

**AN ANALYSIS OF THE CONCURRENT AND PREDICTIVE VALIDITY OF
CURRICULUM BASED MEASURES (CBM), THE MEASURES OF ACADEMIC
PROGRESS (MAP), AND THE NEW ENGLAND COMMON ASSESSMENT
PROGRAM (NECAP) FOR READING**

By

Kristina J. Andren

B.A. Carleton College, 1997

M.S. University of Southern Maine, 2005

A DISSERTATION

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Psychology

(in School Psychology)

The University of Southern Maine

August, 2010

Advisory Committee:

Rachel Brown-Chidsey, Associate Professor of School Psychology, Advisor

Mark W. Steege, Professor of School Psychology

David L. Silvernail, Professor of Research and Evaluation

© 2010 Kristina J. Andren

All Rights Reserved

**AN ANALYSIS OF THE CONCURRENT AND PREDICTIVE VALIDITY OF
CURRICULUM BASED MEASURES (CBM), THE MEASURES OF ACADEMIC
PROGRESS (MAP), AND THE NEW ENGLAND COMMON ASSESSMENT
PROGRAM (NECAP) FOR READING**

By Kristina J. Andren, M.S.

Dissertation Advisor: Dr. Rachel Brown-Chidsey

An Abstract of the Dissertation Presented
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Psychology
(in School Psychology)
August, 2010

This study examined the concurrent validity of four different reading assessments that are commonly used to screen students at risk for reading difficulties by measuring the correlation of the third grade Measures of Academic Progress (MAP) in Reading with three specific versions of curriculum based measurement: DIBELS oral reading fluency (ORF), AIMSweb ORF, and AIMSweb Maze. In addition, correlations were calculated among each of these measures with the third grade New England Common Assessment Program (NECAP) measure of reading achievement. Multiple regression analyses also were conducted to provide information on the predictive validity of CBM and MAP in determining risk for reading difficulty, as measured by a high stakes assessment (e.g., NECAP). Reading performance data were collected on 137 third grade students in the fall and winter. Significant correlations were found among each measure of reading at each point in time ($p < .001$). Correlations ranged from .972 (DIBELS ORF and

AIMSweb ORF) to .621 (Maze and NECAP). Within each measure, ORF had the highest correlations between fall and winter measures ($r = .952$), followed by the MAP ($r = .872$) and Maze ($r = .746$), respectively. Regression analyses revealed that the MAP assessment in the fall best predicted MAP scores in the winter ($p < .001$), followed by oral reading fluency ($p < .05$). MAP was also the best predictor of NECAP scores for the general population of students ($p < .001$), as well as those students receiving supplemental reading support ($p < .001$). When MAP was removed from the equation, ORF was the most significant predictor of performance on the NECAP for general education ($p < .001$) and at-risk readers ($p < .001$). Educational implications and suggestions for further research are discussed.

ACKNOWLEDGEMENTS

I am grateful to Dr. Rachel Brown-Chidsey for her guidance with every step of this project.

I would also like to thank
Dr. F. Charles Mace
Dr. David Silvernail
Dr. Mark Steege
For helping to design and refine this study.

For listening to our young readers, I thank
Alexis Berry
Rebekah Bickford
Eileen Harris
Danae Kesel
Caroline Wallace

Thank you to the administrators and third grade teachers who cooperated with this project.

To Stefan and Elsa,
por ser vosotros

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	v
LIST OF TABLES.....	vii
Chapter	
1. INTRODUCTION & LITERATURE REVIEW	1
2. METHOD	7
Participants.....	7
Measures	8
Procedures.....	11
3. RESULTS	12
Descriptive Statistics.....	12
Correlation Analysis	13
Regression Analysis.....	16
4. DISCUSSION.....	19
Concurrent and Predictive Validity of Reading Measures	19
Educational Implications	22
Limitations and Future Research	24
5. SUMMARY	25
REFERENCES	26
BIOGRAPHY OF THE AUTHOR.....	31

