; Kim, Petscher, Schatschneider, & Foorman, 2010; Yovanoff, Duesbery, Alonzo, & Tindal, 2005; Wiley & Deno, 2005). The conceptual underpinnings of oral reading fluency emanate from Perfetti's (1985) verbal-efficiency account of reading, whereby a student who reads words efficiently has attentional resources to focus on text meaning. A student who reads slowly and with effort—a dysfluent reader—allocates attention to word identification rather than comprehension. Thus, in the current investigation with Spanish-speaking second-graders, we focused on the fluency with which students read passages for comprehension. Fluency was defined as the total number of words read correctly per minute. Fluency at the end of Grade 2 was the dependent variable, controlling for Grade 1 fluency.

Factors Affecting Early Reading Assessment

Formal reading instruction and reading assessment occur within the educational contexts of classrooms and schools. Although multilevel models that account for the clustering of students within classrooms and within schools have long been acknowledged as adding precision to educational research (Raudenbush & Bryk, 2002), less is known about the need to account for effects on early reading assessments of such schoolwide influences as (a) keeping groups of students together as they move from classroom to classroom (i.e., grouping), (b) mentoring or coaching of teachers, and (c) use of technology in assessments.

Grouping

Students are often grouped according to achievement or ability and such grouping can occur within classes or between classes and even between grades. Researchers analyzing the Early Childhood Longitudinal Study—Kindergarten Cohort (ECLS-K) found that higher-grouped students learned more compared to nongrouped students in the early primary grades (McCoach, O'Connell, & Levitt, 2006; Robinson, 2008; Tach & Farkas, 2006). In a meta-analysis of within-class grouping, Lou et al. (1996) found that lower-ability students had lower achievement gains if they were in homogeneous groupings in their classrooms rather than in mixed groupings. Similar negative findings for lower-ability students are evident in studies of between-class ability groupings in middle and high schools (Gamoran, 1992; Gamoran & Berends, 1987; Hoffer, 1992; Oakes, 2005). In contrast to these negative findings, meta-analyses by Slavin (1987) and Kulik and Kulik (1992) report positive benefits on reading outcomes with both within-class and cross-grade ability grouping. The key to successful ability grouping appears to be flexibility, combined with appropriate curricular revision or differentiation (Tieso, 2003).

Mentoring

The idea of coaching or mentoring—sustained and ongoing support for teachers by a knowledgeable educator—has been around for at least 80 years (Cassidy, Garrett, Maxfield, & Patchett, 2010) and is supported by many professional development organizations. There is a small body of research supporting a positive impact of coaching on teachers' content and pedagogical knowledge (see Sailors & Shanklin, 2010); however, there are far fewer research studies on effects of mentoring or coaching on student achievement. Although a recent study suggests that web mentoring can be effective (Pianta, Mashburn, Downer, Hamre, & Justice, 2008), the large-scale study of Florida middle school coaches showed mixed effects on student achievement

(Lockwood, McCombs, & Marsh, 2010; Marsh et al., 2008), and Garet et al.'s (2008) study of effects of professional development found significant effects on teacher knowledge but not on reading comprehension outcomes in second grade.

More recent studies, however, do show positive effects of coaching on student achievement (Bean, Draper, Hall, Vandermolen, & Zigmond, 2010; Biancarosa, Bryk, & Dexter, 2010; Matsumura, Garnier, Correnti, Junker, & Bickel, 2010; Sailors & Price, 2010). Sailors and Price found coaching a viable form of professional development in improving comprehension instruction and student achievement in Grades 2 through 8 in reading/language arts, social studies, and science. Biancarosa et al. (2010) found that the coaching component of a literacy program called Literacy Collaborative added unique variance above and beyond baseline in students' literacy learning in kindergarten through Grade 2 (effect sizes of .22, .37, and .43 in Years 1, 2, and 3, respectively). Matsumura et al. (2010) randomly assigned 32 urban elementary schools to Content-Focused Coaching (CFC) or business-as-usual literacy coaching. They found that CFC was associated with significantly higher school-level gains on the state reading test for English-language learners compared to non-CFC classrooms (*ES* = .51). Classroom observations and teacher self-report supported the increased use of CFC strategies by CFC teachers. Finally, Bean et al. (2010) examined effects of coaching with 20 Reading First teachers and found that time spent on coaching was associated with a greater number of first and second-graders scoring at proficiency and a reduced number designated as at risk. What all of these recent studies showing positive effects of coaching on student achievement have in common are strong, sustained coaching programs. Given state and national interests in effects of coaching on student achievement, we included mentoring in the current study.

Use of Technology in Reading Assessment

Although the effects of technology on gains in reading achievement have been elusive (e.g., Dynarski, 2007; NICHD, 2000), the potential value of technology-enhanced assessment to paper-based assessment is obvious: paperless test distribution and data collection; greater standardization of test administration; the possibility of enhancing and measuring student motivation with interactive question types; electronic scoring of oral and written response; availability of standardized tools for students (e.g., calculators and dictionaries); and better estimates of the extremes of the score distribution through the use of adaptive testing (e.g., Bridgeman, 2009). However, along with the promise of technology-enhanced assessments comes caution regarding the validity, utility, and credibility of the testing innovations (e.g., Baker, 2003; Russell, Goldberg, & O'Connor, 2003).

Not surprising, the research literature on the value of technology-enhanced assessment assumes that the student uses the computer, not the teacher. Such is typically not the case in early reading assessment: The teacher administers the assessment to the primary-grade child and the child manipulates letters, builds words, listens and responds to questions with or without pictures, reads stories, or reads aloud a list of words. To the extent a computer is involved, it is typically the teacher who uses it for scoring during the assessment or later to transfer data from a booklet into a data management system. Electronic capture of data facilitates data management and reporting of results at class, school, district, and even state levels. Capturing data electronically during the assessment saves time and error compared to post hoc data entry. However, real time, electronic data capture adds expense over exclusive use of paper assessments and using the electronic device may interrupt test administration if too obtrusive.

To examine the value added of technological tools to an early reading assessments in English (*Texas Primary Reading Inventory* [TPRI], 2004–2006), we investigated whether the administration format of the assessment (paper, desktop, handheld) or the type of teacher support (web mentoring or no mentoring) moderated the prediction of first- and second-grade word reading and reading fluency data in a randomized study with 210 urban and rural schools in Texas (Foorman, York, Santi, & Francis, 2008). In the case of the desktop format, teachers collected data using the paper version of the test and scores were entered via a desktop computer into a secure website so that reports could be generated. Likewise, data from the handheld had to be synced with the desktop computer to receive reports from the secure website. The moderator variables of technology format and teacher mentoring proved not to be significant. However, analyses showed that the combination of student pretest and mean of pretest classroom—a proxy for within-grade ability grouping—was a better predictor of outcome than student pretest alone. Moreover, the effect of student scores varied by the pairing of first- and second-grade teachers—a proxy for cross-grade ability grouping—in determining second-grade outcomes. In addition, intraclass correlations (ICCs) at the classroom level were much greater than at the school level, suggestive of some degree of between-class ability grouping. Also, differences in urban schools were twice that of rural schools, suggesting that ability grouping happens more frequently in urban schools.

