THE TIMING OF EARLY READING ASSESSMENT IN KINDERGARTEN

Kristi L. Santi, Mary York, Barbara R. Foorman, and David J. Francis

Abstract. **Under the accountability provisions of the No Child Left Behind legislation, screening for reading risk has become routine in kindergarten. The objective of this study was to examine the effects of the timing of kindergarten assessment and the type of support provided to teachers to translate assessment results to instruction. Sixty-two schools with 201 kindergarten classrooms and 3,635 students in a southwestern state were randomly assigned to administer kindergarten assessment in the fall or in the winter, with teachers receiving onsite or web mentoring. A small, significant effect (d = 0.13) was found for outcomes on a standardized reading test administered at the end of kindergarten when teachers administered the screen in the fall and received web rather than onsite mentoring. A slight, nonsignificant, reduction in reading risk (i.e., reduction in false positives) was apparent. Given these small effects, there is little empirical support for initiating screening in the fall rather than in the winter of kindergarten.**

KRISTI L. SANTI, Ph.D., The Santi Group, LLC, Houston, TX.
MARY YORK, Ph.D., Texas Institute for Measurement, Evaluation, and Statistics, University of Houston.
BARBARA R. FOORMAN, Ph.D., Center for Reading Research, The Florida State University.
DAVID J. FRANCIS, Ph.D., Texas Institute for Measurement, Evaluation, and Statistics, University of Houston.

In the United States there is a long history of screening children in kindergarten for sensory, language, and cognitive abilities in order to refer children with disabilities for early treatment (Fletcher & Satz, 1982). Since the mid-1990s, however, kindergarten has increasingly become full day with a formal curriculum. As the kindergarten curriculum has shifted from one of "developmentally appropriate practice" to one of academic subjects (Deboer, 2002; Jeynes, 2006; Saracho, 1986), assessment has shifted from screening for developmental disabilities to screening for academic risk and diagnosis of learning difficulties.

A major catalyst for this shift in the business of kindergarten has been a shift in federal policy towards increasing accountability. In 2001 the No Child Left Behind (NCLB) act passed, with the goal "to ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging State academic achievement standards and State academic assessments" (U.S. Code 6302 § 1001). Coupled with the accountability provisions of NCLB is the primary-grade component of the legislation, called Reading First, which mandates the use of curriculum and assessments based on scientifically based reading research.

As states wrote assessment plans for screening and diagnosis into their Reading First applications, little consideration was given to the timing of kindergarten assessment: whether screening and diagnosis should start at the beginning of the year or after children had

had a chance to socialize to the academic demands of kindergarten. This study examined this question of the timing of kindergarten assessment by randomly assigning schools to administer a state's early reading assessment at the beginning of the kindergarten year or in the middle.

At Risk in Kindergarten

Students who are at risk for reading difficulties are typically identified due to their lack of response to instruction. The lack of response is attributed to several factors, including background, lack of differentiated instruction, lack of progress in a given reading program, and/or poor performance on reading measures. Historically, schools have often classified underachieving kindergarten students as not "developmentally ready," anticipating that they would prosper in first grade.

However, with the accountability systems in place (i.e., NCLB) and the plethora of research on beginning readers, we know what milestones students are able to reach at each grade level (see National Reading Panel [NRP], 2000; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004; Snow, Burns, & Griffin, 1998). With this knowledge, assessments were developed to help teachers identify students who may be at risk for reading difficulties based on predictive indicators. Students who do not perform as expected on the assessments can be classified as at-risk readers with 90% confidence, and teachers can alter their instruction to help ensure that students are learning to read and not falling through the cracks (Foorman, Fletcher, & Francis, 2004).

Screening in Kindergarten

The benchmark assessment in this study – The Texas Primary Reading Inventory (TPRI; Foorman et al., 2004) – consists of two components, a screen and an inventory. The screen is used to determine risk status, whereas the inventory serves as a diagnostic and provides data in the areas of phonemic awareness, phonics, oral reading, and comprehension to inform teachers of how to target reading instruction. For the purposes of this article, the screen will mainly be discussed.

The two kindergarten screens (one at the middle of the year and the other at the end of the year) were developed from a longitudinal database that involved a large sample of children in kindergarten through grade 2. The kindergarten screens, designed to be administered in the middle and end of kindergarten, were developed to minimize the errors that result from not identifying children who need further assistance. Thus, the screens provide brief assessments (3-5 minutes) that permit identification of children who are not likely at risk for developing reading difficulties.