LIST OF TABLES

Table 1: DEMOGRAPHIC DATA	8
Table 2: MEANS AND STANDARD DEVIATIONS (SD) FOR EACH MEASURE OF READING BY GROUP	12
Table 3: INTERCORRELATIONS BETWEEN READING MEASURES IN SEPTEMBER AND JANUARY, OVERALL SAMPLE	14
Table 4: INTERCORRELATIONS BETWEEN READING MEASURES IN SEPTEMBER AND JANUARY FOR STUDENTS IN TIER 1	15
Table 5: SUMMARY OF MULTIPLE REGRESSION ANALYSIS FOR VARIABLES PREDICTING WINTER MAP SCORES	16
Table 6: SUMMARY OF MULTIPLE REGRESSION ANALYSIS FOR CBM VARIABLES PREDICTING WINTER MAP SCORES	16
Table 7: SUMMARY OF MULTIPLE REGRESSION ANALYSIS FOR VARIABLES PREDICTING NECAP SCORES (N = 137).....	17
Table 8: SUMMARY OF MULTIPLE REGRESSION ANALYSIS FOR CBM VARIABLES PREDICTING NECAP SCORES	17
Table 9: SUMMARY OF MULTIPLE REGRESSION ANALYSIS FOR VARIABLES PREDICTING NECAP SCORES IN SAMPLE OF LOWER PERFORMING READERS	18

Table 10: SUMMARY OF MULTIPLE REGRESSION ANALYSIS FOR CBM VARIABLES
PREDICTING NECAP SCORES IN SAMPLE OF LOWER PERFORMING

READERS 18

Introduction and Literature Review

With recent educational reforms, schools nationwide are experiencing increased demands for assessment, accountability, and evidence-based practices in education. The No Child Left Behind Act (NCLB) of 2001 includes a requirement that academic instruction and assessment methods in both general and special education be scientifically-based (U.S. Department of Education, 2002). The 2004 reauthorization of the Individuals with Disabilities Education Improvement Act (IDEIA) incorporated requirements for Response to Intervention (RTI) initiatives, including scientifically based reading instruction, evaluation of students' response to intervention, and data-based decision-making (IDEIA, 2004). RTI is a three-tiered model of assessment and intervention that involves systematic, data-based decision making. The process includes universal screening, early intervention, and ongoing progress monitoring (Brown-Chidsey & Steege, 2005). There are a wide variety of assessment measures that have been used within an RTI model, which differ in technical adequacy, administration setting and procedures, skills tested, and time and financial requirements.

Curriculum-based measurement (CBM) is a method of formative evaluation within a problem-solving model, which utilizes brief standardized tests of basic academic skills (Deno, 2005; Shinn, 2002). The Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) are computerized adaptive tests that measure academic achievement in the areas of mathematics, reading, and language usage. The MAP is designed to provide academic achievement data, in order for educators to gather school-wide data and develop targeted instruction for individuals and groups of students. The purpose of this study was to examine the correlation of the MAP in Reading with three specific versions of curriculum based measurement using the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) and AIMSweb CBM

reading items. In addition, the predictive qualities of both the MAP and CBM in relation to student performance on the New England Common Assessment Program (NECAP) were examined. This analysis provides information on the concurrent and predictive validity of these four reading measures.

A cornerstone of RTI is a technically sound method for gathering data on student progress, in order to identify students in need of additional instruction and to determine whether the instruction is effective. It is important that the assessment method is valid and reliable, efficient and cost-effective to administer, and sensitive to student growth (Deno, 2005; Shinn, 2002). Curriculum based measurement is one tool for gathering universal screening and progress monitoring data that meets these three requirements (Deno et al., 2009; Stecker, Lembke, & Foegen, 2008). In addition, teachers' use of CBM within a data-based decision-making model has been associated with improvements in student achievement, providing evidence for its utility in guiding instructional decisions (Deno et al., 2009; Stecker, Fuchs, & Fuchs, 2005).

In reading, there are two primary curriculum based measures: oral reading fluency (ORF) and maze tasks. Examples of oral reading fluency probes are the DIBELS and AIMSweb. For these instruments, students are given a grade-level reading passage and are asked to read aloud for one minute. The metric for this assessment is the number of words read correctly in one minute (Good, Gruba, & Kaminski, 2002). Maze procedures use reading probes in which every seventh word is removed and replaced with a choice of three words, one of which is the correct word and two are distracters. Administered in a group setting, students are instructed to read the passage and circle the correct word for each blank space, for three minutes. Scores on a maze

task represent the total number of words circled correctly in the given time period (AIMSweb, 2008).