Purpose of the Current Study

The purpose of the current study is to examine contextual effects of administration format, types of teacher support, degree of instructional grouping, and proportions of students taking Spanish assessments at the class and school level as they affect fluency in second-grade Spanish passages.

Participants were Spanish-speaking students from the urban schools included in Foorman et al.'s (2008) study. Fluency was operationalized as accuracy plus speed in reading Spanish passages.

In addition to first-grade grouping as a covariate and administration format and types of teacher support as moderator variables, we investigated the roles of (a) the proportion of second-grade students taking the Spanish assessment at the classroom and at the school level and (b) first-grade variability in predicting Grade 2 reading fluency outcomes. We included the proportion variable because a larger percentage of students taking the Tejas LEE rather than the TPRI would suggest support for primary language instruction and, therefore, better oral reading fluency in Spanish. With nesting at the classroom level, Branum-Martin, Foorman, Francis, and Mehta (2010) found Spanish maintenance bilingual programs in Texas and southern California outperformed immersion programs in Spanish. Because the decision to instruct and assess in Spanish is a school-level decision in Texas, we wanted to contrast models where students were nested at the school level as well as the classroom level. We included the variance of the fluency scores of the Grade 1 students rather than just the Grade 1 fluency mean as a way to capture the diversity of students' fluency ability. We operationalized the grouping index as the ratio of between-classroom variance and pooled within-classroom variance of Grade 1 fluency. Thus, schools with grouping would have large between-classroom variability and small pooled within-classroom variability. Because of the mixed pattern of results on mentoring described above, we decided to include type of teacher support. We hypothesized that in schools with a high degree of grouping, students' fluency outcomes would only improve with the individualized teacher support provided by web mentoring that included face-to-face, on-site mentoring. In sum, our specific research questions were as follows:

1. How do fluency and approach to grouping in first grade affect second-grade fluency outcomes at the school level? In answering this question, we considered both the mean and variance in fluency at Grade 1 as predictors of Grade 2 fluency outcomes. Similarly, how do the classroom- and school-level proportions of second-grade students taking the Spanish assessment affect second-grade fluency outcomes?
2. How do administration format (paper, desktop, handheld) and/or type of teacher support (no mentoring, web mentoring, on-site plus web mentoring) affect second-grade fluency outcomes at the school-level, and are these effects moderated by the covariates studied in the first research question?

Method

Participants

This study is part of a larger research project in which 255 schools in Texas were randomly selected to participate. Using a stratified random sampling procedure, we selected 150 schools from three large urban areas and 105 rural schools from 6 of the 20 regional education centers. The schools in the 6 randomly selected regional centers were located in nonurban cities, towns, and rural areas that ranged in population from 650,000 to less than 25,000. Because the schools were primarily located in small towns of 25,000 or less, we use the term *rural* to describe them. After randomly selecting schools, principals were invited to participate in the study on the condition that they would be randomly assigned to one of three administration formats (paper, desktop, handheld) and one of three kinds of teacher support (no mentoring, website mentoring, or on-site plus web mentoring). It is important to note that classrooms were not selected to participate; instead, schools were randomly selected and randomly assigned to study conditions. Accordingly, the demographic information for the students in the Tejas LEE participating schools included African American (26.45%), Asian (2.4%), Hispanic (60.35%), Native American (0.02%), and White (10.6%). Once schools agreed to participate and were randomly assigned to their condition, a preliminary analysis of state data was run to ensure that there were no outliers in terms of performance on state assessments. Two hundred four schools with a total of 649 first-grade classrooms agreed to participate in the 2003–2004 school year. In the following year, 2004–2005, 187 schools and a total of 523 second-grade classrooms continued their participation.

The current study includes only those schools from the research project that administered the Tejas LEE assessment during both the 2003–2004 and 2004–2005 school years. Other schools used only the English assessment, the TPRI (2004–2006). Sixty-seven urban schools and no rural schools met the study criteria. Table 1 provides means and standard deviations for the independent and dependent variables by the number of urban schools included in this study by administration format and teacher support. There are a much higher number of schools in the Handheld/Web Mentoring condition because the original study design included twice as many schools in this condition to test the effectiveness of online access to the link between students' reading errors on the TPRI and specific lessons in that school's core reading program. This instructional resource was never implemented, but the schools had already been randomly assigned to the study's conditions. The end-of-the-year Tejas LEE fluency assessment was administered to 1,537 students during both the 2003–2004 and 2004–2005 school years.

Table 1. Grade 2 Fluency School Means and Standards Deviations According to Administration Format, Teacher Support, Grade 1 Fluency Means and Variability, Grade 1 Grouping, and Proportions of Grade 2 Students Taking the Tejas LEE at the School and Classroom Levels

Condition	Grade 2 Fluency	Grade 1 Fluency	Grade 1 Fluency Variability	Grade 1 Grouping Index	Proportion TJL in School	Average Proportion TJL in Classroom	No. of Schools
Paper							
No mentoring	82.4 (9.1)	54.1 (10.4)	24.1 (7.8)	0.16 (0.20)	0.53 (0.07)	0.51 (0.10)	6
Web mentoring	80.3 (5.6)	54.0 (6.3)	19.6 (2.8)	0.44 (0.14)	0.67 (0.17)	0.65 (0.19)	8
On-site + Web mentoring	82.4 (14.9)	56.0 (10.4)	20.8 (3.9)	0.38 (0.14)	0.58 (0.13)	0.57 (0.12)	7
Desktop							
No mentoring	74.5 (7.3)	57.6 (6.9)	19.1 (3.5)	0.22 (0.20)	0.48 (0.13)	0.46 (0.21)	5
Web mentoring	82.7 (5.1)	61.0 (9.2)	23.4 (4.0)	0.30 (0.16)	0.65 (0.11)	0.63 (0.14)	6
On-site + Web mentoring	82.5 (18.6)	53.9 (15.1)	17.4 (6.8)	0.29 (0.31)	0.48 (0.18)	0.45 (0.15)	
Handheld							
No mentoring	92.6 (22.5)	51.0 (8.3)	22.2 (8.7)	0.31 (0.30)	0.40 (0.16)	0.38 (0.14)	7
Web mentoring	77.7 (12.2)	51.4 (10.4)	18.8 (6.4)	0.28 (0.27)	0.47 (0.14)	0.46 (0.15)	17
On-site + Web mentoring	81.8 (18.3)	56.7 (10.3)	18.4 (9.5)	0.28 (0.39)	0.25 (0.12)	0.31 (0.10)	4

Note. TJL = Tejas LEE. Values are *M* (*SD*) unless otherwise indicated.

