Published in final edited form as:

Read Psychol. 2012 ; 33(1-2): . doi:10.1080/02702711.2012.630613.

The Relationship of Error Rate and Comprehension in Second and Third Grade Oral Reading Fluency

Mary Abbott, PhD, Howard Wills, PhD, Angela Miller, PhD, and Journ Kaufman, BA
Juniper Gardens Children's Project University of Kansas

Abstract

This study explored the relationships of oral reading speed and error rate on comprehension with second and third grade students with identified reading risk. The study included 920 2nd graders and 974 3rd graders. Participants were assessed using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) and the Woodcock Reading Mastery Test (WRMT) Passage Comprehension subtest. Results from this study further illuminate the significant relationships between error rate, oral reading fluency, and reading comprehension performance, and grade-specific guidelines for appropriate error rate levels. Low oral reading fluency and high error rates predict the level of passage comprehension performance. For second grade students below benchmark, a fall assessment error rate of 28% predicts that student comprehension performance will be below average. For third grade students below benchmark, the fall assessment cut point is 14%. Instructional implications of the findings are discussed.

According to the National Reading Panel (2000), only 10% to 15% of the students with the most serious reading problems graduate from high school and only 2% complete a four-year college program. For most of these students, reading failure begins early in their educational careers. For example, Stevenson and Newman (1986) reported a .52 correlation between kindergarteners' ability to name the letters of the alphabet and their 10th grade performance on a standardized test of reading comprehension.

In order to address the academic deficits, it has become common for school districts across the country to implement a Response-to-Intervention (RtI) instructional framework that seeks to prevent academic failure through early intervention, frequent progress monitoring, and increased intensive research-based instructional interventions (Fuchs & Fuchs, 2006). A critical RtI component is an assessment-driven methodology for moving students into and out of intervention and special services.

Fuchs, Fuchs, and Compton (2004) report that for students in the primary elementary school years, a curriculum based measure (CBM), which is a quick direct assessment of academic skills that have been taught, may be the best way to measure student responsiveness to instructional intervention. CBM assessments are based on the research work of Deno, the Fuchs, and others (see Deno, 2003 for review) and have a specific set of criteria. CBM assessments (1) become reliable and valid through repeated sampling such as progress monitoring, (2) are a measurable task with a system to count the number of correct and incorrect responses within a fixed-time, (3) have established administration and scoring procedures, (4) are short 1-3 minute assessments, and (5) have easy-to-learn procedures that ensure reliability.

CBM oral text reading fluency assessments have become a common methodology for measuring reading growth. Oral reading fluency has three measurable elements: speed, accuracy, and prosody (Wolf & Katzir-Cohen, 2001). Reading speed and accuracy can be measured as correct words per minute (CWPM). This is the actual timed correct rate at which an individual reads and is measured as a raw score of total words read in a minute, minus the number of errors. Error rate can also be defined as a percentage of errors read. Percent of error is calculated by dividing the raw score of word reading error into the total number of words read. Prosody is the individual's appropriate use of expression, intonation, and rhythm during oral text reading.

Research indicates that oral reading fluency is an accurate developmental indicator of the complex set of skills that document proficient reading (Fuchs, Fuchs, Hosp & Jenkins, 2001). Pikulski and Chard (2005) describe their deep view of fluency development as a continuum of early experiences with print, letters, and sounds that develop the foundation for automatic word recognition. Theoretically, perfecting word recognition skills creates a reciprocal, causal relationship with reading comprehension by lowering the demands on attentional and cognitive word reading processing and allowing more mental resources to be devoted to comprehension (e.g., LaBerge & Samuels, 1974; Perfetti, 1985; Posner & Snyder, 1975). This theoretical framework provides a basis for hypothesizing that the components of oral reading fluency may serve as an indicator of the extent to which lower-level processes such as word identification are automated and the higher-level processing of comprehension is achieved (Fuchs, et al, 2001). A growing number of research-based studies of students in elementary education indicate that CWPM is predictive of comprehension performance. Several of these studies report a significant relationship between CWPM and the prediction of reading comprehension (Buck & Torgesen, 2003; Riedel, 2007; Schilling, Carlisle, Scott, & Zeng, 2007).

Although the use of CBM based on CWPM performance may work well as a quick snapshot of reading progress, for a vast majority of students, students achieving at the lowest reading levels experience a persistent challenge in exhibiting CWPM gains (Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003; O'Connor, Notari-Syverson, & Vadasy, 1998; Torgesen & Hudson, 2006; Uhry & Shepherd, 1997). Fuchs, Fuchs, and Hamlett (1993) report that even when strong reading interventions improve word recognition, the CWPM fluency slopes of students with reading disabilities are less steep, and improve in smaller increments than for students in the average range. Consistently across studies, students who fall behind their peers on CWPM measures, for the most part, remain behind their peers.

This consistent and long-term lag creates a challenge for education professionals who use CWPM as the common measure across all students to determine reading performance and to use RtI as a system for making instructional intervention decisions for students. For the lowest performing students, CWPM fails to provide a sensitive enough tool to measure reading growth or response to intervention, and it also fails to provide information about what instructional intervention might be the most appropriate. A sensitive method at quickly and repeatedly assesses reading growth and progress for students with low reading performance continues to challenge researchers and practitioners (McMaster, Fuchs, Fuchs, & Compton, 2005).

The purpose of this article is to explore the relationship between CWPM oral reading rate, error rate, and comprehension within the context of growth and change in reading behaviors, and determine if the use of a CBM error rate measure is a sensitive measure of reading growth for students with the lowest reading achievement.

Using CBM to Measure Oral Reading Fluency

One CBM that has been used extensively in school districts across the country is the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). DIBELS is a set of individual easy-to-administer 1-minute fluency measures and is used to regularly monitor the development of early reading skills over time through repeated administration of appropriate grade-based reading passages (e.g., Kameenui, et al., 2001). DIBELS Oral Reading Fluency (ORF) is an oral text reading assessment that begins at the middle of first grade. ORF utilizes a pre-determined set of grade-leveled passages. At each assessment occasion, a CWPM *cut score* or *benchmark* is used to determine if sufficient reading growth has occurred (see Table 1). Failure to meet the CWPM cut point or *benchmark* indicates that the student is at risk for reading failure and a candidate for reading intervention.

There is a growing base of research that links DIBELS' CWPM scores to state reading performance assessments and standardized reading assessments, and, therefore, suggests that use of the DIBELS as a quick oral reading CBM would be an appropriate measure for determining reading growth. For example, Buck and Torgesen (2003) report a significant positive relationship between third graders' performances on DIBELS CWPM scores and reading achievement on the Florida Comprehensive Assessment Test-Sunshine State Standards (FCAT-SSS). In addition, Riedel (2007) reports that the CWPM scores of Memphis students significantly predict comprehension performance on the TerraNova standardized assessment (CTB/McGraw-Hill, 2003). In another study with second and third grade Michigan students, Schilling, et al. (2007) report the failure to meet DIBELS CWPM *benchmark* strongly predicts students who performed poorly on the end-of-the-year Iowa Test of Basic Skills (ITBS) comprehension subtest. Finally, Hosp and Fuchs (2005) found evidence that grade level CWPM cut scores for reading were predictive of achievement on the Woodcock Reading Mastery Test (WRMT) for overall reading mastery. In a study with students in the first through fourth grades, results reveal a strong relationship between the reading CWPM and WRMT scores at each grade level.

Although there is evidence that the DIBELS' CWPM raw score provides a quick oral reading CBM measure of reading growth when all students are collapsed in a sample, no published work exists that evaluates the use of DIBELS error rate score to measure reading growth with a disaggregated student sample. Further, there are no published studies that link DIBELS error rate to performance on a comprehension measure exists. Error rate percentage data can easily be collected during an ORF assessment and, in fact, since 2005, DIBELS has included error rate raw scores as part of ORF data collection. Research about the widely-used ORF DIBELS is needed to extend the current research understanding about how error rate scores predict comprehension for students at different levels of risk.

Measuring Oral Reading Error and its Effect on Comprehension

Prior studies about reading errors have focused on a variety of aspects about how error rate affects comprehension. For example, Anderson, Wilkinson, Mason, Shirey, and Wilson (1988) examined the impact of teacher correction of errors during oral reading. They reported that, for children with low error rates, when an oral reading error was followed by teacher correction there was improved learning and memory with good comprehension. However, for students with much greater than average error percentages, the benefit of teacher feedback failed to result in good comprehension. The authors make the case that error rate is a good measure of children's level of reading development and, consistent with other studies, error rate correlates negatively with end-of-year achievement. In another study, Stahl reported that when sufficient guidance and feedback were provided to 3rd graders with reading challenges, comprehension became disrupted at a 15% or greater error

rate (Kuhn & Stahl, 2003; Stahl & Heubach, 2005). Our work in fluency and error rate indicate that average error rates differ by grade level (see Table 2). It is important to clearly understand the relationship between error rate and comprehension within the context of grade level performance.

Over the last four decades much of the work on error rate and comprehension has focused around the model and procedures of miscue analysis with the premise that a less than 10% error rate is appropriate at every grade level (Goodman, Watson & Burke, 2005). During miscue analysis, in order to assess how the reader uses syntax, semantic information, and graphophonics cuing strategies, the teacher closely observes, records, and analyzes oral reading behaviors. Each of the errors is coded in terms of the importance of how each error type disrupts meaning. The goal is to detect how these three cueing systems are used in concert to identify words. Miscue analysis data are collected in as natural a condition as possible, with readers orally reading authentic and complete stories to which they have never been exposed. Different forms of miscue analysis have included pattern analysis with informal reading inventory passages, collection, and analysis of oral reading samples during reading instruction, and detailed information about the use of miscue analysis published in the Reading Miscue Inventory (Goodman & Burke, 1972).

A summary of the qualitative miscue analysis research suggests that the reader who is proficient has the ability to recognize when his or her errors or miscues need correcting; therefore meaning is not disrupted. A reader who fails to be proficient lacks understanding of how to correct miscues; therefore meaning is disrupted (Goodman, 1969). However, the component of the miscue analysis model that promotes use of contextual cueing has been disproven in numerous research studies (e.g., Pollatsek, Rayner, & Lee, 2000). According to McKenna and Picard (2006), benefits of miscue analysis lay not in the model's promotion of use of context clues but in guiding teachers to select appropriate text for instruction, evidence of a student's inadequate decoding skills, and student movement away from using context clues toward decoding skills.

As a CBM assessment for RtI, miscue analysis has reliability challenges in that the use of different reading passages with differing text levels makes comparisons between passage assessments difficult. In addition, the process of having students read an entire passage, and then answer comprehension questions about the passage is time consuming, and therefore, makes the process of progress monitoring more challenging. Further, miscue analysis procedures are complicated and extensively rely on teacher judgment to classify errors. Such procedures create challenges for reliable results across the wide range of individuals collecting data. Finally, there is concern that the use of a 10% error rate as a cut point at every grade level fails to take into account the differences of average error rates at different grade levels (see Table 2). What is needed for an RtI system is a tool that is quick and easy to administer with countable components across similarly complex passages that takes into account the average grade level error percentages.

A challenge with the DIBELS assessments are that the CWPM score alone may fail to distinguish what kind of the intervention might be the most appropriate. For example, a student with a low CWPM and high error rate is likely to have insufficient vocabulary understanding to support good reading comprehension. For these students, interventions should focus on error reduction during text reading. However, for students with a low/below *benchmark* CWPM rate and low error rate, the intervention should focus on specific strategies that quicken the pace of reading. In other words, although the CWPM identifies students who may be at risk for reading failure, it may not provide sufficient information for instructional decisions within an RtI framework. These findings add critical information to the use of RtI in diagnosing reading risk as well as for providing essential instructional

intervention information. Such a study has theoretical as well as practical implications for creating a structure of delivery of reading interventions that facilitate the greatest academic gains. The current study explores the relationships of CWPM, percentage of error rate, and comprehension with second and third grade students with varied reading risk by examining three research questions:

1. Generally, what is the relationship between error rate and reading comprehension on literacy assessments?
2. Generally, is the relationship between error rate and reading comprehension moderated by CWPM performance?
3. Specifically, what is the relationship between error rate and CWPM for below *benchmark* students at risk of reading failure? What is the point at which low CWPM and high error impede comprehension?

Method

Participants

Participants were part of a K-3 five-year Center project that was federally funded through the U.S. Office of Special Education Programs. This project was conducted in 12 elementary schools in a large Midwestern metropolitan area. Of the 12 schools, nine schools were public suburban, one public urban, one urban parochial, and one urban charter. The demographics of the school participants were representative of a diverse population that included Caucasian, African-American, Latino, and Asian students within the metropolitan community. The Center used a randomized experimental/comparison group design to evaluate a preventative intervention model in which evidenced-based curricula were systematically implemented with students at-risk of reading failure as defined by DIBELS' *benchmarks* (Abbott, Wills, Kamps, Greenwood, Kaufman, & Filingim, 2008; Kamps, Abbott, Greenwood, Wills, & Veerkamp, 2008). In the experimental condition, data-driven decision-making was a focal point of the preventative implementation model. Data were used to place and change intervention assignment. All schools participating in the project received reading data student outcome data.

For this analysis, a subset of students from the larger K-3 data base was used. The subsample for this study includes the 920 second grade students and 974 third grade students. Of these students, 47% were minority and 48% received free or reduced lunches. Of the overall sample, 88% of the students were from the suburban schools, 5% from the public urban school, 7% from the urban parochial, and 10% from the urban charter. Because interventions are not central to the questions being asked within this study, participants included all of the second and third graders in both the experimental, and comparison conditions schools. We believe that the academic student performance of this subsample across experimental and comparison schools mirrors typical academic performance found in districts across the country.

Each of the students had been identified on at least one DIBELS assessment occasion (fall, winter, or spring) as below *benchmark* at either the *strategic* or *intensive* level of risk of reading failure. According to prior DIBELS research, the term *benchmark* is used to describe students who have a .80 probability or greater of achieving subsequent reading fluency goals. *Benchmark* range scores suggest appropriate reading progress. Students in the *strategic* range have a .50 probability of reaching the next reading goal, and exhibit *some risk* that recommends intervention. Students in the *intensive* range have a .20 probability of meeting the next reading goal. They are most *at risk* and need intervention. (Good, Simmons, & Kameenui, 2001).

Students chosen for this sample had the following data available for analysis: grade appropriate fall, winter, and spring DIBELS CWPM raw score data points and a yearly WRMT Grade-Based Standard Score on the passage comprehension subtest. Data for the analysis were longitudinally collected over three years. The *first* assessment occasion that a student with parental permission was identified as falling in the *strategic* or *intensive* range, the WRMT was administered. Then in the spring of each school year, the WRMT data continued to be collected. It is important to note that WRMT passage comprehension data collection continued even if, at any point, students' CWPM scores moved out of the *strategic* and *intensive* range into the DIBELS *benchmark* range. Also important to note is that *benchmark* portion of the sample for this analysis is comprised only of students that at some time during the fall, winter or spring occasions of their K-3rd grade year were identified as *strategic* or *intensive*. For example, within the larger K-3 study, if at the winter occasion of the kindergarten year, a student was classified as being in the *strategic* range and had parent permission, a WRMT was given. If throughout the second grade year this same student was performing in the *benchmark* range, the WRMT was still collected and that student is part of the analysis for this study. However, any student from the larger K-3 study who performed at *benchmark* at every CWPM data point was not chosen for this study because only students that were at some point in the *strategic* or *intensive* range were given the WRMT.

Assessment Instruments and Procedures

As part of the federal government funding agreement, the Center collected DIBELS grade level appropriate assessments and the WRMT reading battery (Good, Simmons, & Kameenui, 2001; Woodcock, 1988). DIBELS second and third grade CWPM and error rate raw scores were collected with students in the fall, winter, and spring of the school year. The goals and cutoff scores are criterion-referenced and empirically derived (Good, Kaminski, Smith, Simmons, Kameenui, & Wallin, 2002). Alternative reliability range from .90-.96. Concurrent validity with the TCWPM range from .90-.96 (Good, Gruba, & Kaminski, 2002; Good, Simmons, & Smith, 1998; Hintze, Ryan, & Stoner, 2003). The DIBELS *benchmark* criterion for the end of second grade CWPM rate is 90 and by the end of third grade it is 110 (see Table 1).

The WRMT is a standardized, individually administered, untimed test. The split-half reliability coefficients ranged from 0.86 to 0.99, and its concurrent ranged from 0.78 to 0.92 (Woodcock, 1988). For the purpose of this study, raw score data were converted to Grade-Based Standard Scores ($M = 100$; $SD = 15$). Because the focus of this study was the effect of error rate on comprehension, the WRMT Passage Comprehension subtest was used for analysis.

Trained assessors used the standardized administration and scoring guidelines for the CWPM subtest and WRMT assessments. Assessors were required to pass a 90% reliability criterion on at least two trials with a trained observer on both assessments. The standardized scoring procedures for the oral text reading for DIBELS were the same procedures for second and third grade passages. Students were asked to read a passage aloud and were told that they would receive help if they were to get "stuck" on a word. The students were timed for one minute. If the student did not respond to a word within three seconds, the assessor provided the word and prompted the student to continue reading. At each testing occasion, students read three passages. The middle score passage of the three was selected and recorded. Student errors were also recorded. Errors were defined as omissions, substitutions, and mispronunciations of words within the passage. Error rate was calculated for each passage by calculating percent of words correct (i.e., accuracy). For example, if a student

read 50 words in one minute with five errors, the CWPM would be 45 raw score and the error rate would be 10%.

During the WRMT, the assessor asked students to read a passage silently and then orally respond with the missing word that completes the sentence. Although the administration of the assessment is untimed, students who failed to answer with the missing word were encouraged to provide an answer. If the student was unable to provide the missing word, the assessor had the child move to the next passage. The assessment was discontinued when the student failed to provide a sufficient number of correct answers.

Results

Descriptive Statistics and Zero-Order Correlations

Descriptive statistics and correlations at the student level among all variables used in the analyses are presented in Tables 2 and 3. Means, standard deviations, and correlations are presented for each classification group (*intensive*, *strategic*, and *benchmark*) based on their status at the beginning of the academic year. Comparisons between *benchmark* and *strategic* groups as well as between *benchmark* and *intensive* groups were conducted using z-tests to determine if the magnitude of the relationships were statistically different. These descriptive statistics are presented for both grades two and three.

Data Analysis Plan

Hierarchical linear modeling (HLM) was used because of the nested structure of the data consisting of students nested within schools. A two-level model was used to analyze the data; however, no variables were entered on the school level (level two) of the model. The research questions of interest pertain only to student level (level one) variables. This type of model appropriately controls for the nesting of students within schools and accounts for those shared experiences of students within schools, thus more accurately estimating the standard errors of the model.

Separate models were estimated for each grade level as different students were in each sample. For each sample, first an unconditional model was examined determining the amount of variability in the dependent variable, WRMT passage comprehension, due to school level influences. Then the predictor variables, centered DIBELS CWPM (fall) and centered error rate (fall) were added to the model examining their influence on WRMT passage comprehension. These predictors were centered at their means to reduce collinearity with the interaction term created by multiplying these two variables. Next, this interaction term between DIBELS CWPM and errors was entered into the model to examine the hypothesized moderated relationship. Coefficients were estimated using the PROC MIXED routine in SAS with maximum likelihood estimation.

Receiver Operating Characteristic (ROC) analysis was used to further examine the relationship between DIBELS error rate and WRMT passage comprehension for a sub sample group identified in the regression analysis. ROC analysis is used to predict a dichotomous outcome (below or at/above grade level) and utilizes four possible results. The first possible result is that a high error rate predicts below grade level performance and the student does perform below grade level on the WRMT passage comprehension text (true positive). Second, DIBELS high error rate predicts below grade level performance and the student performs at or above grade level (false positive). Third, DIBELS high error rate predicts at/above grade level performance and student performs at /above grade level (true negative). Last, DIBELS high error rate predicts at/above grade level performance and student performs below grade level (false negative). These four possibilities are used to

define sensitivity, which is the proportion of below grade level performance cases that correctly predicted (true positive / (true positive + false negative), and specificity, which is the proportion of the at/above grade level performance cases correctly predicted (true negative / (true negative + false negative).

In the ROC analysis, a curve is made by graphing sensitivity on the y-axis and 1-specificity on the x-axis. The area under this curve is then calculated; an area near .50 would indicate poor accuracy (equal to chance) and an area of 1 would indicate perfect predictive accuracy. Outcomes reported for these analyses include the sensitivity and specificity at the cut point and the area under the curve as well as the percentage of students correctly identified in our sample. The area under the curve and percentage correctly identified are both indicators of the predictive ability of the selected cut point error rate. All ROC analyses were conducted using SPSS software.

Preliminary Analyses

Fully unconditional HLM models were estimated in preliminary analyses. These random-effects models enable estimation of the between school and within school variances. The analyses determine if there are between schools differences on the outcome variable of WRMT passage comprehension. Although children were also nested within classrooms within the schools, examination of the intraclass correlation (ICC) of unconditional models revealed that there was not a significant amount of variation between classrooms (ICC grade 2 = .005; ICC grade 3 = .02) while there was a significant amount of variation between schools. This is not surprising because the subject area of reading is highly individualized at these early stages and teacher practices were fairly consistent within a school; however reading instruction practices did vary across schools. There was significant between school variation for both grade two, $X^2(13) = 131.65$, $p < .001$ and grade three, $X^2(13) = 51.95$, $p < .001$. The adjusted intraclass correlations were 19% and 10 % respectively.

Primary Analyses

Two multilevel regression analyses were conducted, one for each grade level. Spring WRMT passage comprehension at grade-level two and three was the dependent variable in each model and fall measures of DIBELS CWPM and error rate were the independent variables. Both DIBELS CWPM and error rate were centered at their respective means. In each model the first set of equations included both centered predictors without the interaction terms. In both models, higher DIBELS CWPM was positively associated with WRMT passage comprehension scores, $\gamma_{\text{grade2}} = .14$, $p < .001$; $\gamma_{\text{grade3}} = .12$, $p < .001$ and error rate was negatively related to WRMT passage comprehension, $\gamma_{\text{grade2}} = -33.08$, $p < .001$; $\gamma_{\text{grade3}} = -43.16$, $p < .001$. The second set of equations included the interaction term computed between correct word per minute (CWPM) and error rate (ER). Table 4 includes the coefficients for both the grade two and grade three models which included the interaction terms testing for moderation. Calculations of variance explained using the reduction in variance method (Raudenbush & Byrk, 2002) were 51% for the grade 2 model and 53% for the grade 3 model. A significant CWPM X ER interaction emerged, $\gamma_{\text{grade2}} = .55$, $p < .001$; $\gamma_{\text{grade3}} = .37$, $p < .001$. Figure 1 depicts the predicted WRMT passage comprehension scores as a function of both DIBELS error rate and CWPM for both grade levels. Graphical displays of the interaction effect were plotted using the pick-a-point method choosing values at the mid points of each classification appropriate for the given grade level (e.g. grade two entering DIBELS *intensive* classification ranges from 0-25, the midpoint of 13 was selected to use as a plot point for the graph). For the error rate points we selected points at one standard deviation above and below the sample mean for each grade level as representative of high/low error rate.

We investigated the specific nature of the interaction using a simple slope analysis that tested the significance of the DIBELS error rate at values of DIBELS CWPM at the mid points of each classification. These tests of simple slopes for the grade 2 sample indicated that for students beginning the academic year with higher level CWPM (functioning at the *benchmark* level) there is no interaction with error rate, estimate = -1.73 , $p = .80$. However, for those students functioning at lower levels of DIBELS CWPM there was a significant simple slope, estimate = -15.59 , $p < .001$ for *strategic* and estimate = -27.79 , $p < .001$ for *intensive*. This indicates that those students with lower DIBELS CWPM combined with higher error rates are predicted to score lower on the WRMT passage comprehension test than lower DIBELS CWPM students with lower error rates. This same pattern also emerged in the grade three sample with a simple slope estimate = -4.74 , $p = .66$ for the *benchmark* students and significant simple slopes for the *strategic*, estimate = -28.57 , $p < .001$ and the *intensive*, estimate = -40.31 , $p < .001$.

Further investigation of the interaction term was carried out by calculating the regions of significance for each sample. The region of significance analysis determines the values of the moderator at which the slope of the regression of WRMT passage comprehension on DIBELS error rate transitions from non-significance to significance. Upper and lower bounds of the region were calculated. In the grade-two sample, the lower bound was found to occur at a DIBELS CWPM score of 45 and in the grade three sample the lower bound was found at a DIBELS CWPM score of 87. The upper bound scores in both samples were beyond the typical range of scores, and therefore they are not reported. In other words, the region of significance indicated a 'cut point' of where the effect of CWPM on error rate is significant. This region of significance occurs for those students with lower CWPM, specifically those with scores at or below 45 in the fall of grade two or below 87 in the fall of grade three.

In order to further examine these at risk readers identified by the regions of significance tests on the interaction effect from the regression analyses (and corresponding to those classified in the *intensive* and *strategic* categories by the DIBELS), we chose to use a Receiver Operating Characteristic (ROC) analysis. This analysis was used to examine the relationship between the DIBELS error rate and WRMT passage comprehension in more detail for the *intensive* and *strategic* groups of students. The *benchmark* categorization was not examined because this group was outside of the region of significance for the interaction analysis examined in the previous analysis. Fall error rate was again used to predict spring WRMT passage comprehension as in the previous analysis; the purpose of this follow up analysis was to attempt to determine appropriate cut scores of error rate. In other words, how much error is too much error? For this analysis the WRMT spring passage comprehension score was dichotomized to those students who were performing 'at grade level or higher' and 'below grade level'. Our definition, while subjective, is based on a score of 1 standard deviation below the mean standard score. Therefore, a student scoring at a standard score of 85 or higher was judged to be 'at grade level or higher'. We were interested in using the fall error rate score to predict which students would score below grade level on the WRMT spring passage comprehension test.

Sensitivity and specificity were examined in conjunction to choose a cut point to define high error rate. For the purposes of this analysis, sensitivity refers to the proportion of below grade level cases correctly predicted by fall error rate, and specificity refers to the proportion of students at or above grade level correctly predicted by the fall error rate. There is a trade-off between sensitivity and specificity, increasing one will decrease the other and vice versa. Our goal in determining cut scores was to find a balance between sensitivity and specificity. We chose the point where the difference between sensitivity and specificity was the smallest and where sensitivity was the higher value. In terms of identifying students to target for

reducing error rate it is better to err on the side of over identifying students (identifying more that might not benefit as much from reducing error rate) rather than under identifying students (missing students that would potentially benefit from targeted reduced error rate).

In Table 5, we report the selected cut score and the associated sensitivity/specificity area under the curve as well as the percentage of students in our sample correctly classified as either below or at/above grade level. We first examined the entire at risk group of students (all of those students falling below the cut point identified in the regression analyses), this sample included both those students classified as *strategic* and *intensive* by the DIBELS. We determined a cut score on error rate of 28%. This means that if we selected students performing in the *intensive* or *strategic* level of DIBELS CWPM who also had an error rate of 28% or higher, we could predict their spring WRMT passage comprehension score with a sensitivity of .81 and a specificity of .78. Recall that sensitivity means that we would correctly identify 81% of the students who would perform below grade level and specificity means that we would incorrectly identify 22% of student who, although they were selected because of high error rate and low DIBELS CWPM, they would still manage to perform at or above grade level. We examined our sample using this cut point of 28% and we correctly placed 85% of our sample (this included both correctly identified by low DIBELS CWPM/high error rate who performed below grade level and those identified by low DIBELS CWPM/low error rate who performed at or above grade level). The area under the curve for this group is .87 which is close to 1 (representing perfect predictability) and significantly different from .50 (chance). Both the area under the curve and the percentage of correct classification indicate good predictability using this cut point on error rate. We extended these analyses to examining just the most at risk readers, those classified as *intensive* by the DIBELS, and determined a slightly high cut point, 34%, for this smaller more extreme sample. Sensitivity and specificity are also comparable in the larger at risk sample. Table 5 includes parallel analyses for the grade three sample.

Discussion

The purpose of the study has been to explore the relationship of DIBELS CWPM and error rate to WRMT passage comprehension and to examine how CWPM moderates error rate in relation to reading passage comprehension. In addition, this study extends the analysis to determine to what extent error rate can be used in conjunction with DIBELS CWPM to predict passage comprehension on the WRMT. We sought to assess if the predictive validity of the DIBELS' error rate can be more informative than DIBELS CWPM for the *intensive* group of readers. The relationship between the components of CWPM and comprehension are critical variables in reading instruction as well as in reading intervention evaluation of RtI. The results from this study indicate evidence to support a strong relationship between DIBELS error rate and WRMT passage comprehension and between DIBELS error rate and CWPM for students in the DIBELS-defined *intensive* reading group. Specifically a major finding in this study is that students at moderate and high risk for reading problems may have similar CWPM scores but based on their error rates would be predicted to perform quite differently on a measure of reading comprehension on the WRMT.

There are no surprises in the descriptive analysis that addresses research question one about the relationship between error rate and comprehension. Table 2 shows the mean and standard deviation results for all measures across both second and third grade. For both second and third grade students, the DIBELS CWPM/WRMT passage comprehension profiles of students at the *benchmark*, *strategic*, and *intensive* levels are different. Students at *benchmark* and above have the lowest error rates, the highest CWPM, and highest Grade-Based Standard Score on WRMT passage comprehension. In contrast, for both second and third grade *intensive* groups, students exhibiting the highest error rates, the lowest levels of

DIBELS CWPM, the lowest Grade-Based Standard Scores on WRMT passage comprehension, and the greatest within group variability on all measures.

More information about research question one is explained in the correlational data of the general relationship between DIBELS error rate and CWPM and the WRMT passage comprehension. For both second and third grade students at the *intensive* level there are strong correlations between WRMT passage comprehension and DIBELS CWPM (.68 to .73) and error rate (−.72. to −.74) while correlations are in the moderate range for the students at *strategic* and *benchmark* levels. There is a pattern of slightly stronger correlations between DIBELS error rate and WRMT passage comprehension than between WRMT passage comprehension and DIBELS CWPM. When comparing these results with other studies that looked at the predictability of DIBELS CWPM and state assessments, the *intensive* group has stronger correlations and the *strategic* and *benchmark* groups indicate a weaker overall relationship. For example, Riedel (2007) reports a .57 correlation between second grade DIBELS CWPM and the TerraNova. Buck and Torgesen (2003) report a significant correlation between DIBELS CWPM scores and the reading Florida state assessment ($r=.70$, $p<.001$), and Good, Simmons, and Kameenui (2001) found a .67 correlation between DIBELS CWPM scores and the Oregon Statewide Assessment ($r=.67$, $p<.001$). Our combined results across all student groups provided similar correlations. We believe that the differences in our results are due to the desegregation of the sample by DIBELS guidelines. Some differences may also be seen because state assessments measure global reading ability and our analysis focused only on the WRMT passage comprehension. The important point to note is the strong relationships at the *intensive* range of scores.

In order to address question two about the relationship between DIBELS error rate and WRMT passage comprehension moderated by DIBELS CWPM, we conducted a multilevel regression analysis. For both grades two and three, there is a significant negative relationship of error rate to passage comprehension. An additional interaction effect analysis between DIBELS CWPM and error rate on comprehension was conducted. There is a significant interaction effect with both second and third grade samples for both the *strategic* and the *intensive* groups. What this analysis indicates is that the interaction between DIBELS CWPM and error rates predicts the level of the WRMT passage comprehension. Students in the *strategic* and *intensive* groups with the combination of low DIBELS CWPM and high error are predicted to perform at lower levels on the WRMT passage comprehension subtest than students with low DIBELS CWPM and low error rates. Further analysis finds the specific point of the fall DIBELS CWPM occasion at which that significance occurs. For grade two, the regions of significance occurs below the DIBELS CWPM value of 44 and for third grade the regions of significance appears below the value of 87. That is, when a fall second grader has a CWPM score of 44 or less, their error rate predicts comprehension. Likewise, when a fall third grader has a CWPM score of 87 or less, their error rate predicts comprehension.

It is fascinating that these cut points are in line with the fall DIBELS *benchmark* CWPM cut points of 44 for second grade and 77 for third grade because it enables us to draw conclusions about error rate and comprehension performance with the DIBELS risk categories. As we might expect, comprehension performance is likely to be poor for students in *strategic* and *intensive* range with low DIBELS CWPM and with high error rates. What might not be expected is that when students in the *strategic* and *intensive* range have low DIBELS CWPM with low error rates, comprehension performance is likely to be in the average range. In the analysis of this data set, it is the DIBELS CWPM that is the predictive moderating factor between error rate and WRMT comprehension performance. When DIBELS CWPM is high as in the *benchmark*, DIBELS error rate loses its predictive ability. This is an important interaction effect finding because although the CWPM

classifications of *benchmark* and below *benchmark* (*strategic* and *intensive*) generally predict future reading success, once a student is in the below *benchmark* range, error rate becomes the best predictor of comprehension. Within an RtI system, practitioner knowledge about error rate becomes critical not only for crafting the most effective intervention but also in determining the response to that intervention.

Additional analyses allow us to answer question three that addresses the relationship between DIBELS error rate and CWPM for below *benchmark* readers, and the point at which the combination of low DIBELS CWPM and high error preclude average-level WRMT passage comprehension. The ROC analysis is another important contribution of our study because it pinpoints the DIBELS error rate within risk classification (*strategic/intensive*) at which error is too high to support average range WRMT passage comprehension. Those error rate cut points differ dependent on the grade level and risk classification (see Table 5).

At each grade level, we had two different group combinations. Combined *strategic* and *intensive* is one group and *intensive* only is the second group. We did this for three reasons. First we wanted to learn if the error rate cut points would be different between second and third graders. Second, we wanted to learn if cut scores would be different for the total risk population as opposed to the *intensive* only population. Third we wanted to learn the ability of DIBELS error rate to specifically predict WRMT passage comprehension performance for each of these groups.

Generally, we found that the second grade error rate cut points were higher than the third grade error rate cut points. This is what we would expect to happen. It is normal for first graders who are in the early beginning reading stage to have high error rates and low CWPM. That error falls off quickly by the beginning of second grade for *benchmark* readers. That drop in error rate is followed by a large jump in CWPM during the second grade year (44% in fall to 90% by spring). It makes sense that second graders who are in the *strategic* and *intensive* groups will lag behind normal development and exhibit higher error rates during their second grade year.

The cut scores for second grade are 28% for the *strategic/intensive* group and 34% for the *intensive* group. These are the error rate percentages at which error has become too large to support average range comprehension. We are confident that we can predict this result for the second grade *strategic/intensive* group 85% of the time and for the *intensive* only group, 78% of the time. The cut scores for third grade are 14% for the *strategic/intensive* group and 18% for the *intensive* group. These are the error rate percentages at which error has become too large to support average range comprehension. For the second grade *strategic/intensive* group, we can predict this result 80% of the time and for the *intensive* only group, we can predict this result 78% of the time.

To summarize, for fall second and third grade students in the DIBELS CWPM below *benchmark* range of *strategic* and *intensive*, it is the DIBELS' error rate that best predicts WRMT passage comprehension. Further, for fall second grade students who have below *benchmark* DIBELS CWPM, an error rate of 28% or greater predicts that students will no longer be able to sustain average range comprehension. Further, for fall third grade students who have below *benchmark* DIBELS CWPM, an error rate of 14% or greater predicts that students will no longer be able to sustain average range comprehension. Instructional implications are discussed below.

Study Limitations

There are several limitations to the study. The first is that the data are embedded in another study, and, therefore, we were unable to add other measures such as different measures for reading comprehension. Additionally, while this study includes a range of schools and students, a broader national sample would add to the quality of our sampling and to the generalizability of our findings. Also, students from both the experimental and comparison groups were grouped together for this analysis. Although we do not view this as relevant to the specific questions of the study and believe that the academic results of the subsample mirror typical performance that would be found in districts that promote the use of RtI, it is possible that the scores for students in intervention modified the means of the subsample.

Finally, because every student in the sample for this study was at some point in the larger study classified by the DIBELS CWPM in the *strategic* or *intensive* range, the sample does not include students from the larger study that began and remained in the CWPM *benchmark* range throughout second and third grade. It may be that the exclusion of this *benchmark* population influenced the mean DIBELS error rate and CWPM rate, and correlational data reported for the *benchmark* group. However, the focus of the study reported within this article was about students who exhibit high error rates, which are primarily students in the *strategic* or *intensive* ranges. Spring error rates were 2% for second graders and 1% for third graders at *benchmark* or above in our smaller sample. Therefore, we believe that the exclusion of the large numbers of students from the original data base that began and remained at CWPM *benchmark* levels is negligible.

Instructional Implications

Based on the results of our study, we organize our instructional implications under three different topics. The first area is identification of reading risk. The second is the use of data results to measure instructional progress or RtI, and the third is skills instruction related to components of fluency instruction.

Identification of reading risk and response to intervention—The DIBELS CWPM measure as CWPM provides a predictive measure for future reading success. CWPM offers a global instructional indicator of the need for intervention, a ballpark estimate of the strength of intervention, and an on-going progress monitoring tool to check academic progress. However, it is the error rate that further defines and delineates the level of reading challenge and suggests a roadmap of determining instructional content, the skills to be improved, and potentially provides a more sensitive measure for response to treatment and intervention.

The important RtI message is that it is inappropriate to conclude that a student has failed to *respond* to intervention based on their DIBELS CWPM score alone. Improvement from high error rate to a lower error rate should also be taken into account when making judgments about the success of intervention. A continuous drop in error rate can represent large changes in reading progress.

Skills instruction related to components of fluency instruction—What our results suggest is that for students below DIBELS *benchmark*, in addition to DIBELS CWPM, error rate should be taken into account. For below *benchmark* second graders who have error rates at 28% or more there should be a concentrated instructional effort to lower error rate by focusing on interventions that remediate poor word recognition and provide comprehension strategies. For second graders, when the error rate is below 28%, work on word recognition and comprehension skills is still important. However, working to increase speed should also become part of the intervention. Torgesen, et al. (2003) suggests that in addition to word

recognition factors that hamper fluent reading, slow readers also lack understanding of the balance between speed and accuracy and have a lack of awareness about how to drive oneself to read more quickly. These skills should be systematically taught when error rates are in the range at which they do not severely impact comprehension but do impact CWPM.

For third grade students below DIBELS *benchmark*, the DIBELS' error rate cut point is 14%. At 14% or above, word recognition and comprehension skills should be the focus. For error rates below 14%, instruction about how to read faster with strategies such as repeated reading should become the focus.

For students with large error rates, a miscue analysis might be appropriate. Attention to the errors a child makes provides a straight forward instructional road map for the skills-instructional focus. For example, if most student errors are with high frequency words, then instruction on the 1,000 most frequent words is warranted. The 1,000 most frequent English words represent 90% of the words present in elementary school texts (Carroll, Davies, & Richman, 1971). By teaching these specific 1,000 words to automatic recognition of one second or less for each word, there is an increased chance for student errors to drop and for the speed of reading and comprehension to improve.

If a student misreads multi-syllabic words, skills instruction on syllable division and reading is needed. This is especially true for content area words that require work on word identification as well as word meaning. A strong focus on teaching accuracy of word recognition across as many different contexts as possible is recommended. One of the most successful fluency interventions is repeated reading (Meyer & Felton, 1999; Samuels, 2002). It is effective because it provides repeated exposure to new words as well as improves efficiency to access of known words. Repeated reading practice concentrates on the exposure to limited number of specific words over a relatively short span of time.

Most importantly, for students with high DIBELS error rates simply providing more reading opportunities during regular reading instruction may not be sufficient intensity and duration of instruction to increase their word recognition skills to an acceptable acquisition rate. Students need the opportunity to master early reading skills that are incrementally introduced through small group or individualized *intensive* instruction with an additional "double dose" of time spent daily in the specific intervention (Kamps, et al., 2008; Torgesen, et al., 2006; Vaughn & Linan-Thompson, 2003). Through assessment and ongoing progress monitoring using DIBELS error rate as a critical indicator of growth, educators gain valuable information with which to guide intervention and measure students' growth over time.

References

- Abbott, M.; Wills, HP.; Kamps, D.; Greenwood, CR.; Kaufman, J.; Filingim, D. The process of implementing a reading and behavior three-tier model: A case study in a midwest elementary school. In: Greenwood, CR.; Horner, CR,R.; Kratochwill, T.; Oxaal, I., editors. *Elementary School-Wide Prevention Models: Real Models and Real Lessons Learned*. Guilford Press; NY, NY: 2008. p. 215-265.
- Anderson RC, Wilkinson I, Mason JM, Shirey L, Wilson PT. Do errors on classroom reading tasks slow growth in reading? *The Elementary School Journal*. 1988; 88(3):266–280.
- Buck, J.; Torgesen, J. The Relationship Between Performance on a Measure of Oral Reading Fluency and Performance on the Florida Comprehensive Assessment Test. (FCRR Technical Report #1). Florida Center for Reading Research; Tallahassee, FL: 2003.
- Carroll, JP.; Davies, P.; Richman, B. *Word frequency book*. Houghton Mifflin Co.; Boston: 1971.
- Deno SL. Developments in curriculum based measurement. *The Journal of Special Education*. 2003; 37(3):184–192.

- Fuchs D, Fuchs LS. Introduction to responsiveness to intervention: What, why, and how valid is it? *Reading Research Quarterly*. 2006; 41(1):93–99.
- Fuchs D, Fuchs LS, Compton DL. Identifying reading disabilities by responsiveness-to-instruction: Specifying measures and criteria. *Learning Disability Quarterly*. 2004; 27(4):216–227.
- Fuchs LS, Fuchs D, Hosp MK, Jenkins JR. Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*. 2001; 5(3):239–256.
- Fuchs LS, Fuchs D, Hamlett CL. Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review*. 1993; 22(1):27–48.
- Good, RH.; Gruba, J.; Kaminski, RA. Best practices in using dynamic indicators of basic early literacy skills (DIBELS) in an outcomes-driven model.. In: Thomas, A.; Grimes, J., editors. *Best practices in school psychology*. 4th ed.. National Association of School Psychologists; Washington DC: 2002. p. 699-720.
- Good, RH.; Kaminski, RA.; Smith, S.; Simmons, D.; Kameenui, E.; Wallin, J. Reviewing outcomes: Using DIBELS to evaluate kindergarten curricula, and interventions.. In: Vaughn, SR.; Briggs, KL., editors. *Reading in the classroom: Systems for the observation of teaching and learning*. Brookes; Baltimore: 2002. p. 221-259.
- Good RH, Simmons DC, Kameenui EJ. The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading*. 2001; 5(3):257–288.
- Good R, Simmons D, Smith S. Effective academic interventions in the United States: Evaluating and enhancing the acquisition of early reading skills. *School Psychology Review*. 1998; 27(1):45–56.
- Goodman, K. Analysis of oral reading miscues: Applied psycholinguistics.. In: Gollasch, F., editor. *Language and literacy: The selected writings of Kenneth Goodman*. Routledge & Kegan Paul; Boston: 1969. p. 123-134.
- Goodman, YM.; Burke, C. *Reading miscue inventory: Manual for diagnosis and evaluation*. Macmillan; New York: 1972.
- Hasbrouck J, Tindal GA. Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher*. 2006; 59(7):636–644.
- Hintze JM, Ryan AL, Stoner G. Concurrent validity and diagnostic accuracy of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) and the Comprehensive Test of Phonological Processing. *School Psychology Review*. 2003; 32(4):541–556.
- Hosp MK, Fuchs LS. Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade? *School Psychology Review*. 2005; 34(1):9–26.
- Jenkins JR, Fuchs LS, van den Broek P, Espin C, Deno SL. Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology*. 2003; 95(4):719–729.
- Kameenui, EJ.; Simmons, DC.; Good, RH.; Harn, BA. The use of fluency-based measures in early identification and evaluation of intervention efficacy in schools.. In: Wolf, M., editor. *Dyslexia, fluency, and the brain*. York Press, Inc.; MD: 2001. p. 307-331.
- Kamps D, Abbott M, Greenwood CR, Wills HP, Veerkamp M. Effects of small group reading instruction for students who are most at risk in kindergarten: Two year results for secondary and tertiary level intervention. *Journal of Learning Disabilities*. 2008; 41(2):101–114. [PubMed: 18354931]
- Kuhn MR, Stahl SA. Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology*. 2003; 95(1):3–21.
- LaBerge D, Samuels S. Toward a theory of automatic information processing in reading. *Cognitive Psychology*. 1974; 6:293–323.
- McMaster KL, Fuchs D, Fuchs LS, Compton DL. Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional Children*. 2005; 71(4):445–463.
- McKenna MC, Picard MC. Assessment: Revisiting the role of miscue analysis in effective teaching. *The Reading Teacher*. 2006; 60(4):378–380.
- Meyer M, Felton R. Repeated reading to enhance fluency: old approaches and new directions. *Annals of Dyslexia*. 1999; 49:283–306.

- National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. National Institute of Child Health and Human Development; Rockville, MD: 2000.
- O'Connor RE, Notari-Syverson A, Vadasy P. First-grade effects of teacher-led phonological activities in kindergarten for children with mild disabilities: A follow-up study. *Learning Disabilities Research & Practice*. 1998; 13:43–52.
- Perfetti, CA. Reading ability. Oxford University Press; New York: 1985.
- Pollatsek, A.; Rayner, K.; Lee, H. Encoding printed words crucially depends on phonological coding.. In: Kennedy, A.; Radach, R.; Heller, D.; Pynte, J., editors. *Reading as a perceptual process*. Elsevier; Oxford: 2000.
- Posner, MI.; Snyder, CRR. Attention and cognitive control.. In: Solso, R., editor. *Information processing and cognition: The Loyola symposium*. Erlbaum; Hillsdale, NJ: 1975. p. 55-85.
- Pikulski JJ, Chard DJ. Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher*. 2005; 58(6):510–519.
- Raudenbush, SW.; Bryk, AS. Hierarchical linear models: Applications and data analysis methods. 2nd ed. Sage Publications; Thousand Oaks, CA: 2002.
- Riedel BW. The relationship between DIBELS, reading comprehension, and vocabulary in urban first grade students. *Reading Research Quarterly*. 2007; 42(4):546–562.
- Samuels, S. Reading fluency: Its development and assessment.. In: Farstrup, A.; Samuels, S., editors. *What research has to say about reading instruction*. International Reading Association; Newark, DE: 2002. p. 166-183.
- Schilling SG, Carlisle JF, Scott SE, Zeng J. Are fluency measures accurate predictors of reading achievement? *Elementary School Journal*. 2007; 107(5):429–448.
- Stahl SA, Heubach KM. Fluency-oriented reading instruction. *Journal of Literacy Research*. 2005; 37:25–60.
- Stevenson HW, Newman RS. Long-term prediction of achievement and attitudes in mathematics and reading. *Child Development*. 1986; 57:646–659. [PubMed: 3720396]
- Torgeson, JK.; Hudson, R. Reading fluency: critical issues for struggling readers.. In: Samuels, SJ.; Farstrup, A., editors. *What Research Has to Say About Fluency Instruction*. International Reading Association; Newark, DE: 2006.
- Torgesen, JK.; Rashotte, CA.; Alexander, A.; Alexander, J.; MacPhee, K. Progress towards understanding the instructional conditions necessary for remediating reading difficulties in older children.. In: Foorman, B., editor. *Preventing and Remediating Reading Difficulties: Bringing Science to Scale*. York Press; Parkton, MD: 2003. p. 275-298.
- Uhry JK, Shepherd MJ. Teaching phonological recoding to young children with phonological processing deficits: The effect on sight-vocabulary acquisition. *Learning Disability Quarterly*. 1997; 20(2):104–125.
- Vaughn S, Linan-Thompson S. What is special about special education for students with learning disabilities? *Journal of Special Education*. 2003; 37(3):140–147.
- Wolf M, Katzir-Cohen T. Reading fluency and its intervention. *Scientific Studies in Reading*. 2001; 5(3):211–239.
- Woodcock, R. Woodcock reading mastery tests- Revised. Circle Pines, MN: 1988.

Grade 2

Grade 3

Figure 1.

Predicted values for passage comprehension as a function of error rate at high and low levels and CWPM at midpoints of *intensive*, *strategic*, and *benchmark* classifications

Table 1DIBELS second and third grade CWPM *benchmarks*

Risk	Fall	Winter	Spring
Second Grade			
At risk – <i>Intensive</i>	CWPM < 26	CWPM < 52	CWPM < 70
Some risk - <i>Strategic</i>	CWPM 26 - 44	CWPM 52 - 68	CWPM 70 - 90
Low risk - <i>benchmark</i>	CWPM ≥ 44	CWPM ≥ 68	CWPM ≥ 90
Third Grade			
At risk - <i>Intensive</i>	CWPM < 53	CWPM < 67	CWPM < 80
Some risk - <i>Strategic</i>	CWPM 53 - 77	CWPM 67 - 92	CWPM 80 - 110
Low risk - <i>benchmark</i>	CWPM ≥ 77	CWPM ≥ 92	CWPM ≥ 110

Table 2

Means and standard deviations for all measures

	n	CWPM fall	Error Rate fall	Passage Comprehension
Grade 2:				
<i>Intensive</i>	364	18.61 (9.28)	.32 (19)	88.41 (11.04)
<i>Strategic</i>	270	33.37 (8.61)	.16 (.07)	97.79 (9.04)
<i>benchmark</i>	286	63.76 (18.58)	.06 (.05)	104.74 (8.55)
Grade 3:				
<i>Intensive</i>	383	35.54 (12.88)	.18 (.14)	89.92 (10.45)
<i>Strategic</i>	344	64.22 (7.01)	.06 (.05)	99.60 (6.64)
<i>benchmark</i>	247	98.97 (19.47)	.03 (.03)	103.89 (9.63)

Table 3

Correlations with spring passage comprehension

	CWPM fall	CWPM winter	CWPM spring	Error fall	Error winter	Error spring
<i>Intensive</i>						
Grade 2	.69	.70	.73	-.72	-.73	-.74
Grade 3	.68	.71	.68	-.74	-.74	-.72
<i>Strategic</i>						
Grade 2	.32	.47	.49	-.31	-.32	-.50
Grade 3	.26	.23	.22	-.25	-.30	-.21
<i>benchmark</i>						
Grade 2	.29	.33	.34	-.22	-.28	-.28
Grade 3	.37	.33	.33	-.21	-.16	-.25

Note: all correlations are significant at $p < .001$

Table 4**Multilevel Regression Coefficients**

	Grade 2	Grade 3
Fixed Effects	Coefficient	
Intercept	97.82 ^{**} (.80)	97.86 ^{**} (.55)
Correct Word Per Minute (CWPM)	.25 ^{**} (.03)	.17 ^{**} (.02)
Error rate (ER)	-14.48 [*] (4.60)	-24.90 ^{**} (5.67)
CWPM × ER	.55 ^{**} (.12)	.37 ^{**} (.10)
Random Effects	Variance	
Intercept	6.34 [*] (2.81)	2.75 [*] (1.30)

*
p< .05.

**
p< .01.

Table 5

ROC results for fall error rate and spring passage comprehension

	Cut score	Sensitivity	Specificity	Area under curve	Percentage correctly classified
Grade 2:					
<i>Strategic & Intensive</i>	28%	.81	.79	.87 [*]	85%
<i>Intensive</i>	34%	.78	.77	.85 [*]	78%
Grade 3:					
<i>Strategic & Intensive</i>	14%	.85	.79	.89 [*]	80%
<i>Intensive</i>	18%	.79	.78	.87 [*]	78%

*
p< .001 significantly different from .50