# Contextual effects on predicting risk for reading difficulties in first and second grade

**Barbara R. Foorman · Mary York · Kristi L. Santi · David Francis**

**Abstract** This study utilized early reading assessment data from a randomized trial of 210 urban and rural schools in Texas to examine contextual effects on risk prediction in first and second grade. The primary objective was to examine the roles of (a) individual differences, (b) the grade 1 classroom, and (c) the pairing of first and second grade teachers in determining grade 2 outcomes in word reading and fluency. A second objective was to investigate whether the administration format of the assessment (paper, paper plus desktop, handheld plus desktop) or the level of teacher support (web mentoring, no mentoring) moderated the prediction. These moderator variables proved not to be significant. Subsequent analyses found that a combination of student pretest and mean of pretest classroom was a better predictor than student pretest alone. Additionally, the effect of student scores varied by teacher-pair. On average, intraclass correlations (ICCs) ranged from 6% to 17%. Differences in ICCs at the classroom level were much greater than at the school level, and differences in urban schools were twice that of rural schools.

**Keywords** Early reading · Prevention of reading difficulties · Identification of reading problems · Risk prevention in reading · Prevention of reading disabilities · Early reading assessment

B. R. Foorman (✉)
Florida Center for Reading Research, Florida State University, 227 N. Bronough St., Suite 7250, Tallahassee, FL 32301, USA
e-mail: BFoorman@fcrr.org

M. York · D. Francis
Texas Institute for Measurement, Evaluation, and Statistics, University of Houston, Houston, USA

K. L. Santi
The Santi Group, LLC, Houston, USA

## Introduction

Since the 1990's research on predicting risk of reading difficulties in young children has made major advances. These advances are due to increasing evidence regarding valid and reliable predictors of risk and attention to children's response to intervention as a way to differentiate those children with and without reading disabilities. Throughout this research, the focus has been on predicting risk for *individual* children. To the extent that contextual factors have been considered, they are typically demographic variables such as socio-economic status (often measured by household income and/or mother's educational status) or race/ethnicity, or teacher value-added characteristics (e.g., certification, licensure, years of experience). To our knowledge no research has examined prediction of risk for reading difficulties within the context of classroom and school level variables. In other words, does a child's membership in a particular classroom with a particular teacher and a particular peer group constitute a risk factor that may interact with class membership the following year? To address this question we examined first and second grade data from a randomized field trial of 210 schools in urban and non-urban Texas.

Predicting reading difficulties in young children

The National Research Council's (1998) *Preventing Reading Difficulties in Young Children* reviewed evidence of valid and reliable predictors of reading risk: oral language, including phonological awareness; print awareness; letter name and sound knowledge; rapid naming of letters; and word reading. There appears to be a developmental timetable to these predictors of reading success at the end of first grade: print concepts and letter names predict best in early kindergarten; phonological awareness and letter sound knowledge predict in kindergarten and early first grade; rapid naming of letters predicts at the end of kindergarten and early first grade; word recognition predicts well from first grade and beyond (e.g., Fletcher et al., 2002; Molfese et al., 2006; Scarborough, 2001; Torgesen, 2002). Item response work on the segmenting and blending tasks within the domain of phonological awareness reveals its own developmental sequence (Anthony, Lonigan, Driscoll, Phillips, & Burgess, 2003; Schatschneider, Francis, Foorman, Fletcher, & Mehta, 1999).

   Researchers have built on the research on valid and reliable predictors of reading difficulties to develop early reading assessments. Examples of widely used, evidence-based, early reading assessments are the Observation Survey (Clay, 1993; Denton, Ciancio, & Fletcher, 2006), the Qualitative Reading Inventory (QRI; Leslie & Caldwell, 2005), Phonological Awareness and Literacy Screening (PALS; Invernizzi, Meier, Swank, & Juel, 2002), the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002), and the Texas Primary Reading Inventory (TPRI; Foorman, Fletcher, & Francis, 2004). The Observation Survey and QRI are examples of informal reading inventories where risk prediction is a matter of local judgment rather than statistical comparison. Predictive validity for PALS is based on the Stanford-9 Achievement Test (Harcourt, 1996).

For the TPRI, predictive validity is based on a cut score on the Woodcock Johnson Broad Reading (WJ; Woodcock, McGrew, & Mather, 2001) of six months behind grade level (i.e., approximately the 35th percentile). Through item response analyses, the phonological awareness, letter sound, and word reading items on the TPRI screen predict to this cut score on the WJ Broad Reading so that teachers can quickly tell which students are highly likely to be on or above grade level at the end of the year. Screening instruments tend to set rates of over-identification high in order to offset the danger of under-identifying students definitely at risk. By having the screen detect students *not* at risk, the TPRI avoids this shortcoming (Foorman & Ciancio, 2005). For students performing below the cut score on the TPRI screen, the teacher can administer the full inventory, which takes about 20–30 min, and yields a developmental profile of strengths and weaknesses in phonological awareness, graphophonemic knowledge, fluency, and reading comprehension so that instructional objectives can be set for each student.

The Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002) focuses on the predictive utility of patterns of subskill performance rather than the predictive validity of the performance in relation to cut-scores on well-known, standardized tests, as the PALS and TPRI do. In DIBELS, benchmark scores on the Initial Sound Fluency (ISF) in the middle of kindergarten, Phoneme Segmentation Fluency (PSF) at the end of kindergarten, Nonsense Word Fluency (NWF) in the middle of first grade, and Oral Reading Fluency (ORF) at the end of first, second, and third grades have been determined by examining receiver operator characteristic (ROC) curves which identify trade-offs in sensitivity and specificity for various cut-off scores. These benchmark scores have been used to designate levels of risk: (1) *low risk* means that 80% or more of students with that pattern of performance are likely to achieve benchmark goals; (2) *at risk* means that 20% or fewer students with that pattern of performance are likely to achieve subsequent goals; (3) *some risk* means that 50% of students are likely to achieve benchmark goals and that, therefore, no clear prediction of risk is possible. For students categorized as at risk, substantial intervention is recommended (see Good, Simmons, Kame'enui, Kaminski, & Wallin, 2002, for more information).

In sum, early reading assessments are widely in use across the United States to (a) identify students at risk for reading difficulties, (b) make instructional recommendation, and (c) monitor learning and outcomes. The potential for using this system of assessment-driven instruction to prevent reading disabilities from occurring and to provide early intervention to students at risk of a reading disability has been documented (e.g., Foorman, 2003; McCardle & Chhabra, 2004; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001; Wanzek & Vaughn, 2006) and is beginning to be realized under the response-to-intervention framework discussed next.

Response to intervention as a way to refine risk predictions

Response-to-intervention (RtI) is a new process for evaluating students' eligibility for special education services and serves as an alternative to identifying students on the basis of a discrepancy between IQ and achievement. This alternative was

explicitly mentioned in the 2004 reauthorization of the Individual with Disabilities Education Improvement Act (IDEA, 2004). The legislation stated that local educational agencies ''shall not be required to take into consideration whether a child has a severe discrepancy between achievement and intellectual ability'' and ''may use a process that determines if the child responds to scientific, research-based intervention as a part of the evaluation procedures'' (IDEA; Pub. L. No. 108–446 § 614 [b] [6] [A]; § 614 [b] [2 & 3]).