Timing of Risk Status

Research on the timing of kindergarten assessment is sparse. Several studies have investigated the effects of child-centered issues such as behavior, relationships, play, school readiness, and/or psychological adjustments to kindergarten and students' ability to work within a kindergarten setting. The research in this area uses standardized measures of ability, behavior ratings, and other checklists to determine students' school readiness and then compares readiness with students' academic abilities at the end of kindergarten or grade 1. A hallmark of this research includes the debate over whether this approach to determining kindergarten readiness (e.g., the Froebel model) is a better predictor of school readiness and academic achievement than standardized tests (Jeynes, 2006).

A considerable amount of reading research on risk prediction focuses on phonological skills. Schatschneider et al. (2004) reviewed studies examining which prereading skills are most predictive of future reading success. They also extended the literature by employing the statistical analysis most appropriate for making such predictions (i.e., dominance analysis). The Schatschneider study contributes to the substantial body of evidence indicating that phonological skills are among the best predictors of future reading abilities (see also NRP, 2000; Scanlon & Vellutino, 1996; Scarborough, 1998; Snow et al., 1998).

We also have a wealth of knowledge from the literature in special education and reading disabilities suggesting that early detection helps deter later reading problems. For example, Francis, Shaywitz, Stuebing, Shaywitz, and Fletcher (1996) found that the reading achievement of reading disabled and low-achieving students followed a deficit model of development as opposed to a development lag model. In their longitudinal study, these authors found no difference between the reading growth of students who were identified as low achieving versus the students identified as having a learning disability. Furthermore, the two groups of poor readers did not catch up to their peers from ages 7 through 17. Moreover, Torgesen et al. (2001) found that, with intensive intervention, students ages 8 to 10 could increase their accuracy in reading but could not catch up in their fluency rates.

However, early intervention research indicates that the vast majority of children can learn to read successfully (Foorman & Al Otaiba, in press; Torgesen, 2002). Research has also shown that students who are poor readers at the end of grade 1 almost never reach an average reading level by the end of their elementary career (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Juel 1988; Torgesen & Burgess, 1998).

Considering that results from prevention studies show that 70-90% of the at-risk children (students in the bottom 20%) in kindergarten through grade 2 can

learn to read in the average range, our next step is to determine how early the identification process can start (Foorman, Breier, & Fletcher, 2003; Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; Torgesen, 2000).

Given the consensus among researchers that reading-related skills in kindergarten predict more complex reading skills in future grades, the next logical question is how early students should be tested to determine risk status. In this particular area, little research has been conducted. Critical factors such as the timing of assessment and the type of support needed to train teachers in how to analyze and use data when administering a screening and diagnostic assessment are the two focus areas of this paper. Two types of support were provided: web mentoring or onsite mentoring by master teachers. Onsite mentoring or coaching, while not widely studied in a large, scale-up context, has been found successful when working to improve teachers' knowledge base, thus increasing early student literacy (see Neuman & Cunningham, 2008, for review). Taking advantage of new technologies and widely available Internet access in classroom settings, we also provided a web mentoring category in an effort to efficiently reach more teachers.

Research Question

Screening students at the beginning of kindergarten may lead to improved reading performance at the end of the school year. In early years of implementing Reading First, some schools deviated from traditionally agreed-upon practice and began administering the TPRI (Foorman et al., 2004) at the beginning of kindergarten rather than at the middle of the year. Data from these schools indicated that students whose teachers used the TPRI at the beginning of the school year performed at a higher level on the end-of-year reading measure than students whose teachers screened at the middle of the year. Based on these anecdotal data and the widespread use of early reading assessments in kindergarten, we decided to conduct a randomized study to examine when to assess kindergarten students and how best to support kindergarten teachers in their use of early reading assessments. Therefore, this randomized study examined the effect of the timing of administering screening assessments in kindergarten and of different forms of teacher support on kindergarten students' success on end-of-year reading assessments.

METHOD

Participants

A total of 62 schools, 201 teachers, and 3,635 kindergarten students participated in this yearlong study. The schools were from both rural (population between 300 and state median) and urban (population = 735,000+) settings across a large southern state. Schools were randomly assigned to one of two times for screening (beginning of the year and middle of the year) and one of two types of teacher support (onsite mentoring or web mentoring).