Assessment Instrument

More than 92% of the schools in the state of Texas currently administer the Tejas LEE and/or TPRI in kindergarten through Grade 3 and have done so for more than 10 years. Professional development on how to administer these assessments and on how to translate assessment data to instruction is handled by school districts in Texas. Because of this fact as well as the stated goals of the overall study, the research team did not conduct training in the area of proper administration of the assessment nor did the research team check fidelity of the administration of the assessment.

Teachers administered the Tejas LEE (Tejas LEE, 2004–2006) to Spanish-speaking students in their classrooms. The Tejas LEE is a Spanish early reading diagnostic assessment written in authentic academic Spanish. It explicitly is *not* a translation of the English-language TPRI. The Tejas LEE and the English-language TPRI were developed for the purpose of providing educators with instruments that fulfill the early reading assessment requirement established by Texas Education Code Section 28.006 and the federal No Child Left Behind legislation. Linan-Thompson, Bryant, Dickson, and Kouzekanani (2005) report that on the Tejas LEE subtests reliability ranges between .78 and .91 and correlations with the Woodcock-Johnson (Woodcock & Johnson, 1989) were in the good to excellent range. The inventories in Grade 1 and Grade 2 are aligned with the Texas state curriculum standards and consist of the following components: phonological awareness (Grade 1), letter–sound identification (Grade 1), reading accuracy (Grades 1 and 2), reading fluency (Grades 1 and 2), reading comprehension (Grades 1 and 2), and spelling (Grade 2).

This study focuses on the reading fluency component, which is the ability to read grade-appropriate text with accuracy and speed. Students are placed into one of five passages in Grade 1 and one of six passages in Grade 2 based on empirical links between accuracy on a word list and accuracy in a passage. The student reads the assigned passage to completion. The teacher then asks the student to answer questions related to the reading. Although the word list is used to appropriately place students into an instructional or independent reading passage, there may be a few cases where the students are incorrectly placed into a passage. If the student reaches the frustrational level, the teacher directs the student to read another story, the story located just before the passage the student was reading. The fluency score used for analysis was words correct per minute in the passages read for comprehension. The scores were adjusted using linear equating to reflect the difficulty of the story passage, making the scores comparable across students. We used the fluency scores from these passages as our dependent variable rather than correctness of the responses to the five comprehension questions because reliability is low with such few questions.

Classroom teachers administer the Tejas LEE individually to each student in their classroom at three different times during the course of the school year: at the beginning of the year around mid-September, at the middle of the year around mid-January, and at the end of the year around mid-April. The Tejas LEE provides both screening and diagnostic

information. Like each portion of the assessment, teachers are given three levels of proficiency as a basis to assist with providing additional support to those students who need it. The three levels are developed, expected, and needs intervention.

Intervention Conditions

Teacher support. Schools were randomly assigned to one of three teacher support conditions: no mentoring, website mentoring, and on-site plus web mentoring. Teachers in schools assigned to the no mentor condition had no formal support in interpreting assessment data. Classroom teachers in schools assigned to the website mentoring condition had access to a website that provided guidance on how to group students based on assessment data. Schools in the on-site plus website mentoring condition had mentors as part of the research team who went to the school site and worked directly with classroom teachers on how to (a) set up centers, (b) interpret Tejas LEE assessment data, (c) access the website for grouping information, and (d) plan small-group lessons using Tejas LEE data. Teachers who were randomly assigned to the desktop and the handheld administration formats (discussed below) also had the option of communicating with their mentors via emails and phone calls. In addition, they had access to the same website as the website mentoring condition.

All teachers who had an on-site mentor were visited either before, during, or after school based on the needs and available times of the teachers. This time was spent analyzing the data, setting up classrooms for differentiated instruction, and providing reading-related instructional tips as needed. To enlist more teachers, our mentors created newsletters for each school that highlighted the teachers who were successfully incorporating instructional routines from the Intervention Activities Guide (part of the Tejas LEE assessment package). Teachers enjoyed the positive attention the newsletters generated, and more teachers became actively engaged in the mentoring portion of the study as a result. The mentors also randomly observed classroom instruction to determine not only if the reading instruction was data-driven but also whether or not the teachers were teaching all areas of reading as measured by the assessment. During the first year of the study, we realized (a) the need for a smaller teacher–mentor ratio and (b) the need for more bilingual mentors. In the second year, we hired more retired bilingual coaches to address both needs.

Fidelity of the mentoring conditions was ensured (a) by log data maintained by the on-site mentors, who were members of the research staff; (b) by observations made by the third author; and (c) by download information, which was tracked via login information from the website. It was apparent from the download data that the school's reading coaches, rather than classroom teachers, were primarily the ones who logged in to the website to download class summary sheets. Per the school's request and to keep in line with Reading First initiatives, school reading coaches had access to the student data.

We used an intention-to-treat (ITT) method of data analysis when analyzing the data for this study. In the ITT model, subjects are compared according to the treatment and comparison conditions to which they were randomly assigned. In medical literature, ITT provides "a pragmatic estimate of the benefit of a change in intervention policy rather than of potential benefit to patients who receive intervention exactly as planned" (Hollis & Campbell, 1999, p. 670). In large randomized trials, complex multilevel interventions are rarely implemented exactly as planned (see Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005). In the current study, the known factors were the assignment to administration condition and the access to mentoring. Although it did appear that schools may have done more, or less, than what was required by the condition to which they were assigned, the data were analyzed according to the ITT model and no claims were made as to actual implementation.

Administration format. Schools were randomly assigned to one of three administration formats: paper, desktop, or handheld. The paper condition required that the teachers use the Tejas LEE kit as they were accustomed to using it, by completing one record sheet per student and then transferring the individual scores to a class summary sheet for evaluation. The desktop condition required that the teacher administer the assessment with the Tejas LEE kit and then use a desktop computer to enter the item-level scores to a secure website. On completion of data entry, the website electronically transferred the scores to the class summary sheet for evaluation. In the handheld condition, teachers administered the Tejas LEE according to the test procedures but instead of recording scores onto the paper copy of the individual student record sheets, they tapped the screen for each correct or incorrect response during the actual administration of the assessment. Once the assessment was complete, the teachers would sync the handheld device to a computer with an Internet connection and retrieve the data by logging in to a secure website.