Numerous studies have demonstrated a relationship between CBM data and student performance on standardized measures of achievement (Deno et al., 2009; Hintze & Silbergitt, 2005; Silbergitt, Burns, Madyun, & Lail, 2006). These findings are important because they suggest that brief CBM can help predict performance on high stakes tests, thereby providing opportunities for early identification and intervention. Ardoin et al. (2004), for example, investigated the correlation between oral reading fluency, maze, a group-administered achievement test, and reading subtests of the Woodcock-Johnson Tests of Achievement, Third Edition (WJ-III). Seventy-seven third grade students were given all four assessments, and correlations, t-tests to measure differences in correlations, and multiple regression analyses were conducted. All correlations between ORF, maze, and the WJ-III subtests were statistically significant. ORF was more closely related than the maze to the WJ-III, and the addition of the maze did not significantly increase the predictive power of ORF. The authors concluded that although both curriculum-based measures correlated significantly with the WJ-III, ORF was a better predictor of overall reading achievement and reading comprehension.

Wiley & Deno (2005) compared the predictive value of ORF and maze tasks by administering both to a group of third and fifth grade students, and correlating their scores with a state standards test. Moderate correlations were found between both CBM measures and the state assessment. Furthermore, combining ORF and maze increased the predictive power. This finding was not true, however, for English language learners (ELLs). For ELLs, maze was a better predictor than ORF, while ORF better predicted statewide test scores for non-ELL

students. Overall, these results support the use of CBM in reading for screening and progress monitoring.

Similarly, Deno et al. (2009) investigated the use of a maze task as a universal screening measure by examining the relationship between performance on the maze and a standardized test of reading. Correlations between the two reading measures ranged from .61 to .77. In addition, school-wide data indicated that maze scores increased steadily with each grade level over the course of two school years, providing support for its use as a progress monitoring measure. The authors concluded that given its evidence of validity and utility in identifying students at risk, and its group administration format, maze procedures are efficient, effective, and provide clear data as a universal screening measure within a school-wide RTI model.

Several studies have examined the correlation between ORF and standardized reading tests over a longer time period, and found similar results. In a replication of a study by Stage & Jacobsen (2001), which found significant correlations between oral reading fluency and a state reading test for fourth grade students, McGlinchey & Hixson (2004) examined a larger sample of students and tracked performance over eight years. The researchers found moderately strong correlations (.67 on average) between a single ORF probe and performance on the Michigan Educational Assessment Program assessment of reading. In another longitudinal study, Keller-Margulis, Shapiro, & Hintze (2008) found moderate to strong correlations between CBM data and achievement test scores in reading one and two years later. Evidence from longitudinal studies suggests that the magnitude of the correlation between ORF and state reading achievement test data tends to decrease with advancing grade levels (Silberglitt, Burns, Madyun, & Lail, 2006).

There is a strong research base providing evidence that measures of oral reading fluency (Ardoin & Christ, 2008; Baker et al, 2008; Hintze & Silbergitt, 2005) and maze procedures (Begeny & Martens, 2006; Graney, Missall, Martinez, & Bergstrom, 2009) are sensitive to student growth over time. These findings suggest that CBM can be administered on an ongoing basis to track student growth. When used for this purpose, the slope, or rate of growth, in addition to the student's reading level, can predict performance on standardized reading measures (Baker et al., 2008; Keller-Margulis, Shapiro, & Hintze, 2008). In a study on the relationship between ORF level and slope and high stakes reading tests, Baker et al. (2008) found that the best fitting predictive model for performance on high stakes reading tests included ORF level and slope, and the state reading assessment scores from the previous year. Taken together, these studies provide strong evidence for the predictive value and measurement sensitivity of CBM in reading.

Standardized achievement tests have also been used to assess students' reading skills. The Northwest Evaluation Association (NWEA) has developed a computerized achievement test known as the Measures of Academic Progress (MAP). The MAP is a group-administered, computerized adaptive test that measures academic achievement in the areas of mathematics, reading, and language usage. A comparison of student performance on computerized adaptive tests and paper-and-pencil achievement tests indicated no significant differences in reading scores (Kingsbury, 2002). Overall, test modality had very little observed effect (less than one scale score) on student performance. The Measures of Academic Progress in Reading include items assessing word recognition and vocabulary, reading comprehension, and literary analysis. Scores on the MAP are reported as Rasch Unit (RIT) scores, percentiles, and growth scores.

According to studies published by the NWEA (2004), there is evidence of concurrent validity of the MAP with numerous state achievement tests. For measures of reading in third grade, the correlations range from .66 with the Texas Assessment of Knowledge and Skills to .87 with the Stanford Achievement Test, 9th Edition. Another study, examining the predictive value of the MAP on the Delaware State Testing Program, found a correlation of .54 between the two reading measures (Hall-Michalcewiz, 2008). DeLong (2007) investigated the relationship between the MAP and state assessments in a different way. This study compared scores on the Indiana Statewide Testing for Educational Progress-Plus (ISTEP+) for schools that utilize MAP testing and those that do not. Results indicated no significant correlation between ISTEP+ scores and the use of the MAP or the level of implementation of MAP testing.