At the start of the study, school data were analyzed to determine if school status would interfere with the outcomes. Schools were comparable based on state assessment outcomes, teacher experience, and school socioeconomic (SES) levels. However, the statistical analyses included only students administered both the

Table 1

Size of Sample (Students, Classrooms, and Schools) Administered TPRI Screening at End of Year by Study Condition

		Onsite Mentor			Website Mentor		
Area	Timing	Students	Classrooms	Schools	Students	Classrooms	Schools
Rural	Fall	655	37	11	680	37	11
	Winter	537	35	13	607	33	10
Urban	Fall	361	18	4	281	14	4
	Winter	226	11	4	288	16	5

Table 2

Size of Sample(Students, Classrooms, and Schools) Administered GRADE Word Reading and TPRI Screening at End of Year by Study Condition

		Onsite Mentor			Website Mentor		
Area	Timing	Students	Classrooms	Schools	Students	Classrooms	Schools
Rural	Fall	620	36	11	530	30	8
	Winter	515	34	12	551	31	9
Urban	Fall	73	4	1	207	11	3
	Winter	107	5	2	164	10	4

TPRI screening and the Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001) word reading during the spring administration. One school was excluded from the analysis due to small sample size. It was a small rural school with one kindergarten classroom that reported TPRI screening and GRADE word reading scores for fewer than five students.

Table 1 displays the number of students, classrooms, and schools in each condition included in the study. Unfortunately, one district discontinued participation following random assignment. However, teachers in this district still administered the TPRI, which was being used throughout this district prior to the study. The district agreed to provide the TPRI data to the researchers, but did not allow the students to be tested on the other study assessment. Consequently, 868 of the students from 12 schools and 40 classrooms were not administered the end-of-year standardized word reading assessment in the spring, and thus their data were not used in the final analysis.

Table 2 displays the number of students, classrooms, and schools in each condition that were administered the standardized word reading assessment (GRADE + TPRI) in the spring.

Instruments

The Texas Primary Reading Inventory (TPRI; Foorman, Fletcher, & Francis, 2004) was the screening and diagnostic assessment used for this study. Overall, all indices of reliability were found to be in the excellent and good range; screen = alpha .92, diagnostic = .89; screen test-retest .87, diagnostic .60; and screen generalizability .76, diagnostic =.70. As reported in the TPRI Technical Report (Foorman et al., 1997):

> Each of the five screens developed for the TPRI fails to identify less than 10% of the children who end up not reading at expected levels by the end of Grades 1 and 2 (false negative errors). Even when the goal is to identify children who are not at risk, errors involving the over-identification of these children are inevitable (false positive errors). However, errors of this type are viewed as less serious than failing to identify children who are at risk. Moreover, false positive rates were uniformly below 45% for Kindergarten and Grade 1. (p. 8)

The TPRI is administered in a one-to-one setting to each student by the classroom reading/language arts teacher. It is typically given twice a year in kindergarten, in mid-January and mid-April. The screening section determines risk status, and measures letter names/letter sounds and phonemic awareness. The inventory section provides data that inform classroom instruction and measure phonemic awareness, phonics, and listening comprehension.

This instrument is currently administered by teachers in all the schools in the participating state and, therefore, was very familiar to teachers and administrators. It is important to note that teachers were not specifically trained on the TPRI for this study as the teachers attend trainings conducted by their districts or regional support centers on how to administer and use assessment results to inform instruction.

Procedure

The schools were randomly assigned to either administer the assessment at the beginning of the school year

(i.e., October) or at the middle of the year (i.e., January). Schools were then randomly assigned to one of two types of support, onsite or website mentoring. All schools gave the standardized reading assessment Group Reading Assessment and Diagnostic Evaluation (GRADE; Williams, 2001) at the beginning and the end of the school year. The teachers were not given the results of the GRADE during the course of the school year. While teachers could request the data, no teachers made this request.

Administration

Schools assigned to assess TPRI in the beginning of the year. The schools assigned to the October administration assessed their kindergarten students three times a year (October, January, and April) on the TPRI, as prescribed by the administration guidelines. Teachers administered the screening section to all students in the classroom, followed by starting each student at Inventory Task 1 and following the branching rules. The branching rules stipulate that students stay in a domain as long as they receive a passing score (i.e., a score indicating that the student is *developed* on the skill). If a student does not receive a developed score, the teacher stops that portion of the inventory and moves to the next.

At the middle of the year, the teacher started administration of the assessment with the screening section and continued with the first task for which a student showed a score of *"still developing."* This procedure was repeated again at the end of the year.

Schools assigned to assess TPRI in the middle of the year. The schools assigned to the January administration assessed their kindergarten students twice a year (January and April), as prescribed by the administration guidelines. Teachers administered the screening section to all students in the classroom and then administered the inventory to each student according to the branching rules. At the end of the year, the teacher started administration of the assessment with the screening section and continued with the first task for which a student showed a *"still developing"* score from the middle-of-year administration. All schools administered the end-of-year assessment according to the assessment guidelines.