Analysis

The data were analyzed using ANCOVA. The outcome was the school mean fluency score at the end of Grade 2. The covariates included means and variability in end of Grade 1 fluency scores; an index of first-grade grouping; and the proportion of Tejas LEE students. The treatment factors in the model were administration format and teacher support. In addition to these treatment factors and the set of covariates, we examined the possibility that the covariates moderated the effects of the treatment factors. The grouping index was calculated for each school by taking the ratio of between-classroom variance and pooled within-classroom variance

Table 2. Correlations of Grade 2 Fluency With Covariates

	Grade 2 Fluency School Mean	Grade 1 Fluency School Mean	Grade 1 Fluency Variability	Grade 1 Grouping Index	Classroom Tejas LEE Proportion	School Tejas LEE Proportion
Grade 2 fluency school mean	1.00	0.27* (0.0299)	0.26* (0.0333)	0.20 (0.1028)	0.04 (0.7634)	–0.03 (0.8345)
Grade 1 fluency school mean		1.00	0.50* (<0.0001)	0.20 (0.0999)	0.12 (0.3233)	0.16 (0.1883)
Grade 1 fluency variability			1.00	0.22 (0.0703)	0.19 (0.1339)	0.21 (0.0953)
Grade 1 grouping index				1.00	0.34* (0.0043)	0.34* (0.0043)
Classroom Tejas LEE proportion					1.00	0.93* (<0.0001)
School Tejas LEE proportion						1.00

Note. Values in enclosed parentheses represent *p* values. An asterisk indicates that the correlation is significant at the False Discovery Rate critical value using Benjamini-Hochberg correction at $\alpha = .05$ (Benjamini & Hochberg, 1995; Benjamini & Yekutieli, 2001; What Works Clearinghouse, 2008).

of the Grade 1 fluency scores. In schools where students were grouped by ability, the between-classroom variance would be large and the pooled within-classroom variance would be small. End-of-Grade 1 fluency variability and the grouping index were transformed into square roots to conform to normality assumptions.

Two models were estimated that differed only in the way the proportion of second-grade Tejas LEE students was calculated. The first model calculated the proportion of Tejas LEE students in each classroom in each school. The school average of these classroom proportions was used in the model. The second model calculated the proportion of Tejas LEE students in each school and used that calculation in the model. The models were analyzed using SAS PROC GLM. All two-way interactions were included in the model. All analyses were intent-to-treat in that data were analyzed according to the condition to which schools were originally assigned.

If either of the treatment factors—teacher support and administration format—had a significant interaction with one of the covariates, we checked to see if there were group differences at different values of the covariate (i.e., at the mean, at one standard deviation below the mean, and at one standard deviation above the mean). If so, we wanted to know which groups differed from each other by testing pairwise mean differences.

To determine whether the data met the normality assumption, the residuals for both models were plotted. Examination of the histograms and the *q*–*q* plots showed that the residuals approximated the normal distribution. Power analyses showed enough power at .80 for $\alpha = .05$ to detect main effects and interactions for effect sizes ranging from .40 to .44, given a total sample size between 54 and 64 schools.

Results

The school means and standard deviations of the fluency outcome and covariates according to administration format and teacher support are shown in Table 1. The correlations of the covariates are shown in Table 2. We used the Benjamini-Hochberg (BH) correction (Benjamini & Hochberg, 1995; Benjamini & Yekutieli, 2001; What Works Clearinghouse, 2008) to control false discovery rate ($\alpha = .05$) in multiple testing with positive dependency. The school mean of the spring assessment of Grade 2 fluency was positively correlated with the school mean of the spring assessment of Grade 1 fluency, $r = .27$, $p = .030$ ($BH\alpha_{crit} = .042$), and variability in spring Grade 1 fluency, $r = .26$, $p = .033$ ($BH\alpha_{crit} = .050$). The school mean of the Grade 1 fluency was positively correlated with variability in spring Grade 1 fluency, $r = .50$, $p < .0001$ ($BH\alpha_{crit} = .017$). The grouping index was positively correlated with both proportion of second-grade students taking the Tejas LEE in the school, $r = .34$, $p = .004$ ($BH\alpha_{crit} = .025$), and the average proportion of second-graders taking the Tejas LEE in the classroom for each school, $r = .34$, $p = .004$ ($BH\alpha_{crit} = .033$).

Research Question Regarding Effects of Covariates on Grade 2 Fluency Outcomes

When using the average second-grade classroom proportion of Tejas LEE students in the model, the only significant covariate was the grouping index in first grade, which had a significant interaction with teacher support, $F(2, 32) = 5.60$, $p = .008$. The interaction between the grouping index and teacher support is displayed in Figure 1, with posttest least squares means plotted for the grouping index values at one standard deviation below the mean, at the mean, and at

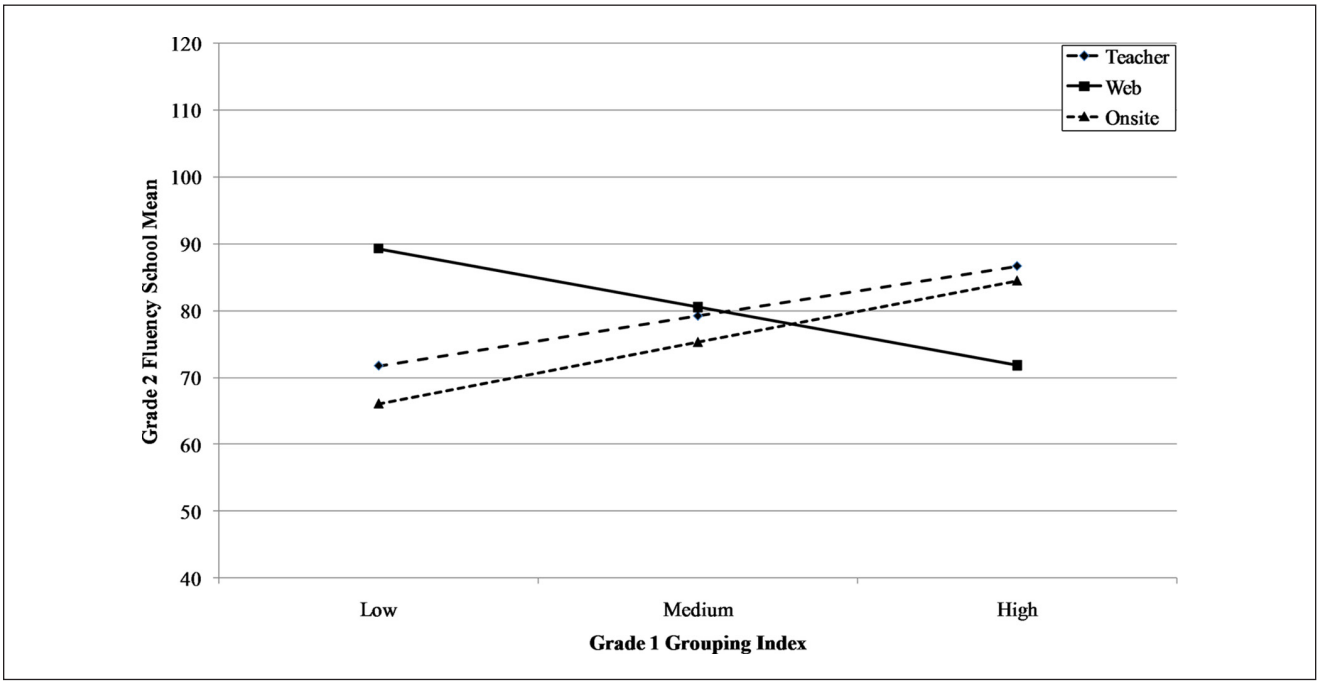

Figure 1. Effect of the interaction of teacher support with grouping index (low = one standard deviation below the mean, medium = mean, high = one standard deviation above the mean) on Grade 2 fluency rate at spring assessment using average proportion of Tejas LEE students in classroom as one of the covariates

one standard deviation above the mean. The graph shows that as grouping increased, the effectiveness of website mentoring decreased. The opposite was true for the no mentoring and on-site plus web mentoring conditions, which were more effective with increased grouping (i.e., more homogeneous classrooms).