Within special education, there are those who insist that RtI concerns *only* identification and classification and that the role of special education be restricted to post-eligibility determination. These researchers encourage the continuation of identification and classification based on discrepancies between achievement and ability, with ability defined by scores on multiple cognitive processing tests rather than a single IQ test (Hale, Naglieri, Kaufman, & Kavale, 2004). However, many other researchers see RtI as a prevention model that stimulates multiple tiers of intervention and reduces the possibility that students referred to special education are academic casualties from inadequate or inappropriate instruction (Fletcher, Coulter, Reschly, & Vaughn, 2004). The prevention model is assumed by the Reading First component of the No Child Left Behind act (NCLB; 2001) and is encouraged under IDEA 2004 by the permission granted districts to use up to 15% of their special education funds for early intervention. Many prevention researchers have developed standard treatment protocols that have been implemented with fidelity (e.g., Al Otaiba & Fuchs, 2006; Mathes et al., 2005; McMaster, Fuchs, Fuchs, & Compton, 2005; Vaughn, Linan-Thompson, & Hickman, 2003; Vellutino et al., 1996; Vellutino, Scanlon, Small, & Fanuele, 2006). A synthesis of recent research on intensive, early interventions reveals higher effects for smaller group size and for an earlier start of intervention (e.g., kindergarten—grade 1; Wanzek & Vaughn, 2006).

In these RtI studies, students identified as at risk for reading difficulties were provided additional reading practice through small-group instruction. Students who remained at risk (i.e., non-responders) were provided additional, more intensive intervention on a pull-out basis. Non-responders after this stage were referred to special education. School districts implementing the RtI framework will need to work out the inclusionary and exclusionary criteria for each layer of intervention. Employing a hybrid system of documenting low achievement in spite of quality instruction coupled with RtI models is recommended (Fletcher, Francis, Morris, & Lyon, 2005).

Contextualizing risk

As schools move to using early reading assessment data to determine risk and guide instruction, it is important to have effective models of risk and prevention. These models do not have to consider only individual risk factors; they can consider *individual* risk nested within *classroom* and within *school* levels of risk simultaneously in a *multilevel* model of risk and prevention (Foorman & Ciancio, 2005). Classroom and school levels of risk are indirectly addressed through value-added models that address NCLB's call for a highly qualified teacher in every classroom.

Student achievement gains are used to measure teacher quality in these models. Value-added studies tend to show substantial variation in the quality of instruction and that most of that variation occurs within rather than between schools (Hanushek, Kain, O'Brien, & Rivkin, 2005). In many studies, teacher quality appears unrelated to advanced degrees or certification (Ballou, 1996; Desimone, Smith, & Frisvold, in press; Hanushek et al., 2005; but also see Darling-Hammond, Berry, & Thorenson, 2001; Goldhaber & Brewer, 2000). Years of experience does relate to achievement gains in some studies but only in the first year of teaching (Hanushek et al., 2005). High quality teachers (i.e., those with high achievement gains) tend to be effective with students of all ability levels (Hanushek et al., 2005); however this relationship appears to be moderated by poverty. That is, disadvantaged students tend to be assigned to weaker teachers in both low-poverty and high-poverty schools, but because low-poverty schools have more high quality teachers, disadvantaged students do benefit from being in wealthier schools (Desimone et al., in press).

In bringing together the individual risk prevention models with the teacher value-added models, we are interested in individual and classroom differences and their interaction. Our specific context for this study was longitudinal, early reading assessment data from a randomized trial of 210 urban and rural schools in Texas in grades 1 and 2. The early reading assessment instrument was administered in three randomly assigned formats (paper, paper plus desktop, or handheld plus desktop) and according to two randomly assigned levels of teacher support (web mentoring or no mentoring). Our specific research questions were:

1. Which predicts student word reading and fluency outcomes in grade 2 better—student pretest in grade 1 or a combination of student pretest and mean of the pretest (grade 1) classroom?
2. Does administration format (paper, paper plus desktop, handheld plus desktop) and/or level of teacher support (web mentoring, no mentoring) moderate the prediction?
3. What is the role of teacher-pairings from grade 1 to grade 2? Does the effect of student score vary by teacher-pair? Is the slope the same or does it vary randomly at the teacher-pair level?

## Method

### Participants

Two hundred ten schools in Texas were randomly selected to participate in the study. Using a stratified random sampling procedure, we selected 105 schools from 3 large urban areas and 105 rural schools from 6 of the 20 regional centers (see Table 1). The schools in the six randomly selected regional centers were located in non-urban 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

**Table 1** Number of proposed and actual schools in the study design according to administration format, teacher support, and area type (urban vs. rural)

| Teacher support | Paper | | Paper + Desktop | | Handheld + Desktop | | Total |
|---|---|---|---|---|---|---|---|
| | Urban | Rural | Urban | Rural | Urban | Rural | |
| Administration format of early reading assessment (TPRI) | | | | | | | |
| No mentoring | | | | | | | |
| Proposed | 15 | 15 | 15 | 15 | 15 | 15 | 90 |
| Year 1 | 11 | 12 | 12 | 9 | 8 | 15 | 67 |
| Year 2 | 10 | 11 | 10 | 7 | 10 | 15 | 63 |
| Web mentoring | | | | | | | |
| Proposed | 15 | 15 | 15 | 15 | 30 | 30 | 120 |
| Year 1 | 12 | 15 | 11 | 15 | 25 | 26 | 104 |
| Year 2 | 12 | 10 | 9 | 10 | 25 | 27 | 93 |
| Total schools | | | | | | | |
| Proposed | 30 | 30 | 30 | 30 | 45 | 45 | 210 |
| Year 1 | 23 | 27 | 23 | 24 | 33 | 41 | 171 |
| Year 2 | 22 | 21 | 19 | 17 | 35 | 42 | 156 |

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, paper plus desktop, handheld plus desktop) and one of two kinds of intervention support (no mentor or website mentor). One hundred seventy-one schools with a total of 542 first grade classrooms agreed to participate in the 2003–2004 school year. In the following year, 2004–2005, 156 schools and a total of 435 second grade classrooms continued their participation. Table 1 provides the number of proposed and actual urban and rural schools in the study design by administration format and intervention support.

According to free and reduced price lunch data, 51.47% of the students in the rural schools and 73% of the students in the urban schools were from economically disadvantaged families. Free and reduced lunch is commonly used as a proxy measure for poverty. Plan eligibility is determined annually by the U.S. Department of Agriculture using Federal Poverty Guidelines. Students with a household income 130% and 185% of the Federal Poverty Guidelines are eligible for free and reduced lunch respectively (Federal Register/Vol. 68, No. 49/Thursday, March 13, 2003/ Notices, pp. 12029). The racial/ethnicity breakdown of the students in the rural schools was: 9.16% African American; 35.86% Hispanic; 43.98% Caucasian; and 11% Other. For the students in the urban schools, the racial/ethnicity breakdown was: 20.67% African American; 63.67% Hispanic; 13.33% Caucasian; and 2.33% Other. Gender ratios were balanced in both types of schools.