Teacher Support

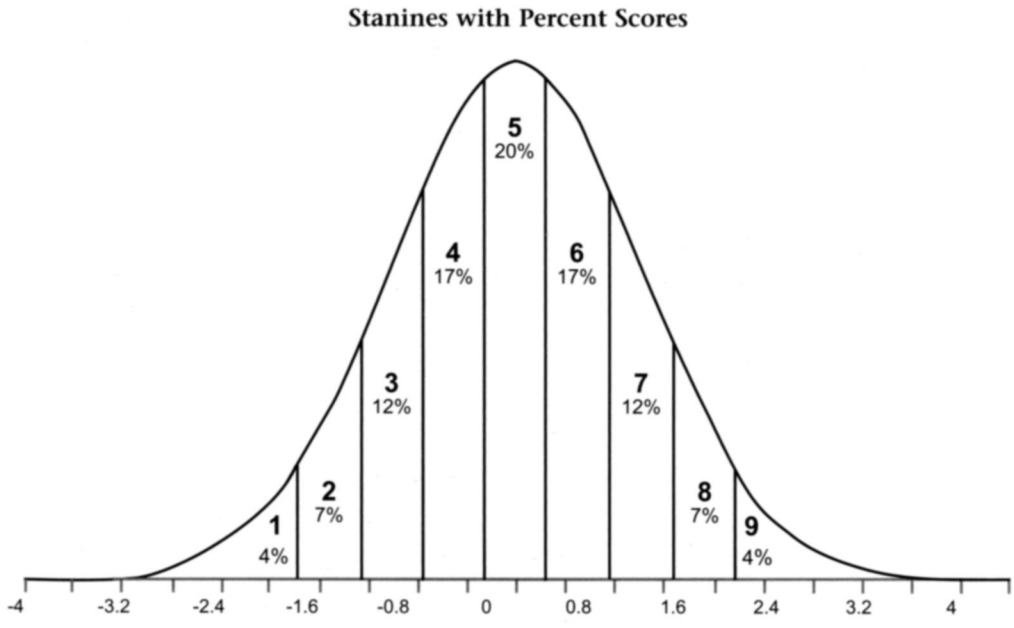
Onsite +website support. One level of support, onsite mentoring, assigned a master teacher (referred to as the mentor) to work with the classroom teachers at their school during a designated time that was convenient for the teacher. The mentors assisted the teachers in setting up centers, interpreting TPRI assessment data, and planning small-group lessons using the TPRI data (for a more detailed description, see Santi, York, Foorman, & Francis, in press). Additionally, teachers could communicate with the mentors via electronic mail and/or phone calls at any time during the course of the school year. The teachers also had access to the mCLASS:TPRI website. The mCLASS:TPRI website allowed teachers access to classroom summary sheets in color, skill profiles, an interactive tool to group students and link activities to the small-group assignment, and a teacher's lounge for exchanging ideas, questions, and comments.

Website-only support. The second method of support, website mentoring, provided assistance to teachers through the use of technology. Teachers accessed the mCLASS:TPRI website to review individual and class summary sheets, use the skill map, and utilize the grouping tool to plan for small-group, differentiated instruction. Teachers also communicated with each

Table 3
Study Design Timeline

Timing First Assessment	Measure	Timeline Fall	Winter	Spring
Fall	TPRI	X	X	X
Winter	TPRI		X	X
Fall & Winter	GRADE	X		X

Figure 1. Illustration of percentage allocation under each stanine.



other via the teacher's lounge to ask questions and share ideas. Teachers in this condition did not receive individual mentoring from the research staff.

Design and Analysis

Study design. The primary focus of this study was to determine if the timing of the first administration of the screening portion of the assessment affected student outcomes at the end of the year. We also examined whether teacher support and the area (urban or rural) in which the schools were located moderated that effect. All analyses were intent-to-treat, in that schools' data were analyzed according to the condition to which they were originally assigned.

The study used a 2x2x2 factorial design with two levels of assessment timing (initial TPRI assessment in October or January, which will be referred to as the beginning and middle of the year, respectively), two levels of area type (urban or rural), and two levels of teacher support (onsite or web mentoring). The data were nested at three levels: student, classroom, and school. The first set of analyses examined the effect of study conditions on GRADE word reading scores assessed at the end of the year. The second set of analyses examined the effect of study conditions on the proportion of students that were at risk at the end of the year using the GRADE word reading and the TPRI screening assessments.

Predictors. The predictors were the timing of the first administration of the TPRI, area type, teacher support, and all two-way interactions.

Outcomes. The outcome for the first set of analyses was the GRADE word reading subtest administered at the end of the year during the 2005-2006 school year. The GRADE word reading raw score was the number of words out of 10 correctly selected on the word reading subtask. Based on scores from the spring assessment, the GRADE word reading scale has a Cronbach's alpha of 0.75. We assumed a Poisson distribution (a discrete probability distribution that expresses the probability of a number of events occurring in a fixed period of time if these events occur with a known average rate and independently of the time since the last event) for the GRADE word reading scores because they have discrete values from 1 to 10; furthermore, plotting of the data shows that it follows a Poisson distribution.