To determine whether differences in teacher support varied significantly at low, medium, and high values of the grouping index, we examined pairwise mean differences in fluency across types of teacher support at specific values of the grouping index, namely, at one standard deviation below the mean of the grouping index and at one standard deviation above the mean of the grouping index. We did not test differences at the mean of the grouping index because the main effect test of mentoring already showed us that there were no differences at the mean. To maintain the family-wise Type I error rate at the .05 level, the critical alpha level was set at .025. There were no significant mean differences at one standard deviation below the mean or above the mean of the grouping index.

When including the school proportion of second-grade students taking the Tejas LEE as a predictor, the only significant covariate was the grouping index in first grade, which had a significant interaction with teacher support, $F(2, 32) = 5.03$, $p = .013$. The interaction is displayed in Figure 2. The trends are similar to those of the first analysis using average second-grade classroom proportion. The graph shows that as variability between classrooms increased (i.e., grouping increased), the effectiveness of website mentoring decreased. The opposite was true for the no mentoring and on-site plus web mentoring conditions, which were more effective as variability between classrooms increased.

To determine whether teacher support differences varied significantly at low, medium, and high values of the grouping index, we examined pairwise mean differences in fluency across types of teacher support at specific values of the grouping index, namely, at one standard deviation below the mean of the grouping index and at one standard deviation above the mean of the grouping index. To maintain the family-wise Type I error rate at the .05 level, the critical alpha level was set at .025. At one standard deviation below the mean of the grouping index, there were significant differences (a) between no mentoring and website mentoring, $t(32) = -2.37$, $p = .024$, with a Cohen's d effect size of –.84, and (b) between website mentoring and on-site plus web mentoring, $t(32) = 2.69$, $p = .011$, with a Cohen's d effect size of .96. However, there were no differences between the no mentoring and the on-site plus web mentoring conditions at either one standard deviation below or one standard deviation above the mean of the grouping index.

Research Question Regarding Treatment Factors and Moderation of Treatment Effects

When using the average second-grade classroom proportion of Tejas LEE students in the model, administration format was significant, $F(1,32) = 4.0$, $p = .03$. The least squares

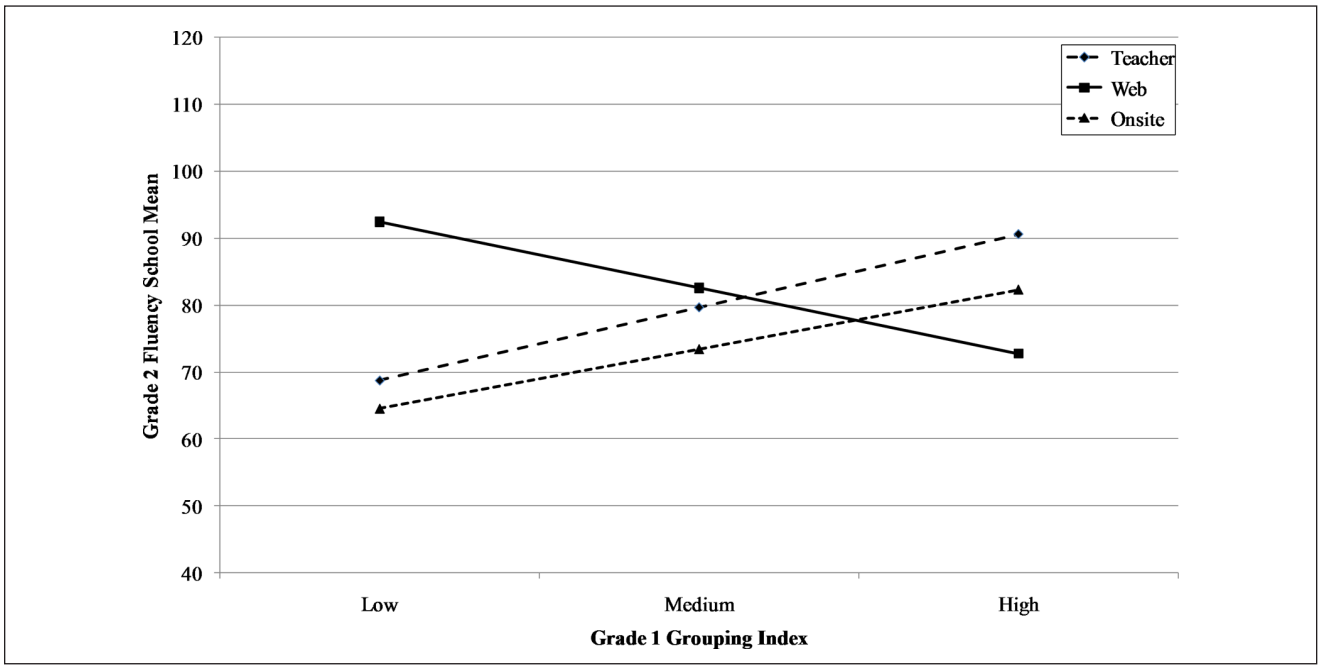

Figure 2. Effect of the interaction of teacher support with grouping index (low = one standard deviation below the mean, medium = mean, high = one standard deviation above the mean) on Grade 2 fluency rate at spring assessment using proportion of Tejas LEE students in the school as one of the covariates

means for paper, desktop, and handheld was 85.95, 71.92, and 82.91, respectively. Using Fisher's LSD, we examined all pairwise comparisons. Fisher's LSD is appropriate when there are only three groups to compare and the main effects test is significant. There was a significant difference between paper and desktop administration formats, $t(32) = 2.19$, $p = .036$, with a Cohen's d effect size of .78; however, there were no significant differences between paper and handheld or between desktop and handheld. Grade 2 fluency performance was better for schools using the paper format compared to the desktop administration format.

When using the school proportion of Tejas LEE students in the model, there were no significant main effects of the treatment factors.