Assessment instrument

The Texas Primary Reading Inventory (TPRI, 2006) was the assessment administered by the classroom teachers in this study. The TPRI was developed as a result

of a legislative mandate in 1997 to provide a diagnostic reading instrument for K-2 teachers in Texas (Foorman et al., 2004). In first and second grades, the screen consists of phonological awareness and its theoretically related construct of letter-sound knowledge in the beginning of grade 1 and word reading at the beginning and end of grade 1 and beginning of grade 2. The inventories in grade 1 and grade 2 are aligned with the Texas state curriculum standards and consist of the following components: (1) phonemic awareness in grade 1 (four tasks involving blending and segmenting onset-rimes and phonemes and deleting initial and ending sounds); (2) graphophonemic knowledge (the recognition of letter-sounds and five word-building tasks targeting initial and final consonants, medial vowels, and initial and final blends in grade 1; a spelling dictation test in grade (2); (3) reading accuracy and fluency in grades 1 and 2 (the ability to read grade appropriate text accurately and fluently); (4) reading comprehension in grades 1 and 2 (five questions assessing the understanding of what has been read). Students are placed into one of five passages based on empirical links between accuracy on a word list and accuracy in a passage.

Classroom teachers administer the TPRI individually to each student in her classroom at three different times during the course of the school year. At the beginning of the year, mid-August, the students are administered the screen and the inventory. The middle of the year assessment is administered at mid-January and consists of the inventory section only. The end of the year assessment is administered in mid-April and consists of the screen and inventory in grade 1 and just the inventory in grade 2. The inventory was designed to assess skills the students should acquire by the end of the school year; therefore, the teacher administers the tasks within the inventory until students reach a ''still developing'' level. For example, the teacher individually administers the screening section and, based on the student's performance moves to either the first task within the inventory (phonemic awareness in grade 1 or graphophonemic knowledge in grade 2) or to the reading accuracy, fluency, and comprehension portion of the inventory. At the next administration interval the teacher starts each student at their first ''still developing'' task within the phonemic awareness and/or graphophonemic knowledge portion of the inventory and also administers the reading accuracy, fluency, and comprehension portions to all students.

Intervention conditions

*Assessment format*

Teachers were randomly assigned to one of three administration formats: paper, paper plus desktop, or handheld plus desktop. The paper condition required that the teachers use the TPRI kit as they were used to using it, by completing one record sheet per student and then transferring the individual scores to a class summary sheet for evaluation. The paper plus desktop condition required that the teacher administer the assessment with the TPRI kit and then transfer the item-level scores to a secure website. Upon completion of data entry, the website electronically transferred the scores to the class summary sheet for evaluation. The handheld plus desktop condition administered the TPRI according to the test procedures but

instead of recording scores onto the paper copy of the individual student record sheets, the teachers 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 into a secure website.

*Teacher support*

The goal of this study condition was to determine whether web mentoring would assist teachers in using data to inform classroom instruction and would result in improved student outcomes. Teachers in this condition accessed a website that provided guidance on how to group students based on assessment data. Web mentoring was contrasted with the standard condition of teachers working alone to interpret assessment data. Teachers were randomly assigned to web mentoring or no mentoring and remained in the assigned condition throughout the study.

Design and analysis

*Study design*

The primary focus of this study was to determine if a student's score in year one alone (i.e., in grade 1 alone) predicted that student's score in year two (i.e., in grade 2) or if a combination of that student's score and the mean of the classroom that the student belonged to in year one predicted that student's score in year two. We also examined whether administration format (i.e., paper, paper plus desktop, or handheld plus desktop) and teacher support (i.e., web mentoring or no mentoring) moderated that prediction. All analyses were intent-to-treat in that schools' data were analyzed according to the condition to which they were originally assigned. In addition, we wanted to determine the role of the year one to year two pair of teachers who instructed students as they moved from grade 1 to grade 2. Namely, did the effect of the student score vary by teacher-pair or not? In other words, was the slope the same or did it vary randomly at the teacher-pair level?

To examine these questions, this study used a $3 \times 2$ factorial design with 3 levels of administration format (paper, paper plus desktop, or handheld plus desktop) and 2 levels of teacher support (web mentoring or no mentoring).

The data were nested at three levels: the student level, the teacher-pair level, and the school level. The teacher-pair level was the pairing of a student's pretest and posttest teacher. For example, in looking at a student's grade 1 score predicting that student's grade 2 score, we paired that student's grade 1 teacher with his or her grade 2 teacher. Because the moderator variables were randomly assigned at the school level, a student that changed schools from pretest year to posttest year was not included in the study.

*Predictors*

The predictors were TPRI subtasks administered to the students at the end of the year during the 2003–2004 school year when they were in first grade. The grade 1

word reading predictor was a composite (i.e., total correct) of words read correctly out of eight on the screening and out of 15 on the word list subtask. The grade 1 fluency predictor was the student's fluency rate (i.e., words read correctly per minute) in the passage reading subtask.

*Outcomes*

The outcomes were TPRI subtasks administered to the students at the end of the year during the 2004–2005 school year when they were in second grade. The grade 2 word reading outcome was the number of words out of 20 read correctly on the word list subtask. The grade 2 fluency outcome was the student's fluency rate in the passage.

*Model specification*

To address the questions of the study, we compared several models. The models differ in the specification of the slope coefficient, random or fixed, and in the specification of the predictors. Below are the equations corresponding to each model:

$$Y_{ijk} = \beta_{0jk} + e_{ijk} \tag{1}$$

$$Y_{ijk} = \beta_{0jk} + \beta_1(X_{ijk} - X'') + e_{ijk} \tag{2}$$

$$Y_{ijk} = \beta_{0jk} + \beta_{1k}(X_{ijk} - X'') + e_{ijk} \tag{3}$$

$$Y_{ijk} = \beta_{0jk} + \beta_{1k}(X_{ijk} - X'') + \beta_2(X'_k) + e_{ijk} \tag{4}$$

$$Y_{ijk} = \beta_{0jk} + \beta_{1k}(X_{ijk} - X'') + \beta_2(X'_k) + \beta_3 (X_{ijk} - X'') * (X'_k) + e_{ijk} \tag{5}$$

$$Y_{ijk} = \beta_{0jk} + \beta_{1k}(X_{ijk} - X'_k) + e_{ijk} \tag{6}$$

$$Y_{ijk} = \beta_{0jk} + \beta_{1k}(X_{ijk} - X'_k) + \beta_2(X'_k) + e_{ijk} \tag{7}$$

$$Y_{ijk} = \beta_{0jk} + \beta_{1k}(X_{ijk} - X'_k) + \beta_2(X'_k) + \beta_3(X_{ijk} - X'_k) * (X'_k) + e_{ijk} \tag{8}$$

$$Y_{ijk} = \beta_{0jk} + \beta_1(X_{ijk} - X'_k) + \beta_2(X'_k) + e_{ijk} \tag{9}$$

$$Y_{ijk} = \beta_{0jk} + \beta_1(X_{ijk} - X'_k) + \beta_2 (X'_k) + \beta_3(X_{ijk} - X'_k) * (X'_k) + e_{ijk} \tag{10}$$

where:

$$\beta_{1k} \sim N(\beta_1, \sigma_{\beta 1k})$$

$i$ = student; $j$ = posttest classroom; $k$ = pretest classroom; $X'_k$ = pretest classroom mean; $X''$ = grand mean.