The outcomes for the second set of analyses were the school-level proportions of students identified as at risk

at the end of the year using the GRADE word reading and the TPRI screening assessments. For GRADE word reading, at-risk students were those who scored less than 5 on the word reading subtask, which corresponds to a stanine score of 3 or below, which is equivalent to a percentile score of 23 or below (see Figure 1). For the TPRI screening, at-risk students were those who scored less than 8 on the letter sound screen (Cronbach alpha .88) and less than 6 on the blending onset-rimes and phonemes screen (Cronbach alpha .88). The TPRI cutoffs are based on the publisher's recommendations.

Statistical analysis to predict GRADE word reading. The SAS® version 9.1 PROC GLIMMIX procedure using a Poisson distribution was used to determine the effects of study conditions on the GRADE word reading. The predictors (or fixed effects) included three variables: the timing of the first TPRI assessment (beginning or middle of year), the school area type (urban or rural), and the type of teacher support received (onsite or website mentoring), and their respective two-way interactions. In specifying the random effects, the classroom means were allowed to vary within area type. The random effects for classrooms were calculated within area type because preliminary analyses showed a different covariance structure for urban schools and rural schools. Furthermore, preliminary analysis resulted in nonsignificant school-level random effects. As a result, they were not included in the model.

Statistical analysis to predict proportion of at-risk students. The SAS® version 9.1 PROC NLMIXED procedure was used to determine the effects of study conditions on the proportion of students in each school considered at risk at the end of the year. Separate analyses were performed using the GRADE word reading and the TPRI screen to determine at-risk students. The TPRI screen analysis included all 3,635 students who were administered the TPRI during the spring administration. The GRADE word reading analysis included the 2,767 students who were administered the GRADE word reading during the spring administration. An additional analysis using PROC NLMIXED was performed to ensure that no systematic bias existed in the TPRI analysis sample between those who were administered the GRADE and those who were not.

The predictors (or fixed effects) in these analyses included three variables, the timing of first TPRI assessment (beginning or middle of year), the school area type (urban or rural), and the type of teacher support received (onsite or website mentoring), and their two-way interactions, just as in the analysis of GRADE scores. However, in specifying the random effects, only the school means were allowed to vary; that is, random effects at the classroom level were ignored. The interclass correlation was calculated by taking the variance estimate for schools and dividing it by the sum of the variance estimate and the residual estimate.

RESULTS

GRADE Word Reading as Outcome

On average, students at the end of kindergarten identified 7.76 words correctly out of 10 words ($SD = 2.2$) on the GRADE Word Reading subtest. Based on the PROC GLIMMIX analysis, the classroom-level random effects were 0.004 for rural ($SE = 0.001$) and 0.013 ($SE = 0.006$) for urban schools. For the fixed effects, the only significant effect was the interaction between the type of teacher support and the timing of the first TPRI assessment, $F(1,2606) = 7.68$, $p = 0.006$.

To understand the interaction of teacher support and timing of assessment, additional analyses were conducted to determine which simple effects were significant. Simple effects examine the effect of one variable within the levels of another. Four simple effects analyses were conducted to examine (a) whether there was a difference in student outcomes between the students first tested on TPRI at the beginning of the year and those first tested on TPRI in the middle of the year when the teachers were given onsite teacher support, (b) whether there was a difference in student outcomes between the students who were first tested on the TPRI at the beginning of the year and those first tested on TPRI in the middle of the year when the teachers received support through the TPRI website, (c) whether there was a difference in student outcomes between classrooms that received onsite teacher support and those that received support through the TPRI website when the students were first assessed with the TPRI starting at the beginning of the year, and (d) whether there was a difference in student outcomes between teachers that received onsite teacher support and those that received support through the TPRI website when the students were first assessed with the TPRI at the middle of the year. The only significant simple effect was for the timing of assessment within the website mentoring condition of the year, $F(1,2606) = 3.59$, $p = 0.012$.

The mean for GRADE word reading for students first tested on the TPRI at the beginning of the year was 8.13 ($SD = 2.02$), whereas the mean score for students first assessed at the middle of the year was 7.58 ($SD = 2.24$). This difference produces a mean difference of .55 or less than one word. The standardized Cohen's d effect size is 0.13.

