Beyond ORF: Student-Level Predictors of Reading Achievement

Author Info:

Angela I. Canto, Educational Psychology and Learning Systems, Florida State University; Briley E. Proctor, Educational Psychology and Learning Systems, Florida State University; Alicia L. Shafer, Educational Psychology and Learning Systems, Florida State University.

Briley E. Proctor is now at Bainbridge Island School District in Bainbridge Island, WA.
The authors would like to specifically acknowledge and thank the participating volunteer reading coaches, school psychologists, and school district personnel where data were collected. This manuscript is based on data and analyses from the first author's doctoral dissertation.

Correspondence regarding this manuscript should be addressed to Angela Canto, Educational Psychology & Learning Systems, 3210 Stone Building, PO Box 3064453, Tallahassee, FL 32306-4453. Email: acanto@fsu.edu

Abstract

This study explored student-level predictors of reading achievement among third grade regular education students. Predictors included student demographics (sex and socioeconomic status (SES), using free and reduced lunch as proxy for SES), direct observations of reading skills (oral reading fluency (ORF) and word decoding skill (nonsense word fluency/NWF), and academic history (number of prior grade retentions (retentions), Reading/Language Arts grades (reading grade), and attendance rate. Hierarchical linear regression results indicated that ORF and reading grade were statistically significant predictors of high-stakes reading achievement for this sample (model R2=.631). Results replicated previous findings of the predictive value of ORF, above and beyond economic disadvantage and highlighted the influence of low reading grades as an additional key predictor of poor reading achievement, with effect above and beyond that of ORF alone.

Keywords: oral reading fluency, reading, prediction, achievement

Beyond ORF: Student-Level Predictors of Reading Achievement

It is well-known that students' ability to read fluently (accurately, quickly, and with expression) is important for overall academic achievement (e.g., Armbruster, Lehr, & Osborn, 2001; Samuels, 2002). Some degree of automaticity in reading is needed for prompt comprehension of the printed text which helps the reader avoid becoming fixated on pronunciation of isolated words at the expense of understanding the text meaning (Sindelar, Lane, Pullen, & Hudson, 2002; Snow, Burns & Griffin, 1998). Indeed, fluent reading is a known predictor of reading comprehension—the ultimate prize or purpose for reading—with correlations between reading fluency and comprehension ranging between .70 and .90 (Baker, Gersten, & Grossen, 2002). Research consistently indicates that Oral Reading Fluency (ORF)—reading connected text aloud—is a critical indicator of general reading skill (Fuchs, 1995). When teachers use ORF data to establish individual student achievement goals, monitor the effects of instructional programs, and adjust interventions accordingly, student achievement improves (Connor, Morrison, & Petrella, 2004; Shinn, 1995; Shinn, Shinn, Hamilton, & Clarke, 2002; Stecker, & Fuchs, 2000).

ORF measures generally demonstrate strong overall technical adequacy (i.e., reliability and validity) (e.g., Deno, 1985, 1989; Fuchs, 1995; Fuchs, Fuchs, & Maxwell, 1988; Good & Jefferson, 1998; Hosp & Fuchs, 2005; Marston, 1989). As cited in these studies and Marston (1989), reliability measures are generally high with most estimates of test-retest reliability (ranging from .82 to .97) and parallel forms reliability (ranging from .84 to .96) being above .90. Inter-rater reliability estimates for ORF procedures have been achieved at .99 (Tindal, Marston, & Deno, 1983 as cited in Marston, 1989). In validity studies, researchers have concluded that ORF assessment procedures appear to result in data possessing adequate to strong validity overall (Fuchs et al., 1988; Marston, 1989). Additionally data obtained through ORF procedures appear to possess moderate to strong concurrent and discriminant validity with other measures of reading skill including oral passage reading, question-answering tests, recall of text procedures, cloze procedures of reading comprehension (i.e., missing word completion measure), and broader measures of reading comprehension (Fuchs et al., 1988).

Student ORF scores have been used to predict reading achievement on many state adopted criterion-referenced tests of achievement (e.g., Buck & Torgesen, 2003; Hixson & McGlinchey, 2004; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Shapiro, Keller, Lutz, Santoro & Hintze, 2006; Silberglitt, Burns, Madyun, & Lail, 2006; Wanzek, Roberts, Linan-Thompson, Vaughn, Woodruff, & Murray, 2010) as well as nationally norm-referenced tests of achievement (Hixson & McGlinchey, 2004; Klein & Jimerson, 2005; Roehrig, et al., 2008; Schilling, Carlisle, Scott, & Zeng, 2007; Wanzek et al., 2010). The proportion of variance explained by ORF in these studies tends to fall between 36% (e.g., Wanzek et al., 2010) and 64% (e.g., Hixson & McGlinchey, 2004), depending on the study and the predictor variables included in the model. Notably, Kranzler, Brownell, and Miller (1998) reported that ORF is not simply a proxy for underlying cognitive processes including cognitive ability, processing speed, and efficiency but rather contributes unique variance to the prediction of reading achievement.