Discussion

Our objective was to analyze school-level second-grade Tejas LEE fluency outcomes with treatment factors of administration format (paper, desktop, handheld) and types of teacher support (no mentoring, website mentoring, on-site plus web mentoring) and several covariates, and to examine the possible moderating role that the covariates might play in altering the effects of the treatment factors. The covariates were (a) the variability and means of fluency at the end of Grade 1; (b) grouping in Grade 1 as evidenced by the ratio of between-classroom variance and pooled within-classroom variance of fluency scores at the end of Grade 1; and (c) proportion of second-grade Tejas LEE students at the classroom or school level. We were interested in which of the covariates and which of the treatment factors significantly affected fluency outcomes in Grade 2 and whether the effects of the treatment factors were moderated by the covariates. Two general linear models were estimated, one using the proportion of Tejas LEE students at the classroom level and the other using the proportion of Tejas LEE students at the school level.

Research Question Regarding Effects of Covariates on Grade 2 Fluency Outcomes

The only covariate that significantly predicted second-grade fluency outcomes in Spanish was the first-grade grouping index, which interacted with types of teacher support. The grouping index is a ratio of between-classroom to within-classroom variability in student ability. The grouping can be described as ability grouping if variability between classrooms is high relative to variability within classrooms and increases, or remains high, with advancing grades. We found the two-way interaction of grouping with levels of teacher support was significant whether we used the *average second-grade classroom* proportion of Tejas LEE students within each school or the *average school* proportion of second-grade Tejas LEE students. The graph of the interaction showed that as between-classroom variability relative to within-classroom variability due to grouping increased, the effectiveness of website mentoring decreased but the effectiveness of no mentoring or on-site plus web

mentoring increased. Post hoc significance tests were only significant, however, when the proportion of Tejas LEE students was modeled at the school level. The website mentoring condition proved more effective than both no mentoring and on-site plus web mentoring at one standard deviation *below* the mean of the grouping index.

It is not surprising that web mentoring would be more effective in affecting students' fluency scores only when ability grouping was minimal because heterogeneous grouping of students means teachers share the challenge of low-ability students. In interpreting this finding, it is important to remember that this difference between types of teacher support and low grouping was apparent only when calculating schoolwide proportion of students taking the Tejas LEE rather than classroom-level calculation. The average proportion of second-graders taking the Tejas LEE at the school level was slightly larger than the average proportion at the classroom level (.509 vs. .496), with comparable standard deviations (0.17 vs. 0.16). The larger school-level proportion reflects the dispersion of Spanish-speaking students across second-grade classrooms rather than grouping Spanish-speaking students within particular classrooms. The practice of distributing Spanish-speaking students across classrooms creates greater demands across the school for teachers with bilingual skills because the Tejas LEE is administered in Spanish to students participating in bilingual programs. Research has shown that teachers' oral language proficiency in Spanish significantly predicts Spanish reading outcomes in the early grades (Cirino, Pollard-Durodola, Foorman, Carlson, & Francis, 2007).

In sum, the positive impact of web mentoring on Spanish fluency scores was greatest when grouping was lowest and was weakest when grouping was highest. This study adds one more positive study to the literature on mentoring (e.g., Bean et al., 2010; Biancarosa et al., 2010; Matsumura et al., 2010; Pianta et al., 2008; Sailors & Price, 2010) but only when there is little ability grouping and dispersion of second-grade Spanish-speaking students is across the school rather than grouped within particular classes. The more pervasive finding was that students' fluency scores increased under ability grouping when teachers had no mentoring or had an on-site mentor plus access to the website. This was true when fluency data were modeled with the proportion of Tejas LEE students at the class or at the school level. What was common across the no mentor and on-site mentor plus website conditions was that teachers were accessing their students' assessment data themselves rather than coaches doing so, as in the web mentoring condition. When students were grouped by ability, it appeared advantageous to have teachers accessing their own student data directly. Thus, this study adds to the mixed data on grouping in the literature (e.g., Slavin, 1987 vs. Gamoran, 1992) and supports the idea that decisions to group students by ability or by language need to be flexible and accompanied by appropriate levels of teacher support.

Research Question Regarding Treatment Factors and Moderation of Treatment Effects

When the proportion of students taking the Tejas LEE was included in the model as a classroom-level predictor, there was a main effect for administration format (e.g., paper, desktop, handheld). Follow-up contrasts revealed that Grade 2 fluency outcomes were significantly greater in the paper alone condition as compared to the desktop condition. We found a similar advantage of the paper alone (and the handheld administration) compared to the desktop condition in an analysis of English word reading outcomes in kindergarten (Foorman et al., 2008). The disadvantage of the desktop condition is the need to transfer data to the web, which delays receipt of student reports and can introduce error (e.g., Baker, 2003; Russell et al., 2003). The advantage of using a paper and pencil test format over technology is the immediacy of the information received. Teachers do not give the test knowing that they will look at the data later; they are more focused on the student responses. By analyzing the data by hand, teachers are also more aware of how the test results translate directly to instruction (Hamilton et al., 2009). Therefore, a major caution about using technology to enhance early reading assessments is to remember that the technology does not replace the professional development required to explain test administration, score interpretation, and links to instruction.

Teacher support did not affect fluency in a straightforward fashion. Rather, teacher support interacted with elements of school/classroom organization. Specifically, teacher support interacted with the grouping index, but yielded mixed results. Second-grade teachers positively affected students' reading fluency when (a) they administered the Tejas LEE on paper with the associated paper reports in classrooms of bilingual students and (b) they either received web mentoring and had relatively homogeneous classrooms or received on-site or no mentoring and had ability-grouped classes. Technological solutions are expensive and, therefore, the lack of empirical support for desktop and handheld data entry has practical significance. At the same time, the positive effect of web mentoring when ability grouping was low supports the potential of web mentoring as a professional development tool in some school settings (Pianta et al., 2008). Likewise, coaches are expensive and this study joins a body of research showing their minimal impact (Foorman et al., 2008; Marsh et al., 2008). Yet more recent research shows promising results for coaching by illustrating the intensity of coaching needed to significantly effect gains in student achievement. Biancarosa et al.'s (2010) effect sizes increased each year of their 3-year study of the Literacy Collaborative. For reading coaches or teacher mentors to be effective in bilingual settings, they must know how to interpret achievement within the context of the student's language proficiency (Francis & Rivera, 2007) and how to translate these results into instruction (Matsumura et al., 2010). The good

news in bilingual research on Spanish-speaking children is that literacy achievement in Spanish does show evidence of positive transfer to English literacy (Anthony et al., 2009; Branum-Martin et al., 2009, 2010). Educating teachers about this positive transfer would seem an important step to improving literacy instruction for English-language learners.