For all the models the intercept is random at the teacher-pair level nested within the school level. Equation 1 is the random intercepts base model with no predictors. Equation (2) is an ANCOVA model with random intercepts and parallel slopes. This model has a fixed slope with the student score centered around the grand mean as a predictor. The fixed slope constrains the relationship between the predictor and the

outcome to be the same for all pretest and posttest teacher pairing. Equations (3) to (8) have a randomly varying slope at the teacher-pair level nested within the school level. The randomly varying slope allows the relationship between the predictor and the outcome to be different for each pretest-posttest teacher pairing. Equation (3) has one predictor, the student score centered around the grand mean. Equation (4) is similar to Eq. (3) but has an additional predictor, the pretest classroom mean. Equation (5) adds the interaction between the two predictors of Equation (4). Equation (6) has the student score deviated from the pretest classroom mean as a predictor. Equation (7) is similar to Eq. (6) except it adds another predictor, the pretest classroom mean. Equation (8) adds the interaction between the two Eq. (7) predictors. Equation (9) is similar to Eq. (6) except it has a fixed slope. Equation (10) is similar to Eq. (8) except it has a fixed slope. To determine which of the 10 equations has the best model specification, we looked at various fit statistics.

## Statistical analysis

The SAS® version 9.1 PROC MIXED procedure was used to analyze Eqs. (1)–(10) with and without the moderators. The moderators included two variables: the administration format and the type of mentoring received, and their interaction. Preliminary runs showed a different covariance structure for urban schools and rural schools; therefore the PROC MIXED analyses were performed using GROUP = AREATYPE in the RANDOM statement to obtain separate covariance parameters for the urban and the rural schools. We used the maximum likelihood (ML) method of estimating the covariance parameters instead of the default restricted maximum likelihood (REML). When using the fit statistics, REML restricts us to comparing models with nested covariance parameters alone (Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2006, pp. 752–754), which would confine the comparisons to Eq. (2) versus Eq. (3), Eq. (7) versus Eq. (9), and Eq. (8) versus Eq. (10). This restriction does not apply to the ML method, enabling us to compare across the models specified in Eqs. (1)–(10), which vary by both the covariance parameters and the predictors. ML estimates are unbiased when the sample sizes are not small. An alpha level of .05 was used for all statistical tests.

Intraclass correlations (ICCs) were calculated for each variance component. The ICCs were calculated by taking the variance estimate of each intercept and slope, if applicable, and dividing it by the sum of all variance components for that area type (urban or rural) and the residual estimates.

## Results

Grade 1 predicting Grade 2 outcomes

The means and standard deviations for the pretest and posttest variables are shown in Table 2. On average, students at the end of second grade were reading 12.36 words correctly out of 20 words on the list ($SD = 3.33$) and reading instructional-level passages at 89.16 words correct per minute ($SD = 33.97$). At the first grade

**Table 2** Means and standard deviations of predictor and outcome variables

| Variable | N | Mean | SD |
|---|---|---|---|
| *Outcomes* | | | |
| Grade 2 word reading student score | 4,911 | 12.36 | 3.33 |
| Grade 2 fluency student score | 4,465 | 89.16 | 33.97 |
| *Predictors* | | | |
| Grade 1 word reading student score | 5,197 | 0.00 | 5.81 |
| Grade 1 word reading student score deviated from classroom mean | 5,197 | 0.74 | 4.77 |
| Grade 1 word reading classroom mean | 5,243 | 14.97 | 3.56 |
| Grade 1 fluency student score prettest | 5,036 | 0.00 | 31.74 |
| Grade 1 fluency student score deviated from classroom mean | 5,036 | 1.26 | 26.24 |
| Grade 1 fluency classroom mean | 5,240 | 63.24 | 17.90 |

pretest, students were reading 14.97 words correctly out of 23 words ($SD$ = 3.56) and 63.24 words correct per minute ($SD$ = 17.90). These fluency rates are very similar to the Hasbrouck and Tindal (2005) means of 60 words correct per minute at the end of Grade 1 and 90 words correct per minute at the end of Grade 2. The means of 0.00 for the word reading and fluency rate predictors reflect the centering of the scores around the grand mean.

The sample sizes in Table 2 were derived in the following way. There were 9,275 students in the first grade dataset in 2003–2004 and 7,763 students in the second grade dataset in 2004–2005. When combined, these first and second grade datasets consisted of 12,702 unique students, all administered the TPRI (and not a Spanish language assessment instead). However, 4,936 students did not have second grade data, therefore, leaving 7,766 students. Of these, 2,475 students had second grade data but no first grade data, which reduced the dataset to 5,288 students. Of these, 44 students changed schools. Thus, 5,244 students were qualified to be used for analyses. Further reduction in sample size occurred for particular measures because of missing values. For word reading, 372 students were excluded due to missing values; for fluency rate, 842 students were excluded for this reason. Some of these missing fluency values were by test design in that students who could not manage grade-level stories did listening comprehension in grade 1 (184 students) or read first grade passages in grade 2 (277 students). In the case of 50 first graders, teachers went outside test procedures and gave them grade 2 passages to read.

The number of teacher-pairs included in the analyses was 1,336. The reason that this number exceeds the sum of the 427 first grade classrooms and the 396 second grade classrooms (i.e., 823) is that a first grade teacher may be paired with several second grade teachers because not all students move to the same second grade classroom. For example, five students from a particular teacher's first grade classroom might move together to teacher A's second grade classroom, 6 to teacher B's second grade classroom, 4 to teacher C's second grade classroom, and 8 to teacher D's second grade classroom.

## Word reading outcome with moderators

We ran Eqs. (1)–(10) with the moderators to determine if either administration format or level of teacher support predicted Grade 2 word reading outcome. The results showed no main or interaction effects of the moderator variables. Therefore, we proceeded with the analysis of Eqs. (1)–(10) without the moderators.

## Fluency outcome with moderators

We ran Eqs. (1)–(10) with the moderators to determine if either administration format or level of teacher support predicted Grade 2 fluency outcome. The results showed no main or interaction effects of the moderator variables. Therefore, we proceeded with the analysis of Eqs. (1)–(10) without the moderators.

## Word reading outcome without moderators

For grade 1 predicting grade 2 word reading outcome, the PROC MIXED results of Eqs. (1)–(10) produced ICCs for the intercept term ranging from 6% to 30%. By examining the intercept for the teacher-pairs within rural and urban school in the top of Tables 3, 4, and 5, one can see about twice as much variability at the classroom level in urban schools compared to rural schools across models. In contrast, the ICCs for the slope term ranged only from 0.2% to 0.3% and are comparable for rural and urban schools. All variance estimates are statistically significant, with $P$ values ranging from $P < .0001$ to $P = .004$.

As is evident in Table 6, Eq. (5) is the best specified model based on the fit statistics (see Table 6). Equation (5) type 3 tests of fixed effects showed significant main effects for the student score centered around the grand mean, $F(1,4741) = 397.68$, $p < .001$, and for the pretest classroom mean, $P(1,4741) = 33.42$, $p < .001$. There was also a significant interaction between the two predictors, $F(1,4741) = 22.56$, $p < .001$. A graph of this interaction is shown in Fig. 1, with means plotted for student and classroom values at or below the 25th percentile and high student and classroom values at or above the 75th percentile. By examining the figure, we see that students and classrooms with high word reading scores in grade 1 tend to have high word reading scores in grade 2. However, when a student scores low in grade 1, the classroom he or she belongs to does not matter as much as when a student scores high. Thus, a student who has high word reading scores in grade 1 will have higher word reading scores in grade 2 if he or she is a member of a high-scoring classroom.