Proportion of At-Risk Students as Outcome

The loss of subjects from the analysis of the GRADE due to the withdrawal of one district from the study necessitated that we assess the extent to which the

departure of these schools might bias analyses of the GRADE compared to the analyses of the TPRI, which could still be based on the entire population of schools. As a result, we examined the performance on the TPRI of students who were administered the GRADE and those who were not. Results revealed no significant difference in school performance on the TPRI on the basis of students' status as present or missing from the spring GRADE, $t(61) = 0.48$, $p = 0.64$, with an estimate of 0.05 ($SE = 0.10$). Based on the outcome of this comparison, we treat the data for the GRADE as missing at random.

An examination of the cross-tabulation of the GRADE Word Reading and TPRI screening measures looking at at-risk vs. not-at-risk students shows that the two assessments sometimes differ in identifying risk (see Table 4). Specifically, 194 students were identified by the GRADE as being at risk but not by the TPRI; conversely, 58 students were identified by the TPRI as being at risk but not by the GRADE. There is a 0.40 correlation between the two measures.

At risk as measured by GRADE word reading. Overall, the GRADE word reading subtask classified 10.3% of the students as at risk at the end of the year. The percentage of students identified as at risk at the end of the year was 9.5% for students who started TPRI assessment during fall and 11.1% for student who started TPRI assessment during winter (see Table 4). The results of the PROC NLMIXED analysis showed a significant random effect at the school level, $t(49) = 4.43$, $p<.0001$. However, none of the fixed effects had significant main effects or interactions. That is, timing of assessment, area type, and type of teacher support did not affect the percentage of students identified as at-risk students using the GRADE word reading subtask in the spring.

At risk as measured by TPRI screening at the end of the year. The percentage of students identified as at risk at the end of the year by TPRI screening was 5.3% for students who started TPRI assessment during fall and 5.4% for students who started TPRI assessment during winter (see Table 5). The results of the PROC NLMIXED analysis showed a significant random effect at the school level, $t(61) = 4.94$, $p<.0001$. However, none of the fixed effects had significant main effects or interactions. That is, timing of assessment, area type, and type of teacher support did not affect the percentage of students identified as at risk at the end of the year using the TPRI screen.

DISCUSSION

The purpose of the present study was to examine the effect of the timing of kindergarten screening on student outcomes as well as to assess the effects of different approaches to supporting teachers in working with assessment data. In terms of reducing reading problems at the end of the year, the timing of the first assessment seemed to have little effect. Although slightly different results were found when comparing schools on (a) the mean performance on the GRADE word reading and (b)

Table 4

Comparison of Reading Risk at End of Year – TPRI vs. GRADE Word Reading

	End of Year GRADE (frequency)					
	Fall Administration		Winter Administration		Total	
End-of-Year TPRI	At Risk	Not At Risk	At Risk	Not At Risk	At Risk	Not At Risk
At Risk	42	34	48	24	90	58
Not At Risk	94	1,260	100	1,165	194	2,425
End-of-Year TPRI	**End of Year GRADE (percent)**					
At Risk	2.94%	2.38%	3.59%	1.8%	3.25%	2.10%
Not At Risk	6.57%	88.11%	7.48%	87.14%	7.01%	87.64%

Table 5
Reading Risk at End of Year for TPRI and GRADE Word Reading – Fall vs. Winter Start

	GRADE	TPRI (GRADE nonmissing)	TPRI (all)
Fall	136/1,430 (9.5%)	76/1,430 (5.3%)	111/1,977 (5.61%)
Winter	148/1,337 (11.1%)	72/1,337 (5.4%)	99/1,658 (5.97%)
TOTAL	284/2,767 (10.3%)	148/2,767 (5.35%)	210/3,635 (5.78%)

the percentage of at-risk students, for the most part, varying the timing of the first TPRI assessment had little effect on students' reading performance or on the percentage of students identified as at risk.

While the effects of timing were not strong, there was evidence of a small, but statistically significant interaction between timing and type of teacher support. This interaction was found to result from the positive effects of earlier testing for teachers receiving web-based mentoring. Specifically, for teachers receiving web-based mentoring, students performed slightly better on the GRADE word reading at the end of the year if the students had been tested for the first time on the TPRI in the fall ($MN = 8.13$, $SD = 2.022$) compared to in the winter ($MN = 7.58$, $SD = 2.24$). This interaction between timing of initial assessment and type of teacher support was unexpected.

Although not statistically significant, students first tested on the TPRI in the fall were less likely to be identified as at risk on the GRADE at the end of the year (9.5%) than were students first tested on the TPRI in the winter (11.1%). The same was true when risk was defined based on the end-of-the-year TPRI scores (5.3% vs. 5.4%). In so far as any reduction in the risk of reading problems at the end of kindergarten is potentially valuable, this question of possible risk reduction due to the timing of assessment in kindergarten merits investigation in future research. Whether larger effects might result when timing of assessment is manipulated in schools that are less familiar with assessing to inform instruction is an open question.