Limitations

The goal of this study was to examine student-, classroom-, and school-level effects on Spanish oral reading fluency outcomes in second grade. Accordingly, a randomized controlled trial with an intention-to-treat design was desirable. By randomly selecting 6 of 20 educational service centers across Texas and then randomly selecting 67 schools within those 6 centers, data that we could not obtain related to student demographics and teacher characteristics were controlled. We did use public data to report on school race/ethnicity percentages and to verify that there were no outliers in schools' state achievement data. As described earlier, the ITT design means that we analyzed data according to the condition (i.e., administration format or type of teacher support) to which the schools were randomly assigned. Administration format was easy to verify because we received data on paper or electronically from our contractor. However, fidelity to the kinds of teacher support was more difficult to verify. We knew who received on-site mentoring and how many times our mentor visited the school, but there were no independent raters to judge the quality of the mentoring. Although we do know how often the mentors visited each teacher, we know that because of the ratio of mentor to teachers, the time spent mentoring each teacher was limited. It would be beneficial to look at how many hours of mentoring are needed to effect change. Also, in the web mentoring condition, the contractor did not produce data to document how many times the web was accessed and which pages were accessed. Even with those data we would not have known who accessed the website—the second-grade teacher or the coach. In future studies, we can reduce the number of schools and prioritize the collection of such fidelity data.

Conclusion

In modeling effects of administration format and type of teacher support on second-grade Spanish fluency outcomes, we are reminded of the importance of school organizational factors such as grouping of students. When the average classroom proportion of students taking the Tejas LEE was included as a predictor in the model, Grade 2 fluency outcomes were significantly higher when teachers administered the assessment on paper with the immediate paper score reports rather than administering on paper and waiting for electronic score reports. We found this same advantage of paper administration in our TPRI study (Foorman et al., 2008). In addition, when the school-level proportion of students taking the Spanish reading assessment was included as a predictor in the model, the grouping index interacted with teacher support to positively influence fluency outcomes when teachers either (a) received web mentoring and had relatively homogeneous classrooms or (b) received on-site or no mentoring and had ability-grouped classes.

In sum, in explaining how Spanish-speaking students perform on early reading assessments in Spanish, it is important to take into account (a) the type of support provided to teachers to help them translate data to instruction and (b) school organizational factors. Effects of technology were evident only in interactions with these factors. In this study, the school organizational factors relevant to explaining fluency outcomes were (a) the grouping of Spanish-speaking students within particular classrooms or their dispersion across second-grade classrooms and (b) the degree to which students were grouped by ability. We have learned from this study that in either case—whether students are grouped by language or by ability—there are associated tradeoffs that require teacher support.

The implications of these results for educational practice are complex. First, technology enhancement is not a panacea for early reading assessments, especially if it delays receipt of data reports. When teachers administer early reading assessments on paper to individual students, results are immediately available to them. Second, the effectiveness of the type of support teachers receive for early reading assessments may vary with school organizational factors, such as whether students are grouped homogeneously or heterogeneously. In this study, homogeneous grouping was associated with increased fluency outcomes in second grade for two out of three types of mentoring—no mentoring and on-site mentoring. Website mentoring was effective only in the context of heterogeneous grouping. Further research is clearly warranted on the conditions under which mentoring or coaching can help teachers translate assessment results to instruction, especially in bilingual settings.

Declaration of Conflicting Interests

The authors declared no conflicts of interests with respect to the authorship and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research and/or authorship of this article:

This work was supported by grant R305W020001, Scaling up Assessment-Driven Intervention Using the Internet and Handheld Computers, from the Institute of Education Sciences in the U.S. Department of Education.

References

Abbott, D. V. (2008). A functionality framework for educational organizations: Achieving accountability at scale. In E. Mandinach & M. Honey (Eds.), *Data driven school improvement: Linking data and learning* (pp. 257–276). New York, NY: Teachers College Press.

Anthony, J. L., Solari, E. J., Williams, J. M., Schoger, K. D., Zhang, Z., Branum-Martin, L., & Francis, D. J. (2009). Development of bilingual phonological awareness in Spanish-speaking English language learners: The roles of vocabulary, letter knowledge, and prior phonological awareness. *Scientific Studies of Reading, 13*, 535–564.

Baker, E. L. (2003). Reflections on technology-enhanced assessment. *Assessment in Education, 10*, 421–424.

Bean, R. M., Draper, J. A., Hall, V., Vandermolen, J., & Zigmond, N. (2010). Coaches and coaching in Reading First schools: A reality check. *Elementary School Journal, 111*, 87–114.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological), 57*, 289–300.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics, 29*, 1165–1188.

Biancarosa, G., Bryk, A., & Dexter, E. (2010). Assessing the value-added effects of Literacy Collaborative professional development on student learning. *Elementary School Journal, 111*, 7–34.

Branum-Martin, L., Foorman, B. R., Francis, D. J., & Mehta, P. (2010). Contextual effects of bilingual programs on beginning reading. *Journal of Educational Psychology, 102*, 341–355.

Branum-Martin, L., Mehta, P. D., Fletcher, J. M., Carlson, C. D., Ortiz, A., Carlo, M., & Francis, D. (2006). Bilingual phonological awareness: Multilevel construct validation among Spanish-speaking Kindergartens in transitional bilingual education classrooms. *Journal of Educational Psychology, 98*, 170–181.

Branum-Martin, L., Mehta, P. D., Francis, D. J., Foorman, B. R., Cirino, P. T., Miller J. F., & Inglesias, A. (2009). Pictures and words: Spanish and English vocabulary in classrooms. *Journal of Educational Psychology, 101*, 897–911.

Bravo-Valdivieso, L. (1995). A four year follow-up study of low socioeconomic status, Latin American children with reading difficulties. *International Journal of Disability, Development and Education, 42*, 189–202.

Bridgeman, B. (2009). Experiences from large-scale computer-based testing in the USA. In F. Scheuermann & J. Björnsson (Eds.), *The transition to computer-based assessment*: *JRC Scientific and Technical Reports* (pp. 39–44). Luxembourg: European Communities.

Carrillo, M. (1994). Development of phonological awareness and reading acquisition. A study in Spanish language. *Reading and Writing, 6*, 279–298.

Cassidy, J., Garrett, S. D., Maxfield, P., & Patchett, C. (2010). Literacy coaching: Yesterday, today and tomorrow. In J. Cassidy, S.D., Garrett, & M. Sailors (Eds.), *Literacy coaching: Research and practice* (pp. 15–27). Corpus Christi, TX: Center for Educational Development, Evaluation, and Research.

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

de la Colina, M. G., Parker, R., Hasbrouck, J., & Lara-Alecio, R. (2001). Intensive intervention in reading fluency for at-risk Spanish readers. *Bilingual Research Journal, 25*, 503–538.

Denton, C., Hasbrouck, J., Weaver, L., & Riccio, C. (2000). What do we know about phonological awareness in Spanish? *Reading Psychology, 21*, 335–352.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., . . . Sussex, W. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort*. Washington, DC: United States Department of Education, Institute of Education Sciences.

Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI#231).

Foorman, B., York, M., Santi, K., & Francis, D. (2008). Contextual effects on predicting risk for reading difficulties in first and second grade. *Reading and Writing, 21*, 371–394.

Francis, D. J., & Rivera, M. O. (2007). Principles underlying English language proficiency tests and academic accountability for ELLs. In J. Abedi (Ed.), *English language proficiency assessment in the nation: Current status and future practice* (pp. 13–32). Davis: University of California.