## Fluency outcome without moderators

For grade 1 predicting grade 2 fluency outcomes, the PROC MIXED results of Eqs. (1)–(10) produced ICCs for the intercept term ranging from 6% to 26%. By examining the bottom half of Tables 3, 4, and 5, we see that urban schools have about twice as much variability within classrooms as rural schools across models; however, in some models, particularly model 6, the school-level variability is also

**Table 3** Variance components estimates, ICCs, and standard errors for grade 1 predicting grade 2 models 1–3

| Variance component | Model 1 | | | Model 2 | | | Model 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Estimate | ICC | SE | Estimate | ICC | SE | Estimate | ICC | SE |
| *Grade 1 predicting grade 2 word reading* | | | | | | | | | |
| Intercept teacher-pair within school rural | 0.857 | 0.08819 | 0.185 | 0.413 | 0.08432 | 0.092 | 0.315 | 0.07038 | 0.090 |
| Slope teacher-pair within school rural | | | | | | | 0.015 | 0.00327 | 0.003 |
| Intercept teacher-pair within school urban | 2.631 | 0.21846 | 0.301 | 0.882 | 0.16376 | 0.125 | 0.766 | 0.15607 | 0.122 |
| Slope teacher-pair within school urban | | | | | | | 0.013 | 0.00257 | 0.002 |
| Intercept school level rural | 0.604 | 0.06217 | 0.173 | 0.368 | 0.07525 | 0.100 | 0.367 | 0.08199 | 0.099 |
| Intercept school level urban | 1.156 | 0.09600 | 0.334 | 0.387 | 0.07187 | 0.125 | 0.350 | 0.07118 | 0.117 |
| Residual | 8.258 | | 0.189 | 4.115 | | 0.095 | 3.782 | | 0.094 |
| *Grade 1 predicting grade 2 fluency* | | | | | | | | | |
| Intercept teacher-pair within school rural | 82.327 | 0.08427 | 18.401 | 32.604 | 0.07284 | 8.297 | 33.492 | 0.08047 | 8.609 |
| Slope teacher-pair within school rural | | | | | | | 0.036 | 0.00009 | 0.007 |
| Intercept teacher-pair within school urban | 242.233 | 0.19127 | 30.966 | 84.617 | 0.16688 | 13.053 | 77.361 | 0.16392 | 13.026 |
| Slope teacher-pair within school urban | | | | | | | 0.036 | 0.00008 | 0.009 |
| Intercept school level rural | 58.838 | 0.06023 | 17.630 | 27.737 | 0.06197 | 8.671 | 31.169 | 0.06604 | 9.397 |
| Intercept school level urban | 188.518 | 0.14885 | 47.050 | 35.172 | 0.06937 | 13.104 | 43.053 | 0.09122 | 14.574 |
| Residual | 835.725 | | 20.020 | 387.259 | | 9.368 | 351.503 | | 9.107 |

**Table 4** Variance components estimates, ICCs, and standard errors for grade 1 predicting grade 2 models 4–6

| Variance component | Model 4 | | | Model 5 | | | Model 6 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Estimate | ICC | SE | Estimate | ICC | SE | Estimate | ICC | SE |
| *Grade 1 predicting grade 2 word reading* | | | | | | | | | |
| Intercept teacher-pair within school rural | 0.289 | 0.06556 | 0.086 | 0.291 | 0.06647 | 0.086 | 0.924 | 0.17181 | 0.124 |
| Slope teacher-pair within school rural | 0.014 | 0.00314 | 0.003 | 0.014 | 0.00309 | 0.003 | 0.013 | 0.00242 | 0.003 |
| Intercept teacher-pair within school urban | 0.738 | 0.15299 | 0.120 | 0.741 | 0.15315 | 0.120 | 2.189 | 0.30189 | 0.215 |
| Slope teacher-pair within school urban | 0.014 | 0.00281 | 0.003 | 0.012 | 0.00251 | 0.002 | 0.017 | 0.00228 | 0.004 |
| Intercept school level rural | 0.330 | 0.07493 | 0.092 | 0.297 | 0.06800 | 0.085 | 0.697 | 0.12954 | 0.181 |
| Intercept school level urban | 0.301 | 0.06240 | 0.105 | 0.315 | 0.06506 | 0.107 | 1.300 | 0.17922 | 0.336 |
| Residual | 3.773 | | 0.093 | 3.771 | | 0.093 | 3.746 | | 0.096 |
| *Grade 1 predicting grade 2 fluency* | | | | | | | | | |
| Intercept teacher-pair within school rural | 31.771 | 0.07673 | 8.475 | 33.211 | 0.08038 | 8.512 | 81.226 | 0.16548 | 11.088 |
| Slope teacher-pair within school rural | 0.036 | 0.00009 | 0.007 | 0.035 | 0.00009 | 0.007 | 0.040 | 0.00008 | 0.008 |
| Intercept teacher-pair within school urban | 76.549 | 0.16103 | 12.824 | 83.209 | 0.17106 | 13.129 | 181.433 | 0.25516 | 19.047 |
| Slope teacher-pair within school urban | 0.035 | 0.00007 | 0.009 | 0.021 | 0.00004 | 0.007 | 0.064 | 0.00009 | 0.013 |
| Intercept school level rural | 30.689 | 0.06456 | 9.203 | 30.733 | 0.07438 | 9.257 | 66.190 | 0.09309 | 17.324 |
| Intercept school level urban | 47.258 | 0.09941 | 15.197 | 54.007 | 0.11103 | 16.287 | 186.157 | 0.26180 | 43.143 |
| Residual | 351.541 | | 9.108 | 349.186 | | 8.989 | 343.408 | | 8.993 |

**Table 5** Variance components estimates, ICCs, and standard errors for grade 1 predicting grade 2 models 7–10

| Variance component | Model 7 | | | Model 8 | | | Model 9 | | | Model 10 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Estimate | ICC | SE | Estimate | ICC | SE | Estimate | ICC | SE | Estimate | ICC | SE |
| *Grade 1 predicting grade 2 word reading* | | | | | | | | | | | | |
| Intercept teacher-pair within school rural | 0.439 | 0.09535 | 0.091 | 0.442 | 0.09623 | 0.091 | 0.374 | 0.07774 | 0.088 | 0.380 | 0.07920 | 0.088 |
| Slope teacher-pair within school rural | 0.011 | 0.00234 | 0.003 | 0.011 | 0.00229 | 0.003 | | | | | | |
| Intercept teacher-pair within school urban | 0.932 | 0.18257 | 0.129 | 0.913 | 0.18049 | 0.126 | 0.879 | 0.16525 | 0.124 | 0.864 | 0.16388 | 0.123 |
| Slope teacher-pair within school urban | 0.012 | 0.00240 | 0.003 | 0.012 | 0.00232 | 0.003 | | | | | | |
| Intercept school level rural | 0.346 | 0.07520 | 0.096 | 0.347 | 0.07556 | 0.097 | 0.330 | 0.06851 | 0.092 | 0.332 | 0.06909 | 0.092 |
| Intercept school level urban | 0.352 | 0.06896 | 0.120 | 0.340 | 0.06718 | 0.117 | 0.331 | 0.06229 | 0.113 | 0.320 | 0.06062 | 0.110 |
| Residual | 3.807 | | 0.097 | 3.794 | | 0.096 | 4.110 | 0.00000 | 0.094 | 4.088 | 0.00000 | 0.094 |
| *Grade 1 predicting grade 2 fluency* | | | | | | | | | | | | |
| Intercept teacher-pair within school rural | 35.137 | 0.08606 | 8.052 | 34.967 | 0.08575 | 8.000 | 31.652 | 0.07096 | 8.188 | 33.157 | 0.07563 | 8.158 |
| Slope teacher-pair within school rural | 0.041 | 0.00010 | 0.008 | 0.040 | 0.00010 | 0.008 | | | | | | |
| Intercept teacher-pair within school urban | 84.349 | 0.17914 | 12.510 | 85.862 | 0.18286 | 12.507 | 82.604 | 0.16185 | 12.762 | 83.427 | 0.16634 | 12.625 |
| Slope teacher-pair within school urban | 0.062 | 0.00013 | 0.013 | 0.037 | 0.00008 | 0.010 | | | | | | |
| Intercept school level rural | 30.229 | 0.06420 | 9.161 | 30.582 | 0.07499 | 9.210 | 27.435 | 0.06151 | 8.496 | 27.541 | 0.06282 | 8.481 |
| Intercept school level urban | 43.546 | 0.09248 | 14.216 | 41.459 | 0.08829 | 13.739 | 40.788 | 0.07992 | 13.973 | 40.408 | 0.08057 | 13.771 |
| Residual | 342.895 | | 8.892 | 342.200 | | 8.839 | 386.971 | | 9.351 | 377.702 | | 9.131 |
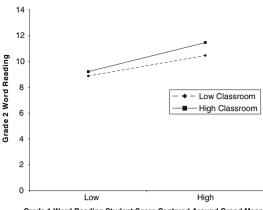
**Table 6** Fit statistics for models 1–10