In the state of Maine, NWEA conducted a study to determine how well RIT scores on the MAP correlated with student performance on the Maine Educational Assessment (MEA), and to identify RIT cut-scores that would predict success on the MEA (Cronin, 2004). Results indicated a correlation of .74 for fourth grade reading scores. This suggests that the MAP has the capacity to identify students at risk for academic difficulties, and given individual subtest scores, determine the subject area(s) where support is most needed. In 2009, Maine changed the state-required test and began using the New England Common Assessment Program (NECAP) instead of the MEA. Despite the widespread use of MAP in Maine schools, no studies documenting the correlation between MAP and NECAP scores were found.

The goals of universal screening and progress monitoring measures are to identify those students at risk for learning difficulties so that they can receive early, targeted intervention, and to evaluate the effectiveness of the instruction. To do this, it is important to identify assessment measures that best describe and predict future reading ability. Previous research has separately

examined the concurrent and predictive validity of CBM and MAP with numerous state reading assessments and standardized achievement measures, yet the direct relationship between CBM and MAP has not been explored. This information is important to school personnel who make decisions about universal and benchmark screening, because the MAP and CBM are often used together for these purposes. This study addressed the following research questions:

1. What is the concurrent validity of the MAP, DIBELS and AIMSweb ORF, and AIMSweb maze measures of reading?
2. What are the correlations of each measure at fall and winter benchmarks?
3. How well do CBM and MAP assessments in the fall predict future reading performance, as measured by the post-test MAP and the New England Common Assessment Program (NECAP), the newly adopted statewide achievement test in reading?

First, it was hypothesized that DIBELS and AIMSweb measures are more closely correlated with one another than with the MAP measure of reading. Second, this study tested the hypothesis that CBM of reading in the form of oral reading fluency and maze predict mid-year scores on the MAP and achievement scores on the NECAP as well or better than fall MAP scores.

Method

Participants

The participants in this study were 137 third grade students from two schools in a suburban public school district in the northeastern United States. Table 1 summarizes data concerning the participants. The two elementary schools enrolled 448 and 199 students, respectively. The racial/ethnic breakdown of the schools was 96% Caucasian, 1% African

American, 1% Hispanic, and 2% Asian. The percentage of students who were eligible for free or reduced price lunch in each school was 19%.

Table 1

Demographic Data

Boys	Girls	Tier 2/3	Special Education	Total
70	67	11	11	137

Data were gathered on 150 students, although 13 students did not participate in every measure at each point in time. Only those students with complete datasets were included in the analysis. Of the final sample, 51% (n = 70) were boys and 49% (n = 67) were girls. Students were coded according to whether they were receiving Tier 1 instruction only, supplemental Tier 2 or 3 reading instruction, or special education services. Tier 1 is the general education curriculum within a regular classroom. Students in Tier 2 and Tier 3 typically received Tier 1 instruction as well as small group reading instruction for 20 – 30 minutes per day outside of the regular classroom. Eleven of the students had Individual Education Plans (IEPs), and 11 students received supplemental Tier 2 or 3 reading instruction. A university human subjects review board approved all procedures.

Measures

Five different measures of reading were used in this study: (a) Dynamic Indicators of Basic Early Literacy Skills (DIBELS) oral reading fluency (ORF), (b), AIMsweb ORF, (c) AIMsweb maze, (d) NWEA Measures of Academic Progress (MAP) reading subtest, and (e) the New England Common Assessment Program (NECAP) reading subtest.

DIBELS. Third grade Dynamic Indicators of Basic Early Literacy Skills (DIBELS) were used to measure oral reading fluency (ORF). Three grade-level reading probes were

administered to each student individually. Each student was given a one-page reading passage and instructed to read aloud for one minute. Performance was measured by the number of words read correctly in one minute. The median of the three scores was used in this analysis (Good, Gruba, & Kaminski, 2002).

AIMSweb. AIMSweb is a commercial assessment and web-based data management system, which consists of oral reading fluency and maze assessments for reading. The ORF measure consisted of three one-minute reading probes that were administered to students individually in a fashion very similar to the DIBELS. Each student was given a one-page reading passage and instructed to read aloud for one minute. Performance was measured by the number of words read correctly in one minute. The median of the three scores was used in this analysis. For the maze assessment, students were given a reading passage in which every seventh word has been deleted and replaced with three multiple-choice alternatives. Administered in classroom groups, students were asked to read the passage silently for three minutes and choose the correct words when they came to the word choices. Performance was based on the number of correct choices made within three minutes. Because the maze task includes three minutes of reading, only one passage was used at a time (AIMSweb, 2008).