One limitation in using ORF, however, is that studies of ORF predictive validity have had mixed results among some ethnic minority subgroups and students of low socioeconomic status (e.g., Buck & Torgesen, 2003; Crowe, Connor, & Petscher, 2009; Hintze, Callahan, Matthews & Williams, 2002; Hixson & McGlinchey, 2004; Hosp, Hosp, & Dole, 2011; Klein & Jimerson, 2005; Kranzler, Miller, & Jordan, 1999). Recently, Hosp, Hosp and Dole (2011) called for additional research noting that while the predictive validity of ORF was generally quite good, it "may not demonstrate consistent levels of predictive validity when focusing on different subgroups" (p. 125). Hosp and colleagues (2011) suggest that the source of this "predictive bias" is difficult to pinpoint. They offered several possible explanations, including the possibility that differences were the result of *a priori* decisions regarding variables included in the prediction models. In sum, ORF research suggests that it is a good overall predictor of reading achievement but that caution may be warranted when interpreting the predictive validity for specific subgroups. The research on predictive validity of ORF may need additional studies to determine the overall pattern (Hosp et al., 2011).

Efforts to improve the prediction of reading achievement by the inclusion of other student-level variables have been rare. The study by Hosp and colleagues (2011), for example, appears to be the only published report examining the relationship between word decoding skill in third grade and third grade high-stakes reading achievement. This is somewhat surprising because it has long been argued that, in addition to oral reading fluency, decoding is also a requisite skill requisite for success on high-stakes measures of reading achievement (Armbruster et al., 2001; Marston, 1989). In fact, text passages on year-end reading achievement tests often include higher-level decodable words (Hiebert, 2002) and decoding ability has been found to be a reliable indicator of persistent reading difficulties (Burke, Hagan-Burke, Kwok, & Parker, 2009). Thus, a measure of decoding may have utility for enhancing prediction of high-stakes reading achievement, but is yet unknown.

In addition to ORF and decoding, researchers are encouraged to explore additional variables that may enhance prediction of student reading achievement. Bishop and League (2006) highlight the importance of using a multivariate screening model of reading achievement. At this time, however, we know little about the impact of other student-level variables on reading achievement. Other variables such as students' reading grades, attendance rate, and prior grade retentions may also explain a significant portion of variance in high-stakes reading achievement scores above and beyond that of ORF. For example, research has shown only rare support for mean differences between sexes on ORF and norm-referenced measures (second grade spring differences between sexes on ORF; Klein & Jimerson, 2005), yet, sex differences have been documented on student grades (Burts, Hart, Charlesworth, & DeWolf, 1993) and grade retention (Jimerson, Carlson, Rotert, Egeland, & Sroufe, 1997; McCoy & Reynolds, 1999). Additionally, variables such as grades and prior grade retentions seem to have intuitive relationships with reading achievement overall; yet, whether the effects of those variables explain additional significant variance over ORF is unknown.

In summary, the purpose of the present study was threefold. First, we were interested in replicating earlier studies on the prediction of high-stakes reading achievement among third grade students using ORF while controlling for student demographics (economic disadvantage and sex). Students' free and reduced lunch status was used as a proxy for SES. It was hypothesized that our findings would be consistent with those reported in earlier investigations on the predictive utility of ORF, controlling for student demographics. Secondly, we wanted to test whether the inclusion of a measure of student decoding would help to improve the prediction model, given that the literature suggests that decoding may still be a factor on achievement on year-end high stakes reading tests. Thirdly, we wanted to explore whether prediction of high-stakes reading achievement among third grade students could be enhanced by the inclusion of additional student-level variables known to be implicated in overall school achievement. Thus, we included in the model data on the student's number of prior grade retentions, attendance rate, and reading grade. These final three variables are data that are readily available to teachers and do not require time or resources for additional direct measurement of student skill. It was hypothesized that the inclusion of these additional student-level variables would increase the proportion of explained variance in the prediction of reading achievement scores.

Methods

Participants

Third grade students (*n* = 145) in a large southeastern school district participated in this investigation. This large metropolitan school district subdivided their schools into five district regions. There is variability in student demographics across these district regions, especially with regard to ethnic diversity and SES (using free and reduced lunch status as a proxy family income indicator). Four elementary schools from each of the five district regions were recruited in order to capitalize on the naturally occurring ethnic and SES variability in the different geographical locations. Both high and low-performing schools with respect to students' scores on the previous year's statewide high-stakes assessment were intentionally selected to ensure variability in achievement scores. Of 20 schools invited, 12 principals subsequently agreed to participate in the present study. Each regular education third grade teacher within the participating schools was then individually invited and all subsequently agreed to participate. Students were eligible for participation if they were enrolled in the participating teacher's classroom as a regular education student.