| Fit statistic | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 | Model 8 | Model 9 | Model 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| *Grade 1 predicting grade 2 word reading* | | | | | | | | | | |
| −2 Res log likelihood | 25,093 | 21,387 | 21,294 | 21,260 | 21,238 | 21,776 | 21,307 | 21,281 | 21,357 | 21,330 |
| AIC (smaller is better) | 25,105 | 21,401 | 21,312 | 21,280 | 21,260 | 21,794 | 21,327 | 21,303 | 21,373 | 21,348 |
| AICC (smaller is better) | 25,105 | 21,401 | 21,312 | 21,280 | 21,260 | 21,794 | 21,327 | 21,303 | 21,373 | 21,348 |
| BIC (smaller is better) | 25,123 | 21,422 | 21,338 | 21,309 | 21,292 | 21,820 | 21,356 | 21,335 | 21,396 | 21,374 |
| *Grade 1 predicting grade 2 fluency* | | | | | | | | | | |
| −2 Res log likelihood | 43,400 | 39,291 | 39,152 | 39,146 | 39,103 | 39,562 | 39,122 | 39,062 | 39,283 | 39,283 |
| AIC (smaller is better) | 43,412 | 39,305 | 39,174 | 39,170 | 39,129 | 39,582 | 39,146 | 39,088 | 39,299 | 39,299 |
| AICC (smaller is better) | 43,412 | 39,305 | 39,174 | 39,170 | 39,129 | 39,582 | 39,146 | 39,088 | 39,300 | 39,300 |
| BIC (smaller is better) | 43,429 | 39,325 | 39,206 | 39,205 | 39,166 | 39,611 | 39,180 | 39,125 | 39,323 | 39,323 |

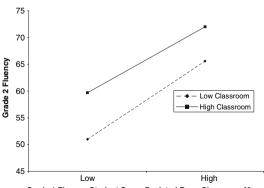**Fig. 1** Interaction of student score centered around the grand mean with pretest classroom mean in predicting grade 2 word reading

substantially larger in urban schools compared to rural schools (i.e., 26% vs. 9%). The ICCs for the slope term ranged from 0.004% to 0.013%, and are comparable for rural and urban schools. All variance estimates are statistically significant, with $p$ values ranging from $p < .0001$ to $p < .007$. As shown in Table 6, Eq. (8) is the best specified model based on the fit statistics. Equation (8) type 3 tests of fixed effects showed significant main effects for the student score deviated from the pretest classroom mean, $F(1,4286) = 559.96$, $p < .001$, and for the pretest classroom mean, $F(1,4286) = 593.82$, $p < .001$. There was also a significant interaction between the two predictors, $F(1,4286) = 65.81$, $p < .001$. This interaction is displayed in Fig. 2, with means plotted for student and classroom values at or below the 25th percentile and high student and classroom values at or above the 75th percentile. Fig. 2 reveals that students and classrooms with high fluency scores in first grade tend to have high fluency scores in second grade. However, when students have high scores in first grade, the classroom he or she belongs to does not matter as much as when a student scores low. Thus, a student who has low fluency scores in grade 1 will have higher fluency scores in grade 2 if he or she is a member of a high-scoring classroom.



**Fig. 2** Interaction of student score deviated from the pretest classroom mean with pretest classroom mean in predicting grade 2 fluency rate

Calculation of social tracking index

In order to better understand the cause of the relatively large ICCs in urban schools relative to non-urban schools, we examined the average number of classrooms in the schools in each setting and the average number of teacher-pairs. Surprisingly, the average number of classrooms was similar. In the 221 urban, first-grade classrooms, the mean number of classrooms per school was 3.16 ($SD$ = 1.59). In the 206 rural, first-grade classrooms, the average was 3.22 ($SD$ = 2.27). In the 204 urban, second-grade classrooms, the mean number of classrooms was 2.91 ($SD$ = 1.33). In the 192 rural, second-grade classrooms, the average was 3.00 ($SD$ = 1.89). In contrast, the average number of teacher-pairs was different: Out of the 622 observed teacher-pairings in urban schools, the average per school was 8.88 ($SD$ = 7.56); out of the 712 observed teacher-pairings in rural schools, the average per school was 11.12 ($SD$ = 12.54). Thus, in spite of similar numbers of first and second grade classes within schools, there was a tendency for students in urban schools to move to the next grade level in intact groups and a tendency for students in rural schools to disperse to new groups.

To better understand the degree of this tracking, we computed a tracking index by looking at 1—the ratio of the number of observed pairings between grades relative to the number of possible pairings (i.e., the product of the number of teachers in the two grades). If all pairings were observed, the index would be 0. If half the possible pairings were observed, the number would be 0.50. The average tracking index for urban schools was 0.14 ($SD$ = .18); for rural schools it was 0.08 ($SD$ = .17). The fact that this index was almost twice as large in urban schools supports the observation that students in urban schools were more likely than students in rural schools to be advanced to the next grade with their same peer group. To see if this social tracking was related to school-level variability in word reading and fluency at the beginning of second grade, we calculated the ICC for each school j, where:

$$\mathrm{ICC_j} = \text{variance of classroom means/variance for school j (total variance)}$$

Then we plotted the urban and rural school-level ICCs for word reading and fluency at the beginning of second grade against the social tracking index. Figure 3 depicts these plots and shows the tendency for greater social tracking from first to second grade to coincide with greater school level variability in word reading and fluency at the beginning of grade 2.