MAP. The Northwest Evaluation Association (NWEA) Measures of Academic Progress (MAP) are computerized adaptive tests that measure academic achievement in the areas of mathematics, reading, and language usage. The test was given in a group setting using computers. The difficulty of each question was based on the student's accuracy on prior questions, so that each test is adjusted to the individual student's performance level. Scores on the MAP were reported as Rasch Unit (RIT) scores, percentiles, and analyses of progress. The RIT scale is an equal interval scale which estimates student achievement based on the difficulty

of individual items. Using this scale, results of the MAP also can be reported as improvement scores, which represent the number of RIT points gained by a student since the previous assessment and the extent to which a student exceeds or falls short of the average growth (Northwest Evaluation Association, n.d.).

The MAP for Reading included five subtests: Word Recognition and Vocabulary, Reading Comprehension – Literal, Reading Comprehension – Inferential/Interpretive, Reading Comprehension – Evaluation, and Literary Response and Analysis. Word Recognition and Vocabulary items measured a student’s ability to use context cues to understand word meanings and relationships between words. Reading comprehension subtests measured students’ ability to recall, identify, classify and sequence stated content, make predictions and inferences, synthesize information, and evaluate, compare, and apply what they have read. Literary Response and Analysis items required students to respond to questions about a story’s characters, themes, plot, and setting (Northwest Evaluation Association, n.d.).

NECAP. The New England Common Assessments Program (NECAP) was developed by New Hampshire, Rhode Island, and Vermont, based on grade level expectations and the requirements of the No Child Left Behind Act (NCLB). Maine joined the program in 2009, and administers the test to students in grades three through eight annually in the fall. The NECAP measures academic achievement in reading, writing, and mathematics. The reading test consisted of multiple-choice and constructed-response questions in six content areas: Word Identification Skills and Strategies, Vocabulary Strategies and Breadth of Vocabulary, Initial Understanding of Literary Texts, Analysis and Interpretation of Literary Text, Citing Information, Initial Understanding of Informational Text, and Analysis and Interpretation of Informational Text, Citing Evidence (Measured Progress, 2009).

Procedures

MAP and CBM data were collected at two time points: September 2009 and January 2010. The MAP was administered in a group setting, located in the computer lab at the students' schools. Standardized procedures, as provided by NWEA, were followed. Within two weeks of each MAP administration, DIBELS and AIMSweb measures were administered. All students were given these tests in the same one-week period. DIBELS and AIMSweb ORF were administered individually, in a single session, in private testing rooms located in the students' schools. Researchers used administration procedures as published in the AIMSweb and DIBELS administration and scoring guides. AIMSweb maze assessments were administered in a group setting in the students' regular classrooms. Researchers used administration procedures as published in the AIMSweb administration and scoring guide. All curriculum based measures were administered by school psychology graduate students and faculty. The NECAP was administered to all students in October 2009, according to the procedures specified by NECAP and the Maine Department of Education

In order to ensure that all individually administered reading assessments were conducted in the same manner with each student and according to the administration methods, interobserver agreement (IOA) was calculated. To compute IOA 20% of the oral reading fluency (ORF) CBM administrations were observed and scored by a second researcher. The second researcher sat behind the first one and listened to the ORF session and scored the student's oral reading in the same exact manner as the first researcher did. The dual scores for the IOA sessions were compared to determine the number of times that the two scores matched (agreements), and how many times they differed (disagreements). The occurrence of agreements, and disagreements were calculated on a word-by-word basis. IOA was then calculated by dividing the total number

of agreements by the number of agreements plus disagreements and multiplying by 100 to determine the percentage of agreements. Total IOA across the sample of double-scored ORF reading sessions was 98%.

Results

Table 2 presents descriptive statistics for each measure of reading by instructional group. On average, this sample exceeded national norms by 10 points on oral reading fluency measures (Good et al, 2002; AIMSweb, 2010b), and four points on the MAP (NWEA, 2008). Performance on the maze task was equal to or less than national averages by one point (AIMSweb, 2010a). When compared to state averages on the NECAP, the participants in this study achieved a higher mean score by four points (NECAP, 2010). Notable were lower scores of students receiving Tier 2 or 3 instruction, even when compared to those receiving special education services.