Implications for Practice

Educators and policy makers have expressed widespread interest in developing and implementing mentoring programs. For example, the American Association of Teacher Education Survey of teacher educators, school teachers, and university and school administrators found that mentoring was identified as the most critical issue confronting educators (Buttery, Haberman, & Houston, 1990). NCLB encourages schools to mentor teachers on how to use assessment to inform instruction with the end goal of increasing student achievement. Teacher certification programs also encourage schools to mentor new teachers about using assessment data to improve student outcomes. Teacher mentoring programs have been implemented widely and discussed extensively (Bey & Holmes, 1990, 1992; McIntyre, Hagger, & Wilkin, 1993). However, as found in this study, simply placing a face-to-face mentoring program in the schools does not guarantee improved student outcomes.

Currently, the timing of kindergarten screening is not informed by empirical evidence. The present study showed no obvious benefit of testing at the beginning of the school year, except when teachers were also provided access to website mentoring to assist them in translating the information obtained from assessment into instructional practice. It is possible that the benefits of earlier assessment are negated when teachers do not have immediate access to support on data interpretation and the use of data to guide instruction. In the onsite mentoring condition, the need to coordinate between the availability of the mentor and the timing of the assessment may limit the effectiveness of placing assessments at the beginning of the kindergarten year. Some states delay kindergarten screening to give students time to acclimate to school and to avoid overidentification of students with limited prior school experience. This study attempted to assess whether the choice of when to screen students for reading difficulties had consequences for students' reading outcomes.

Decisions about students' future status based on early

assessments are subject to two types of error: false-positive errors (i.e., deciding students are at risk, when they are not) and false-negative errors (i.e., deciding students are not at risk, when, in fact, they are). With ubiquitous early intervention, it is impossible to estimate the false-positive error rate, as false-positive errors are indistinguishable from positive intervention results. In contrast, false-negative errors remain important and can be accurately estimated. Thus, we focus more on the failures of the screening system to identify reading problems than on false-positive error rates, due to the focus on early intervention in these schools.

Under NCLB, many states are choosing to screen students at the beginning of kindergarten. For example, the state of Florida has legislated kindergarten screening within the first month of school in order not to miss 5-year-olds who may be at risk of reading problems. Our data show a small, nonsignificant reduction in the number of children identified with potential reading problems in the spring when screening occurs in the fall rather than in the winter of kindergarten. Given the relatively small percentage of students identified as at risk in this study in either condition, the timing of kindergarten assessment appears to have little bearing on student outcomes, and thus is a choice based more on convenience to the teachers, students, and schools than on any concern for student outcomes.

However, one cannot ignore the context in which this study was conducted, and the limits that context may place on the generalizability of the study's findings. The small number of at-risk students in this study indicates that kindergarten instruction in this broad sample of Texas schools is meeting the instructional needs of most children in the schools. Texas has had a reading initiative in the primary grades since 1997, along with state-mandated early assessment of reading development for screening and progress monitoring. If this study were replicated in another state without a 10-year history of assessment-driven instruction, the percentages of at-risk kindergartners might be larger, and the effect of moving screening assessments from winter to fall might be different. At the same time, it must be added that the present study showed no deleterious effects of beginning assessment in October. Thus, schools that prefer to begin assessment in the fall, consistent with what is likely taking place in grades 1 and 2, will find some support for this practice in the outcomes of the present study. Of course, these results must also be considered in light of the context in which the study was conducted and might be different in a state where fewer children attend preschool, or where there is less emphasis on reading instruction in kindergarten.

Limitations of the Study

This study has limitations. First, although the study employed random assignment at the school level, sample size in terms of the number of participating schools was relatively small and somewhat unevenly distributed across urban and rural areas. The rural schools represented 45 of the 62 schools in the study. The student count was large, with 3,635 total kindergarten students and 201 kindergarten teachers.

Second, the study looked at student outcome based on the assessment most commonly used by teachers in the state where the study was conducted and added a standardized, whole-group-administered test at the end of the year and, therefore, should be replicated with different screening measures and individually administered, standardized outcome measures. The GRADE word reading outcome is somewhat limited as an assessment of word reading, but has the advantage of being group administered. In particular, the GRADE potentially has a floor effect at the beginning of kindergarten. These issues also lend the low correlation found between the GRADE and the TPRI.

A case could be made for taking a somewhat broader perspective on assessing student outcomes and monitoring actual instructional interventions provided by teachers as a result of student assessments at the initial time point. Obviously, if teachers do not act on student assessment results, the effects of altering the timing of student assessments bear on the predictive validity of the assessment, and not on the efficacy of tying assessment and instruction together. This point warrants further investigation, as this study did not answer the question of fidelity of teachers' use of assessment data to inform instruction.