The required sample size to detect a large effect (Cohen's *d* = 0.8) was calculated based on a two-tailed linear multiple regression (random model) with a confidence level of .95 and a statistical power of .80 and 8 predictor variables, indicating that the researchers needed to obtain at least 102 participants (Faul, Erdfelder, Buchner, & Lang, 2009). Acknowledging the potential for a low return rate of consent forms, 32 regular education third grade students were randomly selected from each of the 12 schools using a random numbers chart and a total of 384 consent packets as approved by our university's institutional review board were sent home with students in their backpacks. Of the 384 informed consent packets distributed, 192 consent forms (or 50.3 %) were returned. Of those received, 186 parents/legal guardians consented (96.9% of consent forms returned), 6 parents/legal guardians declined participation (3.1%). Five participants were no longer enrolled in the participating school at the conclusion of the study (2.7% attrition). The demographic composition of the final sample is summarized in Table 1. Socio-economic status (SES) was characterized in this study via a dichotomous variable: *economically disadvantaged* (i.e., students receiving either free or reduced lunch price benefits) and *non-economically disadvantaged* (i.e., students that did not apply or were ineligible for free or reduced lunch price benefits).

Only students for whom reading skill performance data (ORF & decoding/Nonsense Word Fluency (NWF)) were available were retained in the final analysis, resulting in 145 cases for analysis. It was determined that the loss in sample size and concomitant loss in power in eliminating cases with missing data was preferable over imputing those values. Thus, the multiple regression results are based on data from 145 participants.

Instruments

Instruments used in the present study included a year-end high-stakes measure of reading progress for grade 3 (*Florida Comprehensive Assessment Test; FCAT*), the ORF (oral reading fluency) and NWF (nonsense word fluency) subtests from *Dynamic Indicators of Basic Early Literacy Skills* (*DIBELS*) assessment system (Good & Kaminski, 2002), and a brief survey administered to each participant's teacher to obtain the participants' third quarter reading grade. ORF and NWF subtests were used in unaltered form from the *DIBELS* assessment system (Good & Kaminski, 2002) and administered and scored following the standardized administration and scoring procedures provided for the instrument. Technical adequacy of ORF is reported above; information regarding NWF and *FCAT* is described below. The remaining data (e.g., demographics, attendance) were obtained via query to the district's student database records.

NWF is a decoding task whereby the student reads aloud a series of vowel-consonant or consonant-vowel-consonant nonsense words. This subtest assesses the student's ability to blend phonemes, requiring both knowledge of letter-sound correspondences and articulation skill. First grade January NWF scores appear to possess strong predictive validity for end-of-first-grade ORF scores (.82) (Good & Kaminski, 2002). Predictive validity appears weaker for end-of-second-grade ORF scores (.60) and for the Woodcock-Johnson Psycho-Educational Battery (Woodcock, McGrew, & Mather, 2001) Total Reading Cluster score (.66). The instrument's authors did not intend for the NWF subtest to be administered to third grade students and, therefore, there is currently no data to examine the reliability, validity, and predictive utility for this grade level. Nonetheless, as discussed, we were specifically interested in including a measure of decoding given that it is a requisite skill for overall reading achievement of new words, especially for struggling readers in third grade. For this study, the second grade benchmark NWF probes were used intact with no modifications.

Student scores from the *FCAT* Reading subtest were used as a general measure of reading achievement consisting of 50 to 55 multiple choice questions at the time this study was conducted. Students were provided informational (subject-matter centered) or literary (fiction, nonfiction, poetry, or drama) text passages and asked to answer questions to assess students' ability to construct meaning from the texts. Scores on the *FCAT* are reported in terms of scaled scores (range 100-500) and achievement level (range 1-5) (Florida Department of Education (FDOE), 2001, 2004). The parallel forms reliability for the *FCAT* was above .90 for grades 4, 5, 8, and 10 (FDOE, 2001) and correlations between the *FCAT* and SAT-9 two measures ranged from .70 to .81 (FDOE, 2001). The mean Reading *FCAT* for third grade standard curriculum students (non-ESE students) was 317.22 (sd = 56.97) for the year in which this study was conducted. Reliability as measured by Cronbach's alpha was strong at .89 for this administration of the *FCAT*.

Procedures

ORF and NWF subtests were administered within a two-week interval in early December, approximately 14-16 weeks prior to the springtime high-stakes assessment of reading achievement. Volunteer school psychologists and school-based reading coaches administered the subtests, all of whom had received a minimum of six hours of formal in-service training in the administration and scoring of the selected *DIBELS* subtests. Each participant was read a scripted assent form prior to administration.

Twenty percent (n = 36) of the protocols from both subtests were randomly selected for reliability checks by the lead author. Results of the reliability checks are as follows: NWF = .72; ORF = 1.00. Errors were noted in the scoring of NWF, including addition errors, neglect of reporting the maximum correct number of phonemes per line, and omission of completion time if under 1 minute. The lead author re-scored each NWF protocol and NWF protocols that did not note completion (8.8%; n = 16 of 181 students tested) were deemed spoiled and eliminated from analysis.

Sex, SES, attendance rate, and number of prior grade retentions were retrieved from the school district's database. Student attendance rate was obtained by dividing the number of days the student was enrolled by the number of days the student was present for the academic year. The sample median attendance rate was .97 (IQR = .039). With regard to grade retention, of the 145 students used in the regression analysis, 37 students (25.5%) had been retained at least once. Of those retained, 10 (6.9%) were retained in Kindergarten, 12 (8.3%) in first grade, 8 (5.5%) in second grade, and 33 (22.8%) in third grade. Twenty-seven of those students had been retained once, 10 students retained twice. An additional 16 students were retained at the conclusion of the study (14 of whom failed the *FCAT*).

Teachers were provided a questionnaire on which to report each participant's third quarter reading grade with self-addressed stamped envelopes provided for return. Of those distributed, 28.2% of the questionnaires were not returned. The school district database only retained the final reading grade for the academic year, deleting the 9-week quarter grades from the database. Therefore, in cases where the third quarter grade was unavailable, the final reading grade was used.

The purpose of this study was: 1) to replicate earlier studies using ORF to predict reading achievement among third grade students, while controlling for student demographics (economic disadvantage and sex); 2) to test whether the inclusion of a measure of student decoding would help to improve the prediction of reading achievement; and 3) to test whether the inclusion of additional student-level variables known to be implicated in overall school achievement—student's number of prior grade retentions, attendance rate, and reading grade—improve the prediction model. While there are several possible avenues of analysis one could use to explore these questions, hierarchical regression was utilized to better understand the individual and additive effects of each predictor variable or variable set.

When interpreting the results from hierarchical regression analyses, the order of entry of variables into the model should be based on sound empirical or theoretical reasoning (Keith, 2006). While several alternatives exist, the following order was used to address the stated purposes of this study. SES and sex were entered in the first block as control variables to control for the effects of these demographics on achievement. ORF was then entered second into the model to determine its effect on reading achievement when controlling for the aforementioned student demographics (replication of prior studies). NWF was entered third in the model to test the added predictive value of decoding on the reading achievement test, above and beyond that of ORF. The remaining student level variables (retentions, attendance rate, and reading grade) were then entered into the fourth and final block to explore the whether the inclusion of these additional student-level variables would increase the proportion of explained variance in the prediction of reading achievement scores above and beyond demographics, ORF, and decoding skill.

Results

The inter-correlation matrix of predictors is provided in Table 2 with associated tests of significance of the relationships between variables using $\alpha = .01$. Significant correlations were found between the *FCAT* reading measure and ORF, NWF, SES, number of prior grade retentions (retentions), and reading grades. Of interest, the significant negative correlation between the reading *FCAT* score and retentions indicated that students who were retained one or more times performed significantly poorer on the outcome reading measure. With regard to student demographics, SES was significantly correlated with ORF, NWF, retentions, and reading grade indicating that students with economic disadvantage were significantly more likely to perform worse on ORF and NWF measures, had been retained at least once, and had poorer reading grades than the group of students that were categorized as not economically disadvantaged. Sex was not significantly correlated with any other variables included in the model. ORF was significantly positively correlated with NWF and reading grades and significantly negatively correlated with retentions. Similarly, NWF was significantly positively correlated with reading grades.

Hierarchical Regression Results

A case analysis was conducted to evaluate the presence of potential outliers exerting excessive influence on the regression results. One outlier was identified; however, a subsequent sensitivity study revealed that the outlier was not exerting excessive influence on the model R^2 (change in R^2 = .011). Thus, the observation was retained and the reported results reflect the inclusion of all participant data (n=145). The model was run with all variables, retaining the studentized model residuals. A visual inspection of the scatter plot of the studentized model residuals versus predicted Y values revealed no indications of any violations of correct fit of a linear model, constant variance, or normality assumptions required for the legitimacy of the regression results.

Hierarchical linear regression results are provided in Table 3. In summary, the addition of the demographic controls into the first block revealed that only SES was significantly predictive of Reading *FCAT* scores, $\Delta R^2 = .192$, $F(2, 142) = 16.92$, $p < .001$. The addition of ORF into the second block confirmed the significance of the relationship between ORF on Reading *FCAT* scores even when controlling for student demographics (SES and sex), $\Delta R^2 = .334$, $F(1, 141) = 52.26$, $p < .001$. The effect of SES in the second block remained statistically significant ($p < .001$). The addition of the third block revealed no significant effect of NWF, above and beyond ORF and student demographics, $\Delta R^2 = .003$, $F(1, 140) = 39.39$, $p < .35$, while the significance of SES and ORF remain unchanged. In the fourth and final block, the addition of the remaining student-level variables (retentions, attendance rate, and reading grades) significantly increased the explained variance in *FCAT* Reading, $\Delta R^2 = .102$, $F(3, 137) = 33.47$, $p < .001$, with reading grade offering the only significant unique contribution in this step ($p < .001$). In the final model SES was not significant (albeit marginally) while the significant effect of ORF observed in earlier blocks remained significant. Overall, results indicated that a multivariate model of student level predictors of reading achievement was robust and an improvement over a model that included ORF and student demographics in isolation. The final model with all predictor variables resulted in a model $R^2 = 0.631$ (Adjusted $R^2 = 0.612$), a fairly large coefficient of determination, indicating that approximately 63% of the total variability of reading achievement scores could be explained using this model.

Discussion

The relevance of this paper rests in the use of regression to identify additional student-level factors, including word decoding skill, academic history, and demographics that may contribute to success on a comprehensive statewide third grade reading achievement test above and beyond ORF. The authors hypothesized that the inclusion of additional student-level variables would improve the overall model, increasing the proportion of variance in reading achievement. This hypothesis was supported. In the final model, the effects of ORF and reading grades were significant in predicting year-end reading achievement, significant in spite of including student economic disadvantage and prior grade retentions into the prediction model—two factors often implicated in poor reading achievement. Thus, these results are encouraging given that it points to factors that can be addressed with instruction and may help ameliorate learning or performance deficits associated with disadvantage and grade retention issues.

As such, ORF continues to be an important factor in predicting reading achievement and continued focus on students' ability to read fluently appears warranted. Armbruster, Lehr, and Osborn (2001), report that the use of frequent oral reading monitored by a teacher or parent (coined "repeated readings") is an effective activity for improving reading fluency and overall reading achievement. Results herein support continued use of interventions that would target reading fluency, perhaps using locally-derived or state-derived benchmarks rather than national benchmarks for ORF to predict year-end assessment success may be more useful (Brown, 2008). In their 2001 study, Crawford, Tindal and Stieber found that "a reading rate of 119 words per minute virtually ensured that a student passed the [Oregon] statewide reading test" (p. 319). This equated to 94% of their sample. In this study, 93% of students that read 113 correct words per minute on ORF subsequently achieved a passing score on the state's year-end reading assessment. It is yet unclear how often individual teachers are establishing and using local benchmarks.

A measure of decoding was specifically added to this study with the hypothesis that decoding remains an important skill for success on high stakes year end reading assessments, which typically include higher-level decodable words. Of note, NWF is not typically administered in third grade and benchmarks are unavailable for this period (Good & Kaminski, 2002). Thus, the inclusion of NWF in this study with third graders was exploratory; however, the hypothesis for its importance was not supported by the data. A post-hoc analysis revealed that if entered first, NWF was a significant predictor (as would be expected from the correlation matrix), but its effect was negated as soon as ORF entered the model. The insignificance of NWF could be contributed to the high correlation between ORF and NWF such that NWF did not add any unique contribution. Nonetheless, NWF simply did not appear to be an important factor independent of the effect of ORF in third grade.

In contrast, the finding that reading grades did uniquely contribute to the prediction of reading achievement above and beyond ORF, and in the context of all of the other variables in the model was unexpected. The predictive utility of teachers' assigned reading grades has not been widely discussed in the literature on predicting reading achievement. It is conceivable that students' reading fluency skill in general contributed, at least in part, to the letter grades assigned to students for reading/language arts; however, by putting ORF in the hierarchical regression analysis first we were then able to explore how grades *added* to that predictive power. It is indeed likely that participating teachers at different schools (or even within a school) may use different criteria to determine a student's reading grade (e.g., may include data on students' participation, work completion, vocabulary, and spelling tests). Nonetheless, ORF alone was not as predictive of reading achievement as was a model that included reading grades.

A *post-hoc* analysis revealed that approximately 4% of the participants in this study who earned a third-quarter reading grade of an A failed the Reading *FCAT*. Moreover, approximately 8% of those who earned a B failed, 44% who earned a C failed, 67% who earned a D failed, and 78% who earned an F failed the state's year end assessment of reading achievement. Perhaps in teachers' constructions of reading grades the teachers are picking up on something above that of reading fluency which is contributing to overall reading achievement. Previous research regarding accuracy of teachers' assessments of reading skill indicates that teachers are good judges of a variety of reading skills. Feinberg and Shapiro (2003) noted that students' assessed oral reading fluency skill was highly correlated with teachers' predictions of oral reading fluency rate ($r = .70$). Additionally, Bates and Nettelbeck (2001) examined the accuracy of teacher judgments in reading achievement among students with and without classroom behavioral problems. Results of this study indicated that teachers remained accurate judges of reading accuracy ($r = .77$) and reading comprehension ($r = .62$), despite the presence of classroom behavior problems. While students with behavior problems tended to perform more poorly on the reading measures, the teachers did not underestimate reading skill (Bates & Nettelbeck, 2001).

In sum, results of this study: 1) support previous findings of the predictive value of ORF, even when controlling for student demographics; 2) do not support the use of a decoding measure to improve the prediction of reading achievement; and 3) highlight the unique influence reading grades on the prediction of reading achievement. From a cost-benefit analysis perspective, over-identifying students as needing additional intervention may be preferable than under-identifying under-achieving students (Roehrig et al., 2008). Glover and Albers (2007) discuss both pros and cons to over and under-identification (e.g., increased burden on programming resources), but agree that under-identification is a greater risk when the consequences are more high-stakes, such as in year-end achievement testing.

Limitations

Limitations in this study were rooted primarily in the lack of consenting participants, missing data, and subtest administration adherence. Although the return rate of consent packets sent home to parents was consistent with average return rates for mailed or sent-home documentation, the final sample may still represent systematic bias toward families who are possibly more involved in their children's education or more conscientious in completing requested documentation. While results may not be universally generalizable to all third grade students, the random sampling method was a strength in this study.

Missing data also posed some difficulty for this study; teacher survey return rates were not as high as expected, resulting in missing third quarter reading grades for several students. To compensate, the final reading grades were used to replace missing values as described above. Additionally, of 181 consenting participants, ORF and NWF data were collected for 145 participants with absenteeism as the most common cause of missing assessment data. While a variety of strategies and statistical techniques are available to researchers, each with pros and cons (see Baraldi and Enders, 2010), we elected to retain the 145 cases that had both ORF and NWF data, accepting the minimal loss in statistical power. Nonetheless, using the 145-case subset still could have created a biased subset of the original 181 cases and is offered as a limitation of this study.

Subtest administration error for NWF was mildly problematic in the present study (8.8% of protocols with errors). The NWF subtest is used less frequently than ORF in this district and is reported by some testers as more difficult to administer and score given that exact pronunciation of individual phonemes is required for score credit. Perhaps including protocol for inter-observer reliability checks during administration would have been helpful in pinpointing the specific source of the problems associated with that measure.

Lastly, with 24 of the third grade students in the sample repeating third grade, it is possible that these students had seen the oral reading fluency passages used in this sample at some point prior to this study, potentially affecting the results. However, this is unlikely given that these students (who were already not achieving well academically) were able to decode, comprehend, and/or recall the passages in any great detail that would substantially alter the ORF scores for those passages. Additionally, latter passages of the benchmark assessment probes were used in this study that at that time were not being used by the schools for progress monitoring.

Future Directions

While this study targeted regular education third grade students, it may be of interest to replicate the present study with a larger, more diverse sample, increasing generalizability of results and allowing for additional comparisons within and across grades with the identified predictors. With a larger, more diverse sample, one might also explore the prediction model for specific subgroups. It may be useful to analyze current reading risk models using ORF benchmarks to delineate cutoffs that appropriately identify students at risk for failure across groups, with due caution in interpretation of any group differences. It is plausible that the significant predictor variables for students who are English language learners (Wiley & Deno, 2005; Yeo, 2010) or students with Specific Learning Disabilities may differ than those found to be significant in the present study with regular education students.

The present study was focused on Reading *FCAT* achievement. Previous research by Buck and Torgesen (2003) examined the correlation between ORF and Math *FCAT* achievement as well and found a significant positive correlation between the two (r = .54, $p < .001$). Similarly, the predictor variables in the present study could be applied to predict math achievement on year end measures of achievement. In lieu of ORF, using silent curriculum-based measures of reading such as maze measures may also prove useful in predicting state assessments of math achievement (Jiban & Deno, 2007). Lastly, it would be interesting to further dissect reading grades such that we can better understand how teacher evaluations map onto reading skills that are important for grade level assessments of reading achievement. This area appears to be ripe for further research.

References

Armbruster, B. B., Lehr, F., & Osborn, J. (2001). *Put reading first: The research building blocks for teaching children to read*. Washington, DC: Partnership for Reading.

Baker, S., Gersten, R., & Grossen, B. (2002). Interventions for students with reading comprehension problems. In M. R. Shinn, H. M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 731-754). Bethesda, MD: National Association of School Psychologists.

Baraldi, A. N. & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48,* 5-37. doi:10.1016/j.jsp.2009.10.001

Bates, C. & Nettelbeck, T. (2001). Primary school teachers' judgments of reading achievement. *Educational Psychology, 21,* 177-187. doi:10.1080/01443410020043878

Bishop, A. G. & League M. B. (2006). Identifying a multivariate screening model to predict reading difficulties at the onset of Kindergarten: A longitudinal analysis. *Learning Disability Quarterly, 29,* 235-252.

Brown, S. E. (2008). Comparing methodologies used to predict outcomes on high stakes tests with curriculum-based measurement. *Dissertation Abstracts International Section A: Humanities and Social Sciences.*

Buck, J. & Torgesen, J. (2003). *The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test* (Technical Report No. 1). Tallahassee, FL: Florida Center for Reading Research and Florida State University. Retrieved from http://www.fcrr.org/technicalreports/TechnicalReport1.pdf

Burke, M. D., Hagan-Burke, S., Kwok, O., & Parker, R. (2009). Predictive validity of early literacy indicators from the middle of Kindergarten to second grade. *The Journal of Special Education, 42,* 209-226. doi:10.1177/0022466907313347

Burts, D. C., Hart, C. H., Charlesworth, R., & DeWolf, D. M. (1993). Developmental appropriateness of kindergarten programs and academic outcomes in first grade. *Journal of Research in Childhood Education, 8*, (1) 23-31.

Connor, C. M., Morrison, F. J., & Petrella, J. N. (2004). Effective reading comprehension instruction: Examining child by instruction interactions. *Journal of Educational Psychology, 96*(4), 682-698. doi:10.1037/0022-0663.96.4.682

Crawford, L., Tindal, G., & Stieber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment, 7,* 303-323. doi:10.1207/S15326977EA0704_04

Crowe, E. C., Connor, C. M. & Petscher, Y. (2009). Examining the core: Relations among reading curricula, poverty, and first through third grade reading achievement. *Journal of School Psychology, 47,* 187-214. doi:10.1016/j.jsp.2009.02.002

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219-232.

Deno, S. L. (1989). Curriculum-based measurement and special education services: A fundamental and direct relationship. In M. R. Shinn (Ed.), *Curriculum-based* measurement: Assessing special children (pp. 1-17). NY: The Guilford Press.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160. doi: 10.3758/BRM.41.4.1149

Feinberg, A. B. & Shapiro E. S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. *School Psychology Quarterly, 18,* 52-65. doi: 10.1521/scpq.18.1.52.20876

Florida Department of Education (2001). *FCAT briefing book*. Retrieved August 13, 2003, from http://www.firn.edu/doe/sas/*FCAT*/pdf/*FCAT*_brief.pdf.

Florida Department of Education (2004). *Assessment and accountability briefing book*. Retrieved January 16, 2006, from http://www.firn.edu/doe/sas/*FCAT*/pdf/*FCAT*aabb.pdf

Fuchs, L. S. (1995). Best practices in defining student goals and outcomes. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology III* (pp. 539-546). Washington, DC: The National Association of School Psychologists.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9,* 20-28. doi: 10.1177/074193258800900206

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology, 45,* 117-135. doi:10.1016/j.jsp.2006.05.005

Good, R. H., & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp.61-88). NY: The Guilford Press.

Good, R. H., & Kaminski, R. A. (2002). *Dynamic Indicators of Basic Early Literacy Skills (DIBELS) 6th Edition.* Eugene, OR: Institute for Development of Educational Achievement. (Available at http://*DIBELS*.uoregon.edu/)

Hiebert, E. H. (2002). Standards, assessments, and text difficulty. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (3rd ed.) (pp. 337- 369). Newark, DE: International Reading Association.

Hintze, J. M., Callahan, J. E., III, Matthews, W. J., & Williams, S. A. S. (2002). Oral reading fluency and prediction of reading comprehension in African-American and Caucasian elementary school children. *School Psychology Review, 31,* 540-553.

Hixson, M. D. & McGlinchey, M. T. (2004). The relationship between race, income, and oral reading fluency and performance on two reading comprehension measures. *Journal of Psychoeducational Assessment, 22,* 351-364. doi: 10.1177/073428290402200405

Hosp, M. K. & Fuchs, L. S. (2005). Using CBM as an indicator of decoding, word-reading, and comprehension: do the relations change with grade? *School Psychology Review, 34*(1), 9-26.

Hosp, J. L., Hosp, M. A., & Dole, J. K. (2011). Potential bias in predictive validity of universal screening measures across disaggregation subgroups. *School Psychology Review, 40*(1), 108-131.

Jiban, C. L. & Deno., S. L. (2007). Using math and reading curriculum-based measurements to predict state mathematics test performance. *Assessment for Effective Intervention, 32*(2), 78-89. doi: 10.1177/15345084070320020501

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review, 30,* 420-437.

Jimerson, S., Carlson, E., Rotert, M., Egeland, B., & Sroufe, L. A. (1997). A prospective, longitudinal study of the correlates and consequences of early grade retention. *Journal of School Psychology, 35*(1), 3-25. doi:10.1016/S0022-4405(96)00033-7

Keith, T. Z. (2006). *Multiple regression and beyond.* Boston, MA: Allyn & Bacon.

Klein, J. R., & Jimerson, S. R. (2005). Examining ethnic, gender, language, and socioeconomic bias in oral reading fluency scores among Caucasian and Hispanic students. *School Psychology Quarterly, 20,*(1), 23-50. doi:10.1521/scpq.20.1.23.64196

Kranzler, J. H., Brownell, M. T., & Miller, M. D. (1998). The construct validity of curriculum-based measurement of reading: An empirical test of a plausible rival hypothesis. *Journal of School Psychology, 36,* 399-415. doi:10.1016/S0022-4405(98)00018-1

Kranzler, J. H., Miller, M. D., & Jordan, L. (1999). An examination of racial/ethnic and gender bias on curriculum-based measurement of reading. *School Psychology Quarterly 14*(3), 327-342. doi:10.1037/h0089012

Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18-78). NY: The Guilford Press.

McCoy, A. R., & Reynolds, A. J. (1999). Grade retention and school performance: An extended investigation. *Journal of School Psychology, 37*(3), 273-298.

Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesen, J. K. (2008). Accuracy of the *DIBELS Oral* Reading Fluency measure on predicting third grade reading comprehension outcomes. *Journal of School Psychology, 46,* 343-366. doi:10.1016/j.jsp.2007.06.006

Samuels, S. J. (2002). Reading fluency: Its development and assessment. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (3rd ed.) (pp. 166-183). Newark, DE: International Reading Association.

Schilling, S. G., Carlisle, J. F., Scott, S. E., & Zeng, J. (2007). Are fluency measures accurate predictors of reading achievement? *The Elementary School Journal, 107*(5), 429-448. doi:10.1086/51862

Shapiro, E. S., Keller, M. A., Lutz, L. G., Santoro, L. E., Hintze, J. M. (2006). Curriculum-based measures and performance on state assessment and standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment, 24*(1), 19-35. doi: 10.1177/0734282905285237

Shinn, M. R. (1995). Best practices in curriculum-based measurement and its use in a problem solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology III* (pp. 547-567). Washington, DC: The National Association of School Psychologists.

Shinn, M. R., Shinn, M. M., Hamilton, C., & Clarke, B. (2002). Using curriculum-based measurement in general education classrooms to promote reading success. In M. R. Shinn, H. M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 113-142). Bethesda, MD: National Association of School Psychologists.

Silberglitt, B., Burns, M. K., Madyun, N. H., & Lail, K. E. (2006). Relationship of reading fluency assessment data with state accountability test scores: a longitudinal comparison of grade levels. *Psychology in the Schools, 43,* 527-536. doi:10.1002/pits.20175

Sindelar, P. T., Lane, H. B., Pullen, P. C., & Hudson, R. F. (2002). Remedial Interventions for Students with Reading Decoding Problems. In M. R. Shinn, H. M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 703-729). Bethesda, MD: National Association of School Psychologists.

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

Stecker, P. M., & Fuchs, L. S. (2000). Effecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disabilities Research & Practice, 15,* 128-134.

Wanzek, J., Roberts, G., Linan-Thompson, S., Vaughn, S., Woodruff, A. L., & Murray, C. S. (2010). Differences in the relationship of oral reading fluency and high-stakes measures of reading comprehension. *Assessment for Effective Instruction, 35,* (2) 67-77. doi: 10.1177/1534508409339917

Wiley H. I. & Deno, S. L. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education, 26,* (4) 207-214. doi:10.1177/07419325050260040301

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside Publishing.

Yeo, S. (2010). Predicting performance on state achievement tests using curriculum-based measurement in reading: A multilevel meta-analysis. *Remedial and Special Education, 31,* 412-422. doi:10.1177/0741932508327463

Table 1

Descriptive Statistics

Variable	*n*	%
Sex		
Male	55	37.9
Female	90	62.1
Ethnicity		
Caucasian	51	35.2
African-American	87	60.0
Hispanic	3	2.1
Asian	2	1.4
Other	2	1.3
Socio-economic status (SES)		
Disadvantaged	88	60.7
Non-disadvantaged	57	39.3
Number of Retentions		
0	108	74.5
1	27	18.6
2	10	6.9
Reading Grade		
A	23	15.9
B	51	35.2
C	41	28.3
D	23	15.9
F	7	4.8

Note. Reading grade statistics incorporate replaced values for missing third quarter data.

Table 2
Descriptive Statistics and Inter-correlation Matrix

	FCAT	SES	Sex	ORF	NWF	Retentions	Attendance	Reading Grade
FCAT	—							
SES	-.43*	—						
Sex	-.07	.02	—					
ORF	.69*	-.32*	-.07	—				
NWF	.50*	-.30*	.10	.63*	—			
Retentions	-.36*	.25*	.17	-.29*	-.19	—		
Attendance	.04	-.07	-.03	.00	.06	.01	—	
Reading Grade	.65*	-.36*	-.05	.49*	.42*	-.28*	.05	—
M	307.82	.61	.38	102.61	81.55	.32	.96	2.41
SD	52.43	.49	.49	33.54	41.98	.60	.04	1.08
n	145	145	145	145	145	145	145	145
Range	134 - 446	—	—	0 - 180	6 - 232	0 - 2	.68 – 1.0	0 - 4

Note. Correlations for FCAT, ORF, NWF, Retentions, Reading grade, and attendance are the Pearson product-moment correlation. Point estimates for the dichotomous variables of Sex and SES are the contrast of means between the two groups. Variables are coded as follows: Sex: female=0 and male=1; SES: non-disadvantaged=0 and disadvantaged=1; Reading Grade: F=0, D=1, C=2, B=3, A=4.
* $p < .01$.

Table 3

Hierarchical Linear Regression Results Predicting Reading Achievement

	Model 1			Model 2			Model 3			Model 4		
Variable	B	β	*p*	B	β	*p*	B	β	*p*	B	β	*p*
Step 1:												
SES	-46.34	-.43	<.001	-25.71	-.24	<.001	-24.85	-.23	<.001	-15.22	-.14	.014
Sex	-6.66	-.06	.414	-2.49	-.02	.691	-3.60	-.03	.573	-.62	-.01	.915
Step 2:												
ORF				.96	.61	<.001	.89	.57	<.001	.67	.43	<.001
Step 3:												
NWF							.09	.07	.349	.02	.02	.784
Step 4:												
Retentions										-8.21	-.09	.098
Attendance										5.81	.005	.924
Reading Grade										17.10	.35	<.001
R^2		.19			.53			.53			.63	
Change in R^2					.33			.00			.10	
F		16.92***			52.26***			39.39***			33.47***	
Change in F					99.48***			.88			12.57***	

Note. The ΔR^2 result is the increase in R^2 due to adding each block incrementally, given the other predictor variables. Variables are coded as follows: Sex: female=0 and male=1; SES: non-disadvantaged=0 and disadvantaged=1; Reading Grade: F=0, D=1, C=2, B=3, A=4.

* $p < .05$. ** $p < .01$. *** $p < .001$