Table 2

Means and Standard Deviations (SD) for Each Measure of Reading by Group

Measure		Overall Sample (N = 137)		Tier 1 (N = 115)		Tiers 2 & 3 (N = 11)		Special Education (N = 11)	
		Fall	Winter	Fall	Winter	Fall	Winter	Fall	Winter
DIBELS ORF	Mean (SD)	97 (40)	111 (41)	102 (38)	116 (38)	47 (27)	57 (18)	93 (41)	103 (46)
AIMSweb ORF	Mean (SD)	94 (41)	116 (41)	100 (39)	122 (38)	45 (27)	64 (28)	85 (42)	102 (45)
Maze	Mean (SD)	12 (6)	15 (7)	13 (6)	16 (7)	6 (3)	8 (3)	9 (8)	13 (5)
MAP	Mean (SD)	195 (15)	201 (14)	198 (13)	204 (12)	170 (9)	182 (14)	191 (18)	197 (14)
NECAP	Mean (SD)	350 (10)		351 (9)		336 (8)		350 (7)	

Correlation Analysis

To determine the relationships among the measures of reading, Pearson Product Moment correlations were calculated. These results are presented in Table 3. Significant correlations were found among each measure of reading at each point in time ($p \leq .01$). Not surprisingly, the strongest correlation was between the two measures of oral reading fluency ($r = .972$). The next highest correlations were between MAP and NECAP ($r = .819$), ORF and Maze ($r = .812$), and MAP and ORF ($r = .809$). The lowest correlation was between Maze and NECAP ($r = .621$). In the fall, ORF correlated more strongly with the MAP than the maze, while the reverse was true in the winter. Correlations were also calculated within each measure between fall and winter administrations. ORF had the highest correlations between the fall and winter measures ($r = .952$), followed by the MAP ($r = .872$) and Maze ($r = .746$), respectively.

In addition to performing correlation analyses on the overall sample, the data were disaggregated by Tier 1 and Tiers 2 and 3, and a separate analysis was conducted in order to analyze differences between the two groups (Table 4). For students in Tier 1, or general education, correlations among each measure of reading remained significant ($p \leq .01$), although they were not as high as they were for the overall sample. Overall, all the measures appeared to correlate strongly with each other and themselves when given more than once. The consistently high correlations suggest that all of the measures tapped into a common skill area: reading.

Table 3

*Intercorrelations Between Reading Measures in September and January,**Overall Sample (N = 137)*

	September				January				
	DIBELS ORF	AIMSweb ORF	Maze	MAP	DIBELS ORF	AIMSweb ORF	Maze	MAP	NECAP
September									
DIBELS ORF	-	.972	.761	.809	.952	.945	.778	.767	.718
AIMSweb ORF		-	.760	.807	.947	.949	.773	.762	.714
Maze			-	.683	.763	.740	.746	.652	.621
MAP				-	.806	.803	.743	.872	.819
January									
DIBELS ORF					-	.961	.812	.765	.705
AIMSweb ORF						-	.793	.769	.678
Maze							-	.708	.713
MAP								-	.777
NECAP									-

Table 4

Intercorrelations Between Reading Measures in September and January for Students in Tier 1

(*N* = 115)

	September				January				
	DIBELS ORF	AIMSweb ORF	Maze	MAP	DIBELS ORF	AIMSweb ORF	Maze	MAP	NECAP
September									
DIBELS ORF	-	.967	.730	.779	.945	.942	.753	.726	.663
AIMSweb ORF		-	.729	.778	.937	.943	.742	.725	.666
Maze			-	.624	.723	.706	.723	.591	.579
MAP				-	.765	.758	.724	.852	.785
January									
DIBELS ORF					-	.961	.785	.727	.655
AIMSweb ORF						-	.769	.715	.617
Maze							-	.680	.693
MAP								-	.725
NECAP									-

Regression Analysis

Several regression models were computed to learn which reading scores best predicted later scores. Step-wise with removal multiple regression analyses were conducted with the winter MAP scores as the dependent variable and all other reading scores as the independent variables (Table 5). The results indicated that the MAP assessment in the fall best predicted MAP scores in the winter ($p < .001$), and that oral reading fluency accounted for significant variance in MAP scores beyond performance on the fall MAP ($p < .05$). In a second analysis, in which the fall MAP scores were removed, results indicated that oral reading fluency was a significant predictor of winter MAP performance ($p < .001$). These results are presented in Table 6.