REFERENCES

Buttery, T. J., Haberman, M., & Houston, W. R. (1990). First annual ATE survey of critical issues in teacher education. *Action in Teacher Education, 12*(2), 1-7.

Bey, T. M., & Holmes, C. T. (Eds.). (1990). *Mentoring: Developing successful new teachers*. Reston, VA: Association of Teacher Educators.

Bey, T. M., & Holmes, C. T. (Eds.). (1992). *Mentoring: Contemporary principles and issues*. Reston, VA: Association of Teacher Educators.

Deboer, G. E. (2002). Student-centered teaching in a standards-based world. *Science and Education, 11*, 405-417.

Fletcher, J. M., & Satz, P. (1982). Kindergarten prediction of reading achievement: A seven-year longitudinal follow-up. *Educational and Psychological Measurement, 42*(2), 681-685.

Foorman, B. R., & Al Otaiba, S. (in press). Reading remediation: State of the art. In K. Pugh & M. McCardle (Eds.), *Helping children learn how to read: Outstanding issues and new directions in diagnosis and treatment of reading failure*. San Antonio, TX: Pro-Ed.

Foorman, B. R., Breier, J. I., & Fletcher, J. M. (2003). Interventions aimed at improving reading success: An evidence-based approach. *Developmental Neuropsychology, 24*(2 & 3), 613-639.

Foorman, B. R., Fletcher, J. M., Francis, D. J., Carlson, C. D., Chen, D., Mouzaki, A., Schatschneider, C., Wristers, K., & Taylor, R. (1997). *TPRI technical report.* Retrieved August 19, 2009, from http://tpri.org/Researcher_Information.

Foorman, B. R., Fletcher, J. M., & Francis, D. J. (2004). *Texas Primary Reading Inventory.* New York: McGraw-Hill.

Foorman, B. R., Fletcher, J. M., & Francis, D. J. (2004). Early reading assessment. In W. M. Evers & H. J. Walberg (Eds.), *Testing student learning, evaluating teaching effectiveness* (pp. 81-125). Stanford, CA: The Hoover Institution.

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

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88*(1), 3-17.

Jeynes, W. H. (2006). Standardized tests and Froebel's original kindergarten model. *Teachers College Record, 108*(10), 1937-1959.

Juel, C. (1988). Learning to read and write: A longitudinal study of fifty-four children from first through fourth grades. *Journal of Educational Psychology, 80*(4), 437-447.

McIntyre, D., Hagger, H., & Wilkin, M. (Eds.). (1993). *Mentoring: perspectives on school-based teacher education.* London: Kogan Page.

National Reading Panel. (2000). *Report of the National Reading Panel: Teaching students to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Bethesda, MD: National Institute of Child Health and Human Development.

Neuman, S. B., & Cunningham, L. (2008). The impact of professional development and coaching on early language and literacy instructional practices. *American Educational Research Journal, 46*(2), 532-566.

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

Santi, K. L., York, M., Foorman, B. R., & Francis, D. J. (in press). Mentoring: A framework for success. *Insight.*

Saracho, O. N. (1986). Play and young children's learning. In B. Spodek (Ed.), *Today's kindergarten* (pp. 91-109). New York: Teachers College Press.

Scanlon, D. M., & Vellutino, F. R. (1996). Prerequisite skills, early instruction, and success in first-grade reading: Selected results from a longitudinal study. *Mental Retardation & Developmental Disabilities Research Reviews, 2*(1), 54-63.

Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In B. K. Shapiro, A. J. Capute, & B. Shapiro (Eds.), *Specific reading disability: A view of the spectrum* (pp. 77-121). Hillsdale, NJ: Erlbaum.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96*(2), 265-282.

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

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research and Practice, 15,* 55-64.

Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology, 40,* 7-26.

Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K, Conway, T., & Rose, E. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities, 34,* 33-58.

Torgesen, J. K., & Burgess, S. R. (1998). Consistency of reading-related phonological processes throughout early childhood: Evidence from longitudinal-correlational and instructional studies. In J. Metsala & L. Ehri (Eds.), *Word recognition in beginning reading* (pp. 161-188). Hillsdale, NJ: Lawrence Erlbaum Associates.

Williams, K. T. (2001). *Group reading assessment and diagnostic evaluation.* Circle Pines, MN: AGS.

AUTHOR NOTES

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. We thank our partners at Wireless Generation, and gratefully acknowledge the help of the participating teachers and principals.

Please address correspondence about this article to: Kristi L. Santi, The Santi Group, LLC, P.O. Box 20766, Houston, TX 77225; e-mail: Kristi@thesantigroup.com