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

Gamoran, A. (1992). Syntheses of research: Is ability grouping equitable? *Educational Leadership, 50*(2), 11–17.

Gamoran, A., & Berends, M. (1987). The effects of stratification in secondary schools: Synthesis of survey and ethnographic research. *Review of Educational Research, 57*, 415–435.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., . . . Sztejnberg, L. (2008). *The impact of two professional development interventions on early reading instruction and achievement* (NCEE Report 2008-4031). Washington, DC: National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences.

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

Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). *Using student achievement data to support instructional decision making* (NCEE 2009-4067). Washington, DC: National Center for Educational Evaluation & Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Hoffer, T. B. (1992). Middle school ability grouping and student achievement in science and mathematics. *Educational Evaluation and Policy Analysis, 14*, 205–227.

Hollis, S., & Campbell, F. (1999). What is meant by intention to treat analysis: Surveys of published randomized controlled studies. *British Medical Journal, 319*, 670–674.

IDEL (Indicadores Dinámicos del Ėxito en la Lectura). (2007). Eugene: University of Oregon.

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95*, 719–729.

Kim, Y., Petscher, Y., Schatschneider, C., & Foorman, B. (in press). Does growth rate in oral reading fluency matter in predicting reading? *Journal of Educational Psychology*.

Kulik, J. A., & Kulik, C.-L. C. (1992). Ability grouping and gifted students. In N. Colangelo and G. A. Davis (Eds.), *Handbook of gifted education* (pp. 178–197). Boston, MA: Allyn & Bacon.

Linan-Thompson, S., Bryant, D. P., Dickson, S. V., & Kouzekanani, K. (2005). Spanish literacy instruction for at-risk kindergarten students. *Remedial and Special Education, 26*, 236–244.

Lockwood, J. R., McCombs, S., & Marsh, J. (2010). Linking reading coaches and student achievement: Evidence from Florida middle schools. *Educational Evaluation and Policy Analysis, 32*, 372–388.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apolonia, S. (1996). Within class grouping: A meta-analysis. *Review of Educational Research, 66*, 423–458.

Manrique, A. M. B., & Signorini, A. (1994). Phonological awareness, spelling and reading abilities in Spanish-speaking children. *British Journal of Educational Psychology, 64*, 429–439.

Marsh, J. A., McCombs, J. S., Lockwood, J. R., Martorell, F., Gershwin, D., Naftel, S., . . . Crego, A. (2008). *Supporting literacy across the sunshine state: A study of Florida middle school reading coaches*. Santa Monica, CA: RAND.

Matsumura, L. C., Garnier, J. E., Correnti, R., Junker, B., & Bickel, D. D. (2010). Investigating the effectiveness of a comprehensive literacy coaching program in schools with high teacher mobility. *Elementary School Journal, 111*, 35–62.

McCoach, D. B., O'Connell, A. A., & Levitt, H. (2006). Ability grouping across kindergarten using an early childhood longitudinal study. *Journal of Educational Research, 99*, 339–346.

National Institute of Child Health and Human Development (2000). *National Reading Panel–Teaching children to read: Reports of the subgroups* (NIH Pub. No. 00-4754). Washington, D.C.: U.S. Department of Health and Human Services.

Oakes, J. (2005). *Keeping track: How schools structure inequality* (2nd ed.). New Haven, CT: Yale University Press.

Perfetti, C. A. (1985). *Reading ability.* New York, NY: Oxford University Press.

Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K., & Justice, L. (2008). Effects of web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 23*, 431–451.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.) Thousand Oaks, CA: Sage.

Robinson, J. P. (2008). Evidence of a differential effect of ability grouping on the reading achievement growth of language-minority Hispanics. *Educational Evaluation and Policy Analysis, 30*, 141–180.

Russell, M., Goldberg, A., & O'Connor, K. (2003). Computer-based testing and validity: A look back into the future. *Assessment in Education, 10*, 279–293.

Sailors, M., & Price, L. R. (2010). Professional development that supports the teaching of cognitive reading strategy instruction. *Elementary School Journal, 110*, 301–322.

Sailors, M., & Shanklin, N. L. (2010). Introduction: Growing evidence to support coaching in literacy and mathematics. *Elementary School Journal, 111*, 1–6.

Seymour, P. H. K. (2005). Early reading development in European orthographies. In M. J. Snowling & C. Hulme (Eds.), *The science of reading* (pp. 296–315). Malden, MA: Blackwell.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*, 143–174.

Signorini, A. (1997). Word reading in Spanish: A comparison between skilled and less skilled beginning readers. *Applied Psycholinguistics, 18*, 319–344.

Slavin, R. E. (1987). Ability grouping and student achievement in elementary schools: A best evidence synthesis. *Review of Educational Research, 57*, 293–336.

Tach, L., & Farkas, G. (2006). Learning-related behaviors, cognitive skills, and ability grouping when schooling begins. *Social Science Research, 35*, 1048–1079.

Tejas LEE. (2004–2006). Austin, TX: University of Texas System and Texas Education Agency.

Texas Primary Reading Inventory. (2004–2006). Austin, TX: University of Texas System and Texas Education Agency.

Tieso, C. L. (2003). Ability grouping isn't tracking anymore. *Roeper Review, 26*, 29–37.

Wayman, J. C., & Cho, V. (2008). Preparing educators to effectively use student data systems. In T. J. Kowalski & T. J. Lasley (Eds.), *Handbook on data-based decision-making in education* (pp. 89–104). New York, NY: Routledge.

What Works Clearinghouse. (2008). *What Works Clearinghouse Procedures and Standards Handbook (version 2.0)*. Retrieved from http://ies.ed.gov/ncee/wwc/references/idocviewer/Doc.aspx?docId=19&tocId=10

Wiley, H. I., & Deno, S. L. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education, 26*, 207–214.

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

Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice, 24*, 4–12.

Zigler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 3–29.

About the Authors

Mary J. York, PhD, is a research assistant professor with the Texas Institute for Measurement, Evaluation, and Statistics at the University of Houston. Her current interests include multilevel, growth curve, and structural equations modeling.

Barbara R. Foorman, PhD, is the Francis Eppes Professor of Education and Director of the Florida Center for Reading Research at Florida State University. She is coeditor of the *Journal for Research on Educational Effectiveness* and is internationally known for her work in reading and language development and reading assessment.

Kristi L. Santi, PhD, is the CEO of The Santi Group. Her current interests include conducting research in the area of English-language learners, special education, and teacher professional development.

David J. Francis, PhD, is the Hugh Roy and Lillie Cranz Cullen Distinguished Professor and Chairman of the Department of Psychology and Director of the Texas Institute for Measurement, Evaluation, and Statistics at the University of Houston. He is nationally recognized for his research on early reading and language development in at-risk populations, especially students at risk for reading disabilities and students who are English-language learners.