Table 5

Summary of Multiple Regression Analysis for Variables Predicting Winter MAP Scores

($N = 137$)

Variable	B	SE B	β
Fall MAP	.652	.063	.728
DIBELS ORF	.060	.024	.178

Notes: $R^2 = .77$ ($p < .05$).

Table 6

Summary of Multiple Regression Analysis for CBM Variables Predicting Winter MAP Scores

($N = 137$)

Variable	B	SE B	β
DIBELS ORF	.260	.019	.767

Notes: $R^2 = .59$ ($p < .001$).

A multiple regression analysis also was performed to determine which measures of reading best predicted NECAP scores (Table 7). Results showed that MAP was the best predictor, accounting for 67% of variance in NECAP scores ($p < .001$). In order to determine whether ORF contributed unique variance to NECAP scores when MAP data were not present, an additional multiple regression analysis was conducted. As shown in Table 8, when MAP was removed from the equation, ORF was also found to predict NECAP scores, with a statistically significant amount of variance accounted for by the ORF scores. ($p < .001$).

Table 7

Summary of Multiple Regression Analysis for Variables Predicting NECAP Scores (N = 137)

Variable	B	SE B	β
Fall MAP	.536	.032	.819

Notes: $R^2 = .67$ ($p < .001$).

Table 8

Summary of Multiple Regression Analysis for CBM Variables Predicting NECAP Scores

(N = 137)

Variable	B	SE B	β
DIBELS ORF	.178	.015	.718

Notes: $R^2 = .52$ ($p < .001$).

An additional regression model was used to determine whether the types of instruction that students received predicted reading scores. When type of instruction was included as a separate independent variable in the regression equation, results showed that the level of instruction (Tier 1, Tier 2, Tier 3, or special education) did not contribute unique variance to

performance on the NECAP or winter MAP ($p \geq .05$). Even though students in Tiers 2 and 3 had the lowest reading scores across the board, they were not uniquely identified by these measures.

Finally, a regression analysis was constructed which examined the predictive values of the reading scores for student subgroups. The overall student sample was divided by instructional group, and a multiple regression analysis was performed on the combined sample of Tier 2, Tier 3, and special education students ($N = 22$; Table 9). This analysis indicated that the MAP best predicted NECAP scores for the general population of students, as well as those students receiving additional reading support ($p < .001$). Again, when MAP was removed from the equation, curriculum-based measures also were significant predictors of performance on the NECAP for at-risk readers ($p < .001$). These results are displayed in Table 10.

Table 9

Summary of Multiple Regression Analysis for Variables Predicting NECAP Scores in Sample of Lower Performing Readers ($N = 22$)

Variable	B	SE B	β
Fall MAP	.509	.072	.846

Notes: $R^2 = .72$ ($p < .001$).

Table 10

Summary of Multiple Regression Analysis for CBM Variables Predicting NECAP Scores in Sample of Lower Performing Readers ($N = 22$)

Variable	B	SE B	β
DIBELS ORF	.204	.034	.804

Notes: $R^2 = .65$ ($p < .001$).

Overall, the results indicated that both MAP and ORF scores predicted students' later reading scores. MAP was the best predictor of itself as well as of NECAP scores, however, when MAP scores were not in the regression model, ORF predicted later reading scores at statistically significant levels.

Discussion

There is a strong research base supporting the technical adequacy, predictive validity, measurement sensitivity, and instructional utility of curriculum based measurement (CBM) in reading. CBM has been established as an efficient and effective method for universal screening and progress monitoring. Increasingly, the Measures of Academic Progress (MAP) are being used for the same purposes. This study examined the extent to which these measures predict student reading performance over time, and according to whether students received more intensive reading instruction.

Current data were consistent with previous research that showed a positive correlation between CBM and state achievement measures, and between MAP and state achievement scores. This study expanded the current literature by measuring the correlations and predictive features between CBM and MAP scores. A strong relationship between the two assessment tools was found, providing evidence for their concurrent validity.

When the same analyses were performed on the disaggregated sample which included only Tier 1 students, MAP and CBM measures were found to be less predictive for this group of students. Although a very preliminary finding, this suggests that each measure has more predictive power for students who are at risk for reading problems than for the general student population.