## Discussion

This study utilized early reading assessment data from a randomized trial in Texas schools to examine contextual effects on risk prediction in first and second grade. A primary objective was to determine if a student's word reading or fluency scores in first grade predicted that student's scores in second grade or if a combination of that student's scores and the mean of the classroom to which the student belonged in first grade predicted that student's scores in second grade. We also examined whether
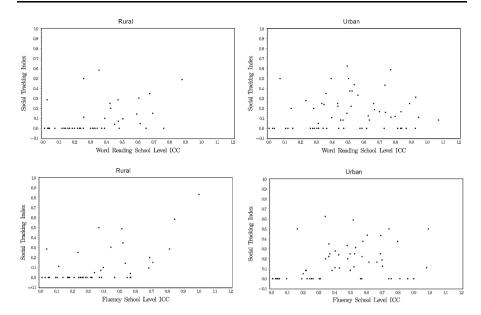
**Fig. 3** School level ICCs plotted against social tracking index for word reading and fluency outcomes

administration format (i.e., paper, paper plus desktop, or handheld plus desktop) and teacher support (i.e., web mentoring or no mentoring) moderated that prediction. All analyses were intent-to-treat, meaning that schools' data were analyzed according to the condition to which they were originally assigned. In addition, we wanted to determine the role of the year one to year two pair of teachers who instructed students as they moved from grade 1 to grade 2. Namely, did the effect of the student score vary by teacher-pair or not? In other words, was the slope the same or did it vary randomly at the teacher-pair level?

Preliminary analyses revealed different covariance structures for urban and rural schools. Therefore, a factor called area type was included in the models and separate covariance parameters were obtained. Additionally, initial analyses of word reading and fluency outcomes indicated that there were no significant effects of the moderator variables (administration format and teacher support). Although disappointing, these results are consistent with a weak impact on achievement gains of technology unless it is used to facilitate specific learning objectives (e.g., Wenglinski, 1998) and of web mentoring (e.g., Sinclair, 2003).

The roles of individual differences and classroom in predicting risk

The answer to the question of whether student pretest in first grade or a combination of student pretest and mean of pretest classroom was a better predictor of grade two outcomes was straightforward: A combination of student pretest and mean of pretest classroom is a better predictor than student pretest alone. This is a very important finding given that nearly all risk prediction is based on student pretest scores alone,

as we saw in the literature reviewed in the introduction. The response-to-intervention (RtI) approach to identifying students for reading disabilities offers a possible exception in that a student's pretest score is used for initial placement into an intervention, but subsequent placement is determined by response in an instructional context.

The answer to the question of whether the effect of student scores varied by teacher-pair is ''yes.'' Slope varied randomly at the teacher-pair level, meaning that the relationship between predictor and outcome was different for each pretest-posttest pairing of teachers. This finding is significant because it underscores that student achievement gains vary as a function of the particular pairing of teachers a student has when moving from first to second grade.

The most common analytic approach to risk prediction is an ANCOVA model with random intercepts and fixed slopes. In this approach, the pretest covariate is the student's deviation from the grand mean and the fixed slope means that student scores are centered around the grand mean. The best fitting models in this study had random slopes and a combination of student pretest score and pretest classroom mean. In the analysis of word reading outcomes, the best fitting model conceptualized student pretest as the student's first grade word reading score centered around the grand mean. In contrast, in the analysis of fluency outcomes, the best fitting model conceptualized student pretest as the student's first grade fluency score deviated from the pretest classroom mean. Graphs of the two interactions indicated that a student who had high word reading scores in first grade had higher word reading scores in second grade if he or she was a member of a high-scoring classroom. In contrast, a student who had low fluency scores in first grade had higher fluency scores in second grade if he or she was a member of a high-scoring classroom.

Thus, in the case of word reading outcomes, the first grade classroom mean interacted with a first grader's ability to read words in comparison to *all* first graders, and high ability was best nurtured in a high-ability classroom environment. This is an individual difference notion of risk interacting with a classroom notion of risk. In the case of fluency outcomes, however, the first grade classroom mean interacted with a first grader's difference from his or her first grade classroom peers. Here the student's rate was best conceptualized as rate relative to classroom rate, and low relative rate was best nurtured in a high relative rate classroom environment. This is a peer difference rather than an individual difference notion of risk interacting with a classroom notion of risk.

It is possible that this subtle difference in the nature of the covariate predicting risk reflects a difference between accuracy and rate measures. Word reading on the TPRI is an untimed measure of the number of words read correctly on word lists. Fluency on the TPRI is a contextualized measure of rate within the reading comprehension task: number of words read correctly in one minute on instructional-level passages. Thus, for a slow reader, being in a classroom of faster reading peers seems to be an intervention all by itself. In short, do what your peers do: read faster. The implications of this finding are enormously important. The primary intervention for fluency recommended by the National Reading Panel (National Institute of Child Heath and Human Development, 2000) is guided repeated reading. What this study

suggests is that the word ''guided'' may be the operative word in the recommended practice of ''guided repeated reading.'' In other words, it is not repeated reading alone that is important; it is repeated readings in the context of faster reading models, in this case faster reading peers in the classroom (although tape recorded or computer recording readings of books may have similar effects).

ICCs and social tracking

For researchers planning group randomized trials in education, the values of ICCs are hotly disputed (e.g., Hedges, & Hedberg, 2006) because the variability between classrooms and between schools affects power calculations, which, in turn, affect the number of classrooms and/or schools needed. On average, the ICCs for the intercept term in the models in this study ranged from 6% to 17%, although the maximum extended to 30% for word reading and to 26% for fluency. These values are higher than the 3% to 5% typically assumed in power calculations (Hedges & Hedberg, 2006). For all models but one, the differences in ICCs between rural and urban schools were much greater at the classroom level than at the school level. Moreover, at the classroom level, the ICCs for urban schools were approximately twice as large as those for rural schools.

Initially we speculated that this difference in ICCs was due to rural schools having fewer classrooms and, therefore, fewer pairings of teachers from first to second grade. However, upon examination, the average number of classrooms in the two types of schools was remarkably similar, yet the average number of teacher pairings was much larger in urban schools. In short, there was more of a tendency for first graders in these urban schools to move to second grade in intact groups than in the rural schools. We created a tracking index to compare the number of observed pairings to the number of all possible pairings at a school. This index was about twice as large in urban schools as rural schools. When we plotted the tracking index for urban and rural schools against the school-level ICCs for word reading and fluency scores at the beginning of second grade, we confirmed a tendency for school-level variability in these reading measures to increase as evidence for tracking increased. Note that we used beginning of the year in Grade 2 rather than end of year data so that we could capture the peer effect separate from the teacher instruction effect evidenced in end of year data. We refer to this as *social tracking* rather than ability tracking because the complexity of comparing observed student level to classroom level scores at the beginning of first and second grade with Monte Carlo simulations of these comparisons is beyond the scope of this paper. Suffice it to say that social tracking was an important part of the context of risk prevention in the randomly selected schools in this study, especially in the urban schools.

We concur with Hanushek et al. (2005) that variation in achievement within schools is much larger than variation between schools and that teacher quality as measured by gains in achievement matters. What we add to this is that peers matter as well in creating a learning environment that interacts with student abilities and skills in beginning reading.

Limitations

Because school was the unit of random assignment, we did not have access to student level data on ethnicity/race, limited English proficiency, or participation in the free and reduced lunch program. Thus, we were unable to examine whether achievement gains were moderated by poverty as Desimone et al. (in press) found. Also, in our analyses the residual variance components were still significant. That means that other measures of student performance in grade 1 could improve the prediction of risk in grade 2 word reading and fluency outcomes.

## Conclusion

In this investigation of risk prediction in early reading we found that one cannot ignore the context that the students are in (i.e., classroom, school) and that the contextual relationship is complex. Models that ignore student-level covariates are missing an important part of the risk prediction, and significant factors exist at the student, classroom, and school levels. Additionally, the model for the pretest data to be used in the analysis is also important. In our data, the interaction of the pretest class mean with student deviation from their first-grade peers, in the case of word reading, or from their first-grade classroom in the case of fluency, was a significant predictor of second grade outcomes. A first grader with high word reading ability maximizes grade 2 outcomes by being in a high word-reading classroom in grade 2. In contrast, a first grader with low fluency is encouraged to read faster by being in a high fluency rate classroom in grade 2. These individual differences interacting with the classroom context were further explained by a social tracking phenomenon especially apparent in urban schools. The tendency for students to move together from first to second grades in urban schools accentuates the contextual effects of learning.

## References

Al Otaiba, S., & Fuchs, D. (2006). Who are the young children for whom best practices in reading are ineffective? An experimental and longitudinal study. *Journal of Learning Disabilities, 39*, 414–431.

Anthony, J. L., Lonigan, C. J., Driscoll, K., Phillips, B. M., & Burgess, S. R. (2003). Phonological sensitivity: A quasi-parallel progression of word structure units and cognitive operations. *Reading Research Quarterly, 38*, 470–487.

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

Clay, M. M. (1993). *Reading Recovery: A guidebook for teachers in training*. Portsmouth, NH: Heinemann.

Darling-Hammond, L., Berry, B., & Thorenson, A. (2001). Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis, 23*(1), 57–77.

Denton, C. A., Ciancio, D. J., & Fletcher, J. M. (2006). Validity, reliability, and utility of the observation survey of early literacy achievement. *Reading Research Quarterly, 41*, 8–34.

Desimone, L., Smith, T., & Frisvold, D. (in press). Has NCLB improved teacher and teaching quality for disadvantaged students. In A. Gamoran (Ed.), *Standards-based reform and the poverty gap: lessons for 'no child left behind'* Washington, DC: Brookings Institute.

Federal Register/Vol. 68, No. 49/Thursday, March 13, 2003/Notices, pp. 12029. Retrieved March 15, 2007 from http://www.fns.usda.gov/cnd/governance/ notices/iegs/IEGsHistorical.htm.

Fletcher, J. M., Coulter, W. A., Reschly, D. J., & Vaughn, S. (2004). Alternative approaches to the definition and identification of learning disabilities: Some questions and answers. *Annals of Dyslexia, 54*(2), 304–331.

Fletcher, J. M., Foorman, B. R., Boudousquie, A., Barnes, M., Schatschneider, C., & Francis, D. J. (2002). Assessment of reading and learning disabilities: A research-based, treatment-oriented approach. *Journal of School Psychology, 40*, 27–63.

Fletcher, J. M., Francis, D. J., Morris, R. D., & Lyon, G. R. (2005). Evidence-based assessment of learning disabilities in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34*(3), 506–522.

Foorman B. R. (Ed.) (2003). *Preventing and remediating reading difficulties: Bringing science to scale*. Austin, TX: Pro Ed.

Foorman, B. R., & Ciancio, D. J. (2005). Screening for secondary intervention: Concept and context. *Journal of Learning Disabilities, 38*, 494–499.

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.

Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis, 22*(2), 129–145.

Good, R. H., & Kaminiski, R. A. (2002). DIBELS *Oral reading fluency passages for first through third grades* (Technical Report No. 10). Eugene, OR: University of Oregon.

Good, R. H., Simmons, D. S., Kame'euni, E. J., Kaminski, R. A., & Wallin, J. (2002). *Summary of decision rules for intensive, strategic, and benchmark instructional recommendations in kindergarten through third grade* (Technical Report No. 11). Eugene, OR: University of Oregon.

Hale, J. B., Naglieri, J. A., Kaufman, A. S., & Kavale, K. A. (2004). Specific learning disability classification in the new Individuals with Disabilities Education Act: The danger of good ideas. *The School Psychologist, 29*, 6–13.

Hanushek, E. A., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality*. (Working paper 11154). Cambridge, MA: National Bureau of Economic Research.

Harcourt (1996). *Stanford-9 Achievement Test*. San Antonio, TX: Harcourt Brace & Company.

Hasbrouck, J., & Tindal, G. (2005). *Oral reading fluency: 90 years of measurement* (Technical Report No. 33). Eugene, OR: University of Oregon.

Hedges, L. V., & Hedberg, E. C. (2006). Interaclass correlation values for planning group randomized trials in education. Retrieved November 13, 2006, from http://drdc.uchicago.edu/what/re-search.html.

Individuals with Disabilities Education Improvement Act of 2004 (IDEA), Pub. L. No. 108–446, 118 Stat 614, (2004).

Invernizzi, M., Meier, J. D., Swank, L., & Juel, C. (2002). *Phonological awareness literacy screening (PALS)*. Charlottesville, VA: University of Virginia.

Leslie, L., & Caldwell, J. (2005). *Qualitative Reading Inventory-4*. New York, NY: HarperCollins.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. O., & Schabenberger, O. (2006). $SAS^{®}$ *for mixed models* (2nd ed.). Cary, NC: SAS Institute Inc.

Mathes, P., Denton, C., Fletcher, J., Anthony, J., Francis, D., & Schatschneider, C. (2005). The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly, 40*, 142–147.

McCardle, P., & Chhabra, V. (2004). *The voice of evidence in reading research*. Baltimore, MD: Brooks Publishing Co.

McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children, 71*, 445–463.

Molfese, V. J., Modglin, A. A., Beswick, J. L., Neamon, J. D., Berg, S. A., Berg, C. J., & Molnar, A. (2006). Letter knowledge, phonological processing, and print knowledge: Skill development in nonreading preschool children. *Journal of Learning Disabilities, 39*, 296–305.

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, DC: U. S. Department of Health and Human Services.

National Research Council (1998). Preventing reading difficulties young children. In C. E. Snow, M. S. Burns, & P. Griffin (Eds.), *Committee on the prevention of reading difficulties young children, commission on behavioral and social science and education*. Washington, DC: National Academy Press.

No Child Left Behind Act of (2001). Pub. L. No. 107–110, 115 Stat.1425.

Rayner, K., Foorman, B., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2*, 31–74.

Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97–110). New York, NY: The Guilford Press.

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

Sinclair, C. (2003). Mentoring online about mentoring: Possibilities and practice. *Mentoring and Tutoring, 11*(1), 79–94.

Texas Primary Reading Inventory (TPRI; 2006–2008). New York, NY: McGraw-Hill.

Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology, 40*(1), 7–26.

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional Children, 69*, 391–409.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early interventions as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88*, 601–638.

Vellutino, F. R., Scanlon, D. M., Small, S., & Fanuele, D. P. (2006). Response to intervention as a vehicle for distinguishing between children with and without reading disabilities: Evidence for the role of kindergarten and first-grade interventions. *Journal of Learning Disabilities, 39*, 157–169.

Wanzek, J., & Vaughn, S. (in press). Research-based implications from intensive early reading interventions. School Psychology Review.

Wenglinski, H. (1998). *Does it compute? The relationship between education technology and student achievement in math*. Princeton, NJ: ETS.

Woodcock, R. W., McGrew, K. S., & Mather N. (2001). *Woodcock-Johnson*®*-III Complete*. San Antonio, TX: The Riverside Publishing Company.