

**Establishing Curriculum-Based Measurement Oral Reading Fluency
Performance Standards to Predict Success on Local and State Tests of
Reading Achievement**

Donald Sibley, Deb Biwer, Amy Hesch
Arlington Heights Public School District 25

Author's Notes:

Requests for reprints should be sent to Don Sibley, NCSP, Dryden Elementary School, 722 S. Dryden Pl., Arlington Heights School District 25, Arlington Heights, IL 60005 or sibley@ahsd25.k12.il.us.

The authors are deeply indebted to Dr. Roland Good of the University of Oregon for his generous assistance with the data analysis for this study. Without his support, statistical expertise and encouragement, this study would not have been possible.

Abstract

Curriculum-based assessment (CBA) and curriculum-based measurement (CBM) procedures are increasingly being utilized to assess student academic skills for educational decision-making, including screening, progress monitoring, entitlement and intervention planning. At the same time, school districts are increasingly compelled to administer local and state standardized achievement tests to gather student performance data. Although the broad purpose of each assessment is the same – measure student performance within the instructional curriculum – there is usually very little understanding of the relationship between the assessment processes. With the media attention focused on local achievement test scores, little attention has been given to the role that CBM can have in predicting student performance on state and local achievement tests. Specifically, there have been relatively few attempts to establish CBM benchmarks - performance levels that can predict success on the local standardized assessment. Recent research to establish curriculum-based oral reading fluency benchmarks has been done in the states of Oregon and Alaska. The primary purpose of this study is to analyze data collected in a suburban school district in northeastern Illinois to evaluate the linkage between established CBM reading fluency benchmarks and state and local district standardized reading achievement in grades three through five. A secondary purpose of this study is to examine the predictive validity of CBM oral reading fluency relative to state and local reading assessments. Results of predictive validity studies demonstrate strong predictive validity for CBM Oral Reading Fluency measures relative to student performance on state and local standardized achievement tests. Analysis of the linkage between established oral reading fluency benchmarks and state and local assessment in Illinois demonstrates high utility for oral reading fluency benchmarks established in other states. Finally, this study suggests oral fluency benchmarks for fourth grade reading fluency that are linked to fifth grade “high stakes” achievement test performance.

Establishing Curriculum-Based Measurement Oral Reading Fluency Performance Standards to Predict Success on Local and State Tests of Reading Achievement

Within recent years there has been a marked increase in the amount of time that students spend in various standardized assessment processes. Many, probably most, school districts in the nation today are engaged in “high stakes” assessment, collecting vast amounts of student performance data, primarily through a combination of local and state assessments. These assessments are generally driven by increased efforts at accountability and/or a need to measure student progress relative to the instructional curriculum. Traditionally, local assessment efforts have relied upon a variety of nationally normed achievement tests such as the Iowa Test of Basic Skills, the Stanford Achievement Test and the Comprehensive Test of Basic skills. In addition, most school districts are compelled to administer state-developed achievement tests that measure student achievement (norm-referenced) or attainment of specific learning standards (criterion-referenced). There are a number of problems associated with this approach to assessment. First, there is often a poor match between the local curriculum and the test content creating problems interpreting the results of the testing. Second, this type of assessment is frequently summative in nature; it occurs at the end of an instructional period when it is too late to impact student performance. Finally, as a result of pressures to report scores that reflect improvements in performance, many districts are revamping curriculum content and revising instructional approaches so that students are more likely to perform well. In the most extreme examples teachers are compelled to “teach to the test”.

Although we are now engaged in more assessment than in previous years, there is evidence that more students are struggling with the curriculum skills (Shapiro, 1996). Between 1978 and 1993, the percentage of students placed in learning disabilities programs across the nation increased from 2.3% to 5.2%, while placement rates in other special education programs have remained essentially stable. Ownby et al. (1985) reported that referrals in a small school system for academic skills problems occurred nearly 5 times as often

as referrals for other concerns. A recent report published by the National Center for Educational Statistics showed that current fourth grade reading scores on a 500-point test administered to 8,000 students showed no gains in performance from 1992. Only thirty-two percent of the students tested were rated as “proficient” or better. Private school students fared slightly better than public school students, according to the report. These are sobering statistics that come at a time when reading proficiency is deemed to be a more essential skill than ever. Data such as these require educators to carefully examine the purpose of the assessments in which our students are engaged.

A different view about the purpose of assessment is needed. According to Salvia and Ysseldyke (1995), assessment is defined as “the process of collecting data for the purpose of 1) specifying and verifying problems and 2) making decisions about students.” In other words, assessment should inform instruction, not simply tell us how students are performing. Furthermore, Salvia and Ysseldyke (1995) identified five types of decisions for which assessment data is usually collected: referral, screening, classification and entitlement, instructional planning and progress monitoring. Unfortunately, few, if any, published academic skills assessment procedures are able to inform each of these decisions. Norm-referenced instruments may be useful for making classification or entitlement decisions, but are of little help in developing instructional interventions. Similarly, criterion-referenced instruments may be helpful in establishing targets for instructional intervention but are of very limited utility in monitoring student progress. Although logic suggests that we need a variety of assessment procedures to collect different types of data in order to inform a range of decisions, that does not appear to be part of the reality of assessment. According to (Goh et al., 1981; Ysseldyke et al., 1981), assessment practices do not show a tendency to vary with the referring concern. Clearly, what is needed is an assessment process that produces data that can be used for the full range of decisions that need to be made about students.

Both general and special education personnel are engaged in collecting data to be used in decision-making about students. The general education teacher evaluates classroom data to determine if all students in his/her classroom are learning at an acceptable rate. Within a problem-solving model, if there is evidence of a possible problem, the teacher will collect additional data to determine the nature of the problem. The teacher might also initiate interventions and collect data to evaluate the impact of such interventions. If the student continues to have difficulty, the classroom teacher may request additional assessment. It is at this point that special education personnel may become involved to answer questions regarding the severity of the discrepancy and to determine eligibility/entitlement. There are frequently questions about what data is necessary to inform the range of decisions suggested by this description. Generally, data from group-administered and group-normed instruments is insufficient and, often, inappropriate.

According to Good et al. (2001), assessment at the primary level that provides data for educational decision making and accountability must include the following features. It must (a) document and account for growth on a continuum of skills, (b) predict success or failure on criterion measures of performance (i.e., state and local assessments), and (c) provide appropriate instructional goals that, if met, will prevent reading failure. This approach to assessment is based on the assumption that assessment must not only inform if students are learning, but also if they are learning at a rate that will allow them to attain the desired criteria on “high stakes” tests. In order to accomplish this task, it is necessary to establish critical benchmarks of student performance. As stated by Good et al. (2001), the purpose of a benchmark is to designate a level of performance on one task that demonstrates favorable odds of fulfilling subsequent goals.

Curriculum-based measurement (CBM) is an essential component in the type of assessment described by Good et al (2001). Curriculum Based Measurement (CBM) has been well established as a reliable and valid

measure of academic achievement in basic skills areas – reading, writing, spelling and arithmetic (Shapiro, 1996; Shinn, 1989). CBM presents a number of advantages over traditional norm-referenced tests. Among these are 1) relevance to instruction by assessing performance with the same materials used for instruction in the classroom, 2) focus on repeated measurement to monitor student progress, 3) sensitivity to change and performance over time and 4) reliable feedback on the effectiveness of instructional interventions (Canter & Marston, 1998; Shapiro, 1996). Specifically, CBM performance can be used to inform school personnel which students may or may not have a high probability of attaining desired standards on state and local assessments. In order for this function to be effective, it is necessary to establish meaningful critical performance levels or benchmarks of CBM performance that are linked to the criterion measure. In practice, benchmarks allow us to examine performance on an earlier measure to predict performance on a later measure.

The procedure for developing benchmarks is outlined by Good et al. (2001). According to Good, the first step is to establish an initial estimate of a goal using a combination of good empirical evidence, theoretical rationale and a judgment of social value (p. 12). For example, an adequate level of reading fluency is a skill that is likely to be judged as high in social value. The second step is to examine the utility of the initial goal in different contexts, with different samples of student data and at different times (p. 12). The outcome of the second step may cause the initial goal to be adjusted up or down. Good states that the “anchor” for the system of benchmark goals is a rate of 40 words correct per minute (wcpm) in the spring of first grade. Subsequent oral reading fluency benchmarks are built from this anchor. This goal is based on the commonly held assumption that all first grade students must read at or above 40 words per minute in the spring of first grade if they are to have a high probability of being capable readers by third grade. Good notes that the figure of 40 wcpm is consistent with a reading trajectory that reflects an adequate rate of progress (p. 13). Further support for the standard of 40 wcpm comes from the instructional placement standards developed by Fuchs, Fuchs and Deno (1982) which suggest that students in grades 1 and 2 should read at the rate of 40 to 60 words per minute in order to benefit from teacher-led instruction; third through fifth graders need to read between 70 to 100 words per minute. There are several other features of benchmarks that need to be considered. Benchmarks need to be ambitious while setting a reasonable target. For example, Good (2001) notes that it is reasonable to establish a goal of 90 words correct per minute for all students in second grade. However, it is unreasonable to expect that 100 percent of second grade students will read 90 wcpm. Expecting that 100 percent of students will meet a benchmark places unreasonable expectations on teachers and fails to account for the natural variability that exists in student ability and performance. In order to be meaningful, benchmarks also need to have global rather than local utility. In other words, there is a need for benchmarks that are meaningful in different settings with different sets of students. Finally, effective benchmarks need to predict attainment of critical performance standards with sufficient “lead time” to inform instruction which can alter the outcomes. Good notes that current assessment practices do not serve as a sufficient “early warning” system for those students who are at risk (p. 25).

The purpose of this study is to examine the utility of applying Oral Reading Fluency benchmarks established in other settings (i.e., Oregon and Alaska) to student performance data from a northeast suburban school district in Illinois. First, this study will examine the intercorrelation of oral reading fluency measures utilized by this district. Second, this study will examine the correlation between oral reading fluency and state (ISAT) and local assessment (Level Tests) instruments. Third, this study will examine the degree to which established benchmarks for second and third grade spring CBM Oral Reading Fluency (ORF) are linked to student performance on state and local assessments. Finally, this study will propose fourth grade reading fluency benchmarks that are linked to fifth grade ISAT and Level Reading tests.

Method

Setting and Participants

Because of the administration schedules of the state and local instruments, there were two groups of students used for this study. Participants in the primary group were 112 students currently enrolled in the fifth grade in two elementary schools located in a suburban school district in Northeast Illinois. The data used for this study was collected from fall 1997 through winter 2001. The participants in the secondary group were 114 students currently enrolled in sixth grade in the same district. The participants in both groups included both general education and special education resource students, including several inclusion students. Within this district approximately 15% of the student population receives some type of special education service. The district is predominantly Caucasian, middle to upper-middle class. The minority population is approximately 4%, split somewhat evenly between students of Asian and Hispanic heritage. Approximately 7 percent of the students within the district are eligible for free or reduced-cost lunches. The student population within this district is quite stable; the student transience rate averaged less than 10% for the years during which the data were collected for this study.

Measures

Three measures of student academic performance were used in the present study: (1) curriculum-based oral reading fluency measures, (2) the Illinois Standards Achievement Test (ISAT) in Reading and (3) the Level Test for Reading. CBM Oral Reading Fluency data are reported for grades two through five. ISAT data are reported for grades three and five. Level Reading data are reported for grades three through five.

Curriculum Based Measurement Oral Reading Fluency: Curriculum-based measurement data have been collected in this district since 1995; local norms for reading fluency were established in 1997. All students participating in this study were administered CBM reading fluency probes following the same four year time line in which reading fluency data were collected once in the fall (October) and once in the spring (late March/early April). Two sets of reading probes were used due to a change in reading curriculum. Probes were randomly selected from the 1991 Houghton Mifflin and 1996 Scholastic Literature series following guidelines presented by Shapiro (1999). An equating study, conducted when the district adopted the Scholastic series, showed no significant differences in performance between the probes for the grade levels targeted in this study. The reading probes were administered individually by trained staff following the standardized procedures outlined by Shinn (1989) and Shapiro (1999). Three 1-minute reading probes were administered to each student; the scores used for this study are the median score from each triad of probes. Probes were scored by trained staff following standardized scoring guidelines in which reading probes were scored for the number of words read correctly per minute. Readers are referred to Shinn (1989) and/or Shapiro (1999) for specific details regarding CBM administration and scoring procedures.

Illinois Standards Achievement Test (ISAT): The Illinois Standards Achievement Test (ISAT) Reading test, introduced in 1998, is administered statewide during the same week in the spring to all students in grades 3, 5, and 8. The ISAT Reading test is divided into three parts, each of which contains single-correct answer multiple choice and extended-response questions. The passage reading portion assessed student skills in reading to gain information and reading for literacy experience by using passages from periodicals, newspapers, short stories and novels. All questions on the test are aligned with the Illinois Learning Standards for Reading which were designed to guide language arts and English curricula in the schools. The description provided by the State of Illinois suggests that this is a criterion-referenced instrument. The

scoring rubric uses a scale of 0 to 4. A score of 0 is given for responses “insufficient to effectively determine evidence of achievement in reading”. Scores of 1 and 2 are awarded for responses indicating “developing levels of achievement in reading”. Students receive a score of 3 for responses indicating “a developed level of achievement” and a score of 4 for responses indicating “a well-developed level of achievement in reading”. Student raw scores are converted to scale scores, which are tied to three performance levels. Student performance is rated as “Exceeding Standards”, “Meeting Standards” or “Below Standards” (ISBE, 1998).

Level Achievement Tests: Level Achievement Tests are administered to all students in the district beginning in the fall of third grade. Level Tests are standardized achievement instruments created from a “blueprint” based on local district goals using items of increasing difficulty which are chosen from a large bank of multiple choice items developed by teams of trained teachers. The tests include enough items at each student’s current level of achievement to ensure reliable assessment of student skills. The difficulty range of the items in each test overlaps the difficulty range of items directly below and above. This is to ensure that a student will be assigned to a test where the middle of the difficulty range closely corresponds to the student’s current achievement level. In contrast to more traditional instruments, the range of difficulty is deliberately rather narrow to diminish both frustration (excessive number of items that are too challenging) and boredom (excessive number of items that are too easy). This feature ensures a higher rate of success and a more positive attitude toward standardized testing while allowing for the creation of shorter, yet more reliable, assessment instruments. The measurement scale used for Level Tests is called a RIT scale, based on Rasch Units, an equal interval scale that offers the advantage of mathematical management (RIT scores can be directly compared to one another). In addition to RIT scores, Level Tests also report student performance in terms of percentile scores and goal scores. In reading, RIT scores range from 150 to 260, with students typically starting at the 170-190 level in third grade and progressing to the 230-260 range by high school. A RIT score of 200 reflects typical performance in the fall of 4th grade. Percentile Scores show the approximate percentage of students in the grade that scored at or below the level of a particular student *in the norming sample*. Unlike RIT scores, which are expected to increase from year to year, percentile scores will remain essentially constant if the student progresses at a rate consistent with the normative sample. Goal Scores reflect relative performance in each goal area compared to the standardization sample, indicating whether the target student’s performance in a goal area is High (exceeds the 66th percentile), Average (33rd to 66th percentile) or Low (below the 33rd percentile) for their age group.

Level Tests offer a number of significant advantages over traditional, nationally normed achievement tests. First, Level Tests are aligned with the local curriculum enhancing the utility of the instrument for informing classroom instruction. Second, Level Tests are calibrated to the curriculum by “assigning a difficulty value to each test item”, in order to develop scale scores that correlate students’ achievement to the specific subject curriculum. Because the test items are deliberately anchored to the local curriculum, the district can reliably monitor student progress from year to year, modify the instrument to reflect changes in the curriculum and maintain consistent normative data. A third advantage of Level Tests is that the instrument is “tailored” to the unique achievement level of each student. Students are initially assigned to the appropriate test level through the use of “locator tests”. Finally, because the test items are directly tied to the instructional curriculum, the use of Level Tests significantly reduces, and may eliminate, the assessment error that results from the poor test-curriculum overlap that plagues many nationally normed standardized achievement tests (NWEA, 1998).

Results

This study consisted of two components. The first component was an analysis of the predictive validity of CBM Oral Reading Fluency measures administered in second through fifth grades. The second component was an evaluation of the linkages between oral reading fluency measures and the state and local standardized achievement instruments utilized by the district. The focus of the second component was evaluating the efficacy of applying CBM Oral Reading Fluency benchmarks established in Oregon and Alaska to data collected in a northeastern Illinois school district.

Inter Correlation of CBM Oral Reading Fluency (ORF) Measures

The initial correlation study examined the relationship between reading fluency probes administered in second through fifth grades. Predictive validity estimates were obtained by correlating performance on the CBM ORF median passage scores in second grade with median passage scores on CBM probes administered in grades three through five. As seen in Table 1, Pearson correlation coefficients demonstrate significant correlation for CBM.

	<i>Grade 2S</i>	<i>Grade 3F</i>	<i>Grade 3S</i>	<i>Grade 4F</i>	<i>Grade 4S</i>	<i>Grade 5F</i>
<i>Fall Grade 2</i>	0.911	0.749	0.713	0.708	0.647	0.699
<i>Spring Grade 2</i>		0.812	0.796	0.792	0.729	0.780
<i>Fall Grade 3</i>			0.889	0.888	0.826	0.877
<i>Spring Grade 3</i>				0.915	0.875	0.916
<i>Fall Grade 4</i>					0.889	0.932
<i>Spring Grade 4</i>						0.907

($p < .0001$ for all cells)

Table 1
Correlation between CBM Oral Reading Fluency, Grades 2 through 5

Correlation of CBM Oral Reading Fluency Measures with ISAT and Level Reading Test

The second correlation analysis examined the relationship between second grade spring ORF, third grade fall ORF and third grade spring ISAT. Predictive validity estimates were obtained by correlating performance on second and third grade CBM ORF and third grade spring ISAT. As seen in Table 2, Pearson correlation coefficients demonstrate significant correlation of CBM ORF with the ISAT.

<i>ORF/ISAT Correlation</i>	<i>Spring Grade 2 ORF</i>	<i>Fall Grade 3 ORF</i>
<i>Grade 3 Spring ISAT</i>	0.627	0.750

($p < .0001$ for all cells)

Table 2
Correlation between CBM ORF and ISAT

The final correlation analysis examined the relationship between spring CBM ORF for second, third and

fourth grades and Fall Level Reading assessment in third through fifth grades. Predictive validity estimates were obtained by correlating performance on second through fourth grade CBM ORF measures with Level Reading assessment results obtained the following fall (third through fifth grades). As seen in Table 3, Pearson correlation coefficients demonstrate significant correlation between CBM ORF and Level Achievement tests for Reading.

<i>ORF/RIT Correlation</i>	<i>Grade 3 Fall Rdg RIT</i>	<i>Grade 4 Fall Rdg RIT</i>	<i>Grade 5 Fall Rdg RIT</i>
<i>Spring ORF Grade 2</i>	0.614	0.622	0.649
<i>Spring ORF Grade 3</i>	x	0.758	0.764
<i>Spring ORF Grade 4</i>	x	x	0.708

(p < .0001 for all cells)

Table 3
Correlation between CBM ORF and Level Reading RIT Score

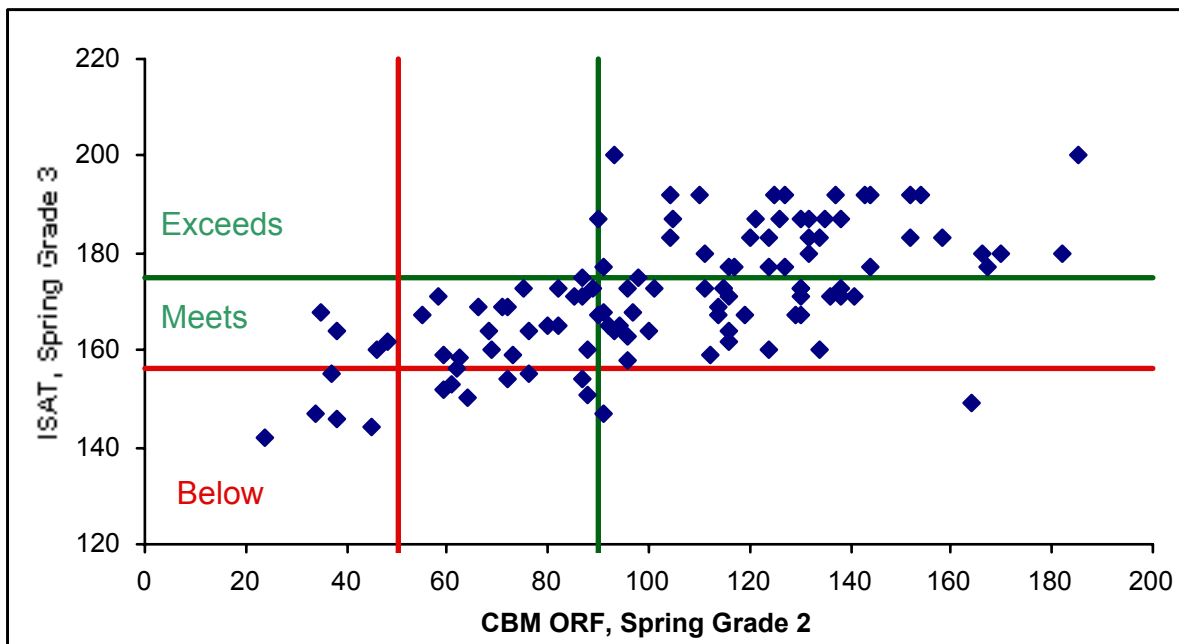
Establishing second, third and fourth grade ORF Benchmarks

The purpose of this component is not to establish independent oral reading fluency benchmarks but to examine linkages between spring ORF benchmarks for second, third and fourth grades and student performance on local and state standardized achievement instruments. Ideally, effective ORF benchmarks should allow us to predict, with some precision, what percentage of students are likely to meet, or not meet, established standards on state and/or local standardized achievement tests. Oral reading fluency benchmarks tend to be reported in pairs. The higher number indicates the fluency rate that predicts that a “very high percentage” of students will meet or exceed standards, while the lower number reflects the fluency rate that predicts a “high percentage” of students will fail to meet standards. This study was “anchored” on the second and third grade reading fluency benchmarks presented by Good et al. (2001) as a result of data-based benchmarks established in Oregon and Alaska. Second grade ORF benchmarks used for this study were 50 and 90 words correct per minute (wcpm). Third grade ORF benchmarks were 90 and 110 wcpm. The final analysis of this study will examine the utility of fourth grade ORF benchmarks of 90 and 130 words correct per minute.

Linkage of Grade 2 CBM Oral Reading Fluency Benchmark Goal and Grade 3 ISAT

The linkage between 2nd grade CBM ORF and third grade performance on the ISAT is illustrated in Figure 1. Students in this linkage were enrolled in the third grade for the 1998-1999 school year. The horizontal lines represent the Illinois standards of “Exceeds Standards” at 175 and “Meets Standards” at 156. A score below 156 is assigned a performance level of “Below Standards”. The vertical line at 90 corresponds to a CBM ORF benchmark goal for April of second grade. Students performing at or above this level of fluency on CBM are likely to meet or exceed the established ISAT standards in the spring of third grade. In this sample, 97% of the students who met the second grade CBM ORF goal met, or exceeded, the stated standards for third grade ISAT. The ISAT outcome is less clear for those students scoring between 50 and 90 on second grade CBM. Of the nine students who read below 50 words correct per minute only 4, or 44%, were able to attain the “meets standards” performance rating on ISAT. Students attaining ORF rates of 90 wcpm or higher have very strong probabilities of “meeting or exceeding” third grade ISAT standards. Similarly, those who read slower than 50 wcpm have a low probability of “meeting or exceeding” third grade ISAT standards.

Figure 1



Linkage between Grade 2 Spring ORF and Grade 3 Spring ISAT

Linkage of Grade 2 CBM Oral Reading Fluency Benchmark Goal and Grade 3 Level Reading Test

The linkage between 2nd grade CBM ORF and third grade performance on the Level Reading Test is illustrated in Figure 2. Students in this linkage were enrolled in the third grade for the 1998-1999 school year. The horizontal lines represent the Level Test performance standards of “High” at 197 or higher and “Average” at 183. A score below 183 is assigned a performance level of “Low”. The vertical line at 90 corresponds to a CBM ORF benchmark goal for April of second grade. Students performing at, or above, this level of fluency on CBM are likely to attain a performance rating of “Average” or “High” on Level Reading Test in the Fall of third grade. In this sample, 71 of 73, or 97% (64% of the full sample), of the students who met the second grade CBM ORF goal of 90 words correct per minute met, or exceeded, the stated standards for the third grade Level Reading Test. The Level Test outcome is less clear for those students scoring between 50 and 90 wcpm on second grade CBM. However, in this particular sample 31 students read between 50 and 90 words per minute. Of those 31 students, 23 or 74% (20% of the full sample), were able to attain the desired performance rating to “Average” on the Level Reading Test. Of the ten students who read below 50 word correct per minute only 4, or 40% (3.5% of the full sample), were able to attain the “Average” performance rating on Level Reading.

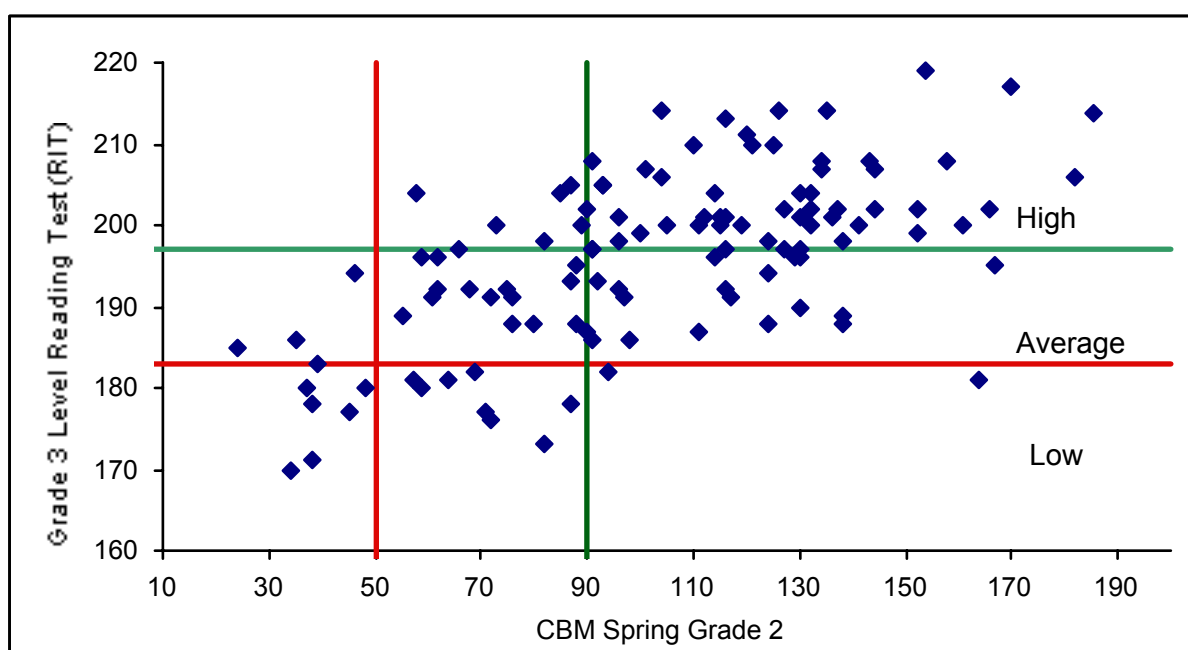


Figure 2
Linkage between Grade 2 Spring ORF and Grade 3 Fall Level Reading Test

Linkage of Grade 3 CBM Oral Reading Fluency Benchmark Goal and Grade 4 Level Reading

The linkage between 3rd grade CBM ORF and fourth grade performance on the Level Reading Test is illustrated in Figure 3. Students in this linkage were enrolled in the fourth grade for the 1999-2000 school year. The horizontal lines represent the Level Test performance standards of “High” at 204 or higher and “Average” at 191. A score below 191 is assigned a performance level Low”. The vertical line at 110 corresponds to a CBM ORF benchmark goal for April of third grade. Students performing at or above this level of fluency on CBM are likely to meet or exceed the established Level Reading Test performance standards in the fall of fourth grade. In this sample, 74 of 74, or 100% (65% of the full sample), of the students who met the third grade CBM ORF goal of 110 words correct per minute attained a performance rating of “Average” or “High” for the fourth grade Level Reading Test. The Level Test outcome is less clear for those students scoring between 70 and 110 wcpm on second grade CBM, although all 31 (27% of the sample) students who read between 70 and 110 words per minute were able to attain the rating of “Average”. Of the nine students who read below 70 word correct per minute only 3, or 33% (2.6% of the full sample), were able to attain an “Average” or “High” performance rating on the fourth grade Level Reading Test.

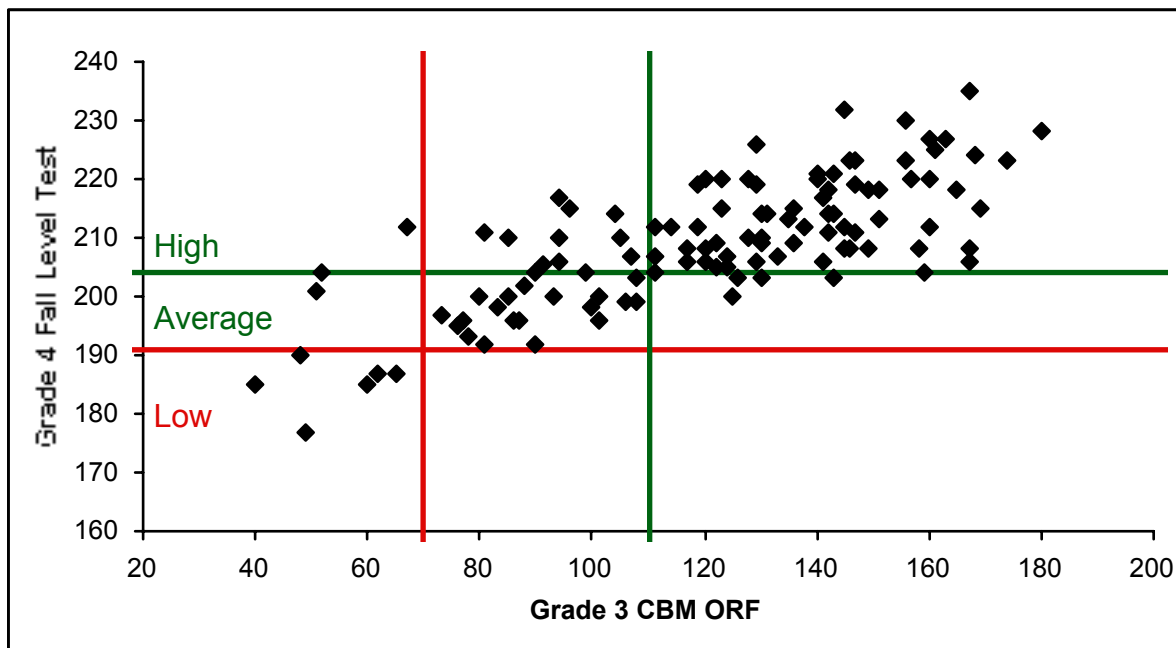


Figure 3
Linkage between Grade 3 Spring ORF and Grade 4 Fall Level Reading Test

Linkage of Grade 4 CBM Oral Reading Fluency Benchmark Goal and Grade 5 Level Reading

Consistent with the purposes of this study, Figure 4 presents suggested benchmarks for fourth grade reading fluency which are linked to student performance on the fifth grade Level Reading Test. Students in this linkage were enrolled in the fifth grade for the 2000-2001 school year. The horizontal lines represent the Level Test performance standards of “High” at 211 or higher and “Average” at 199. A score below 199 is assigned a performance level “Low”. The vertical line at 130 represents a CBM ORF benchmark goal for April of fourth grade. There is a strong probability that students performing at or above this level of fluency on CBM will attain a performance rating of “High” on the Level Reading Test in the fall of fifth grade. In fact, 69 of 69, or 100% (60% of the full sample), of the students who met the fourth grade CBM ORF goal of 130 words correct per minute attained a performance rating of “Average” or “High” for the fifth grade Level Reading Test. The Level Test outcome is less clear for those students scoring between 90 and 130 wcpm on fourth grade CBM. In this particular sample 32 students read between 90 and 130 wcpm. Of those 32 students, 27 or 84% (23% of the full sample), were able to attain performance ratings of “Average” or “High”. Of the 13 students who read below 90 word correct per minute 7, or 54% (6% of the full sample), were able to attain an “Average” or “High” performance rating on the fifth grade Level Reading Test.

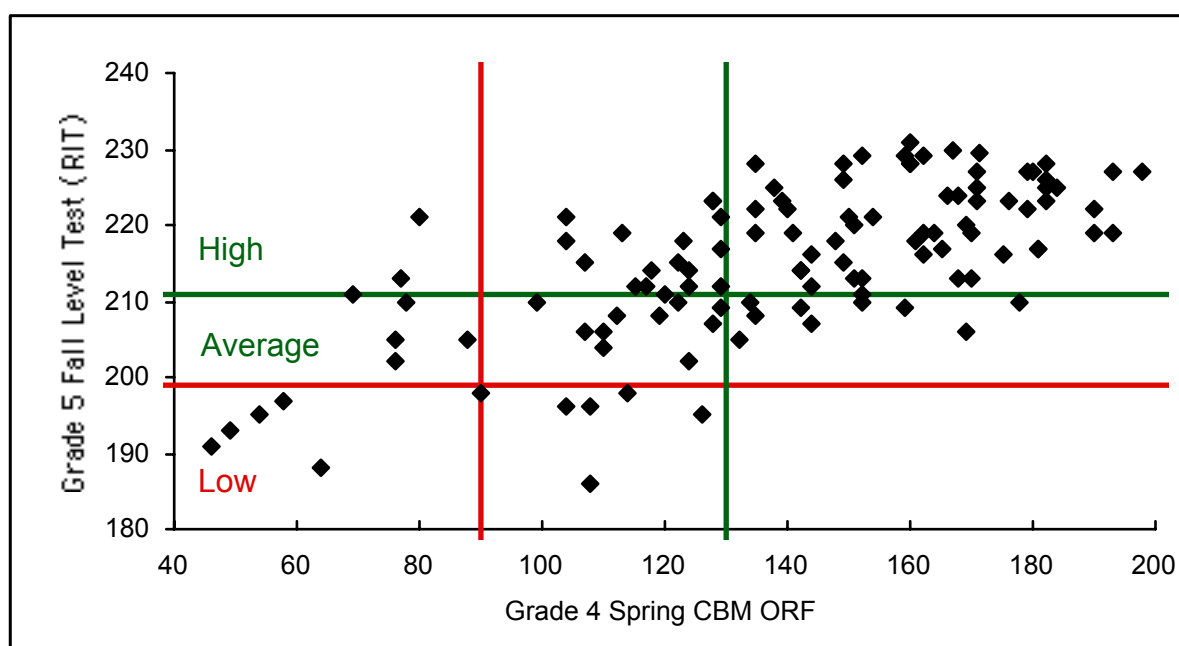


Figure 4
Linkage between Grade 4 Spring ORF and Grade 5 Fall Level Reading Test

Discussion

The primary purpose of this study is to examine the relationship between CBM oral reading fluency measures and state and local achievement tests in reading. Specifically, this study focused on an examination of the “universality” of oral reading fluency benchmarks previously established in Oregon and Alaska.

Before the utility of existing oral reading fluency benchmarks could be examined, however, it was necessary to examine the predictive validity of the oral reading fluency measures currently utilized by the local district.

Inter correlation of CBM Oral Reading Fluency Measures

First, inter correlation of the oral reading fluency measures was examined. Consistent with data presented by Shapiro (1999) and a correlation study conducted last year by the primary author (Sibley et al., 2000), the inter correlation of the CBM oral reading fluency measures is quite strong. The correlation coefficients for the reading fluency measures are also of the magnitude expected.

The second analysis focused on the correlation between oral reading fluency and the ISAT. The result of that analysis establishes a very strong link between CBM oral reading fluency and the state achievement test for reading. In fact, the correlation between the second grade spring CBM oral reading fluency and third grade spring ISAT is sufficiently strong that problem identification and problem certification decisions can be made with the CBM oral reading fluency data. Certainly, the CBM data can be used to identify those students who may not meet established standards on the ISAT.

The third analysis examined the relationship between oral reading fluency and the Level Reading Test. Results of that analysis establish a very strong link between CBM oral reading fluency and the Level Reading Test for all grades examined. The correlation coefficients are of sufficient magnitude that critical student decisions can be made based on CBM oral reading fluency.

Utility of Applying Established Oral Reading Fluency Benchmarks

As noted, the primary purpose of this study is to evaluate the strength of the relationship between oral reading fluency benchmarks and established standardized achievement tests. Previous research in Oregon and Alaska established second and third grade spring oral reading fluency benchmarks of 90 and 110 words correct per minute, respectively. The results of this study support the utility of these fluency based performance standards for predicting student performance on subsequent “high stakes” achievement measures. Oral reading fluency benchmarks established in Oregon and Alaska appear to have a very high level of utility for the population examined in this study. These results also support the attempts to establish global, rather than local, oral reading fluency benchmarks. The results of these analyses are summarized in Table 5.

	Grade 1 Spring	Grade 2 Spring	Grade 3 Spring	Grade 4 Spring
Target Oral Reading Fluency Rates	≥ 40 wcpm	50 - 90 wcpm	70 – 110 wcpm	90 – 130 wcpm

*Table 5
Target Oral Reading Fluency Rates for Grades 2 through 5*

The data presented in this study demonstrate the utility of applying existing second grade reading fluency benchmarks. This study clearly demonstrates that those students who are able to read at or above 90 words correct per minute in the spring of second grade have a very high probability of attaining the “meets or exceeds” standards on the third grade ISAT Reading test. These same students have a very high probability

of attaining the “Average” or “High” performance rating on the third grade Level Test for Reading. Similarly, students who read at or below 50 correct words per minute in the spring of second grade have a low probability of attaining those performance ratings on the third grade ISAT or Level Test for Reading.

This study also demonstrates the utility of applying established oral reading fluency benchmarks for third grade. Students in this study who were able to read at or above 110 correct words per minute in the spring of third grade were able to attain performance ratings of “Average” or “High” on the fourth grade Level Reading Test. In contrast, students who read at or below 70 words correct per minute in the spring of third grade were considerably less likely to attain performance standards of “Average” or “High”.

An additional purpose of this study was to propose oral reading fluency benchmarks for the spring of fourth grade which are linked to attainment of desired ISAT and Level Test standards for fifth grade. The initial benchmarks proposed in this study were established by using slope data developed by Fuchs et al. (1993). The authors estimated 36 instructional weeks between test administration in the spring of third grade and spring of fourth grade. Slope data provided by Fuchs et al. suggest that an improvement of 30 to 36 words correct per minute could be expected, given predicted increases of .85 to 1.0 words per week. Adding 30 words per minute to the third grade benchmarks predicted fourth grade reading fluency benchmarks of 100 and 140 words per minute. When those benchmarks were applied, 100 percent of the students who met that standard also attained at least an “Average” rating on the fifth grade Level Test. On the other hand, 53 percent of those who read below 100 words per minute were able to attain a Level Test rating of “Average” or better.

There is evidence that these benchmarks appear to be overly ambitious. Good et al (2001) suggested that benchmarks need to establish “realistic” standards for instruction. Reading fluency benchmarks that predict 100% of students will meet or exceed criterion standards are, therefore, unrealistic and place undue burden on instruction. Accordingly, the benchmarks were adjusted downward to 130 and 90 words per minute. Curiously, this change had very little impact on the linkage between CBM and the Level Test. With the upper benchmark set at 130 and the lower benchmark set at 90 words per minute, 100 percent of the students who met the upper benchmark also met the Level Test standard of “Average” or better. When the lower benchmark was set at 90 words per minute, 54 percent of the students who read below that level were still able to attain the desired level of performance on the Level Test. The data reported in this study lend strong support to the use of fourth grade oral reading fluency benchmarks of 90 and 130 words per minute. Additional study is needed to determine if the benchmarks proposed in this study are sufficiently stringent but also realistic with regard to instruction.

The instructional implications are quite evident. There is a growing body of data that clearly demonstrate not only the efficacy, but also the necessity, that students be reading connected text at the rate of *at least* 40 words correct per minute in the spring of first grade in order to maintain an adequate rate of growth toward acquisition of reading skills. Furthermore, those children who are able to read at least 50 words correct per minute in the spring of second grade appear to be well on their way to attaining desirable outcomes on “high stakes” standardized achievement tests. Clearly, those children who exceed this minimal level of performance have an even higher probability of achieving the desired performance standards in third grade and beyond. From an instructional standpoint, students who fail to meet this critical early benchmark require intensive instructional interventions to get them back on track. In the absence of such intervention, the data clearly indicate a strong probability that these students will fail to achieve critical performance standards on “high stakes” assessments. Instruction in early literacy and early reading of connected text clearly needs to be directed at attaining the goal of *at least* 50 words correct per minute in the spring of second grade. CBM

procedures can be incorporated for the purpose of progress monitoring and informing about the rate of growth of target students and the effectiveness of the interventions.

The data collected for this study suggests that 50 wcpm may, in fact, be a low estimate for second grade students in the district targeted by this study. Students in that district showed a mean reading rate of 103 wcpm with a standard deviation of 36; 84% read faster than 67 wcpm. Therefore, students reading at 50 wcpm in the target district would be found to be “at risk” for reading failure. Although the benchmarks established in Alaska and Oregon do not have high utility at this grade level, the benchmarks developed for subsequent grade levels do appear to have sufficient utility for essential decision making and appear to provide realistic “targets” for instruction. There is a need for further research within this same setting to determine if the results of this study stand up over time and with different groups of students. In order to have confidence in these results there needs to be evidence that the group used for this study is representative of students in the district rather than a group that is somehow unique. Further research will provide for greater refinement and definition of reading fluency benchmarks.

References

- Achievement Level Tests* (1998). Northwest Evaluation Association. Portland, Oregon.
- Canter, A., & Marston, D. (1998). Curriculum-based measurement: A handout for teachers. In A. S. Cantor & S. A. Carroll (Eds.), *Helping Children at Home and School: Handouts from Your School Psychologist* (pgs.163–164). Bethesda, MD: National Association of School Psychologists.
- Fuchs, L., Fuchs, D., & Deno, S. (1982). *Developing goals and objectives for educational programs. [Teaching guide]*. In Shapiro, E. S. (1996). *Academic skills problems: Direct assessment and intervention*. New York: Guilford Press.
- Fuchs, L., Fuchs, D., Hamlett, C., Walz, L., & Germann, G. (1993). *Formative evaluation of academic progress: How much growth can we expect?* *School Psychology Review*, 22, 27-48.
- Goh., D., Teslow, C., & Fuller, G. (1981). *The practices of psychological assessment among school psychologists*. *Professional Psychology*, 12, 699-706.
- Good, R., Simmons, D. & Kame'enui, E. (2001). *The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes*.
- Understanding your child's ISAT scores: A parent guide to the Illinois Standards Achievement Test*. (1998) Illinois State Board of Education: Springfield, IL.
- Kaminski, R., & Good, R. (1998). *Assessing early literacy skills in a problem solving model: Dynamic indicators of basic early literacy skills*. In M. R. Shinn (Ed.) *Advanced applications of curriculum-based measurement*. New York: Guilford.
- Ownby, R., Wallbrown, F., D'Atri, A., & Armstrong, B. (1985). *Patterns of referrals for school psychological services: Replication of the referral problems category system*. *Special Services in the School*. 1(4), 53-66.
- Salvia, J., & Ysseldyke, J. E. (1995). *Assessment in special and remedial education* (6th ed.). Boston: Houghton Mifflin.
- Shapiro, E. S. (1996). *Academic skills problems: Direct assessment and intervention*. New York: Guilford Press.
- Shinn, M. (1989). *Curriculum-based measurement: Assessing special children*. New York: Guilford Press. (In particular, see chapter by D. Marston.)
- Sibley, D., Bollman, K., Droho, J., Hughes, B., & Otte, L. (2000). *Gender Differences Across Grades and Across Subjects on CBM Measures*. Unpublished research. Paper presentation at National Association of School Psychologist Conference, New Orleans.
- 2000 National Assessment of Education Progress* (2000). National Center for Educational Statistics, Washington, D.C.
- Sibley, D., Biwer, D., & Hesch, A. (2001). Unpublished Data. Arlington Heights, IL: AHSD 25.