The present study also revealed that MAP in reading was the best predictor of performance on mid-year MAP and NECAP scores. Such a result is not surprising because MAP has previously been shown to have adequate reliability and validity when compared with other measures and should correlate with itself. In addition, the features of MAP and NECAP share similarities that CBM do not. Specifically, both the MAP and NECAP contain multiple-choice items in specific areas of reading, including word identification, vocabulary, and understanding and analysis of text. In contrast, CBM measures are timed assessments that measure reading fluency. The MAP and NECAP are administered in a group setting, while ORF assessments are administered to students individually. The similarities and differences in the nature of these assessments may explain the significant findings.

While MAP was the most predictive instrument, measures of oral reading fluency also yielded significant results, adding to the evidence that CBM is a strong predictor of future reading performance, as measured by high stakes tests. These findings could be important for school personnel who make decisions about what universal reading assessments to use to identify students who need additional instruction. For those New England schools which already use MAP as part of their assessment system, these scores should predict how their students will do on the NECAP reading portion. Similarly, for schools using NECAP and CBM, these measures correlate and CBM ORF scores should predict students' later NECAP scores.

An interesting finding was that type of instruction (participation in Tier 1, Tier 2, Tier 3, or special education) did not significantly predict NECAP or MAP performance. In other words, these measures did not uniquely identify students as being at one or another instructional level. This finding could be explained by a number of factors. First, students in this sample generally performed higher on these reading measures than the state average, and a relatively low

percentage of students were receiving supplemental reading instruction through Tier 2 or Tier 3. It may be that the reading scores between instructional groups were not as discrepant as would be typically found in a more heterogeneous sample of readers. Secondly, although the schools that the students attended provided special reading instruction for students who scored low on the fall MAP, the nature of their reading skills at each assessment period was not exhibited in the data. It is possible that this finding indicates a need to review the Tier 2 and 3 reading instruction and confirm whether it is suited to each student's individual need. Another possibility in analyzing this finding is that results may be different if the subgroups of students were more equivalent in size. A larger sample of Tier 2 and Tier 3 students would enhance the statistical properties of the analysis. Finally, this finding may reflect a limitation of the instruments in detecting student placement. The non-specific finding also points to the importance of regularly monitoring student progress with performance criteria, to help ensure that student needs are not missed.

This final point is particularly important in light of the result that ORF best predicted itself at levels higher than any of the other measures. When the Fall and Winter data were compared, students scores on the ORF tasks were more highly correlated than their scores on the MAP. This result points to a key difference between the CBM and MAP items. MAP items are computer adaptive and students do not complete the same items each time. By comparison, the ORF items are highly similar at each testing period. On the DIBELS ORF, students complete similar, but not identical, reading passages at each benchmark screening (e.g., Fall, Winter, and Spring). On the AIMSweb ORF, the reading passages are identical at each screening. The greater similarity of the ORF items at each screening period, as well as the timed nature of the task, makes ORF items able to be used as regular progress measures in ways that MAP items

cannot. Thus, it is likely that the most accurate information about students' reading progress over time will be gained from ORF progress measures.

These results have implications for educators and administrators who make decisions about how to collect student performance data and select those students in need of more intensive reading intervention. One factor to consider is the instructional value that is gained by using one measure over the other. The MAP reading items cover the broad range of subskills required for effective reading at different grade levels, including phonological awareness, vocabulary, and comprehension. Reading scores on the MAP are broken down in four different learning areas, and provide more construct-specific information than CBM as a screening instrument. Nonetheless, the MAP is not timed and thus cannot offer data concerning a student's reading fluency which has been found to be a major predictor of later reading skills. This study did not consider the accuracy with which MAP scores indicated students' reading instruction needs, a factor needing attention in follow up studies.

A second factor that school personnel who select screening assessments need to consider is the time required to administer these assessments, and the instructional time that is lost while students are being tested. CBM in reading can be administered to one student in approximately five minutes. The MAP in reading is administered in a group setting in 50 minutes, although most students typically finish the test in 20-40 minutes (NWEA, 2009). Although the MAP can be administered to large groups of students at one time, it requires more time of the students than CBM does. Data from studies of RTI implementation have shown that students with reading difficulties can make gains in as little as 20-30 minutes per day (Linan-Thompson & Hickman-Davis, 2002). It could be that the best use of student learning time is gained by using CBM as a screening measure because it leaves more time for students who need such interventions.

Finally, the financial cost of assessment is an important consideration for school districts throughout the country. In the school district in which this study took place, the cost per pupil, per year is \$3.00 for AIMSweb and \$12.50 for the MAP. For the 150 students who participated in this study, the total cost for CBM was \$450 and for MAP was \$1875. These costs are not insignificant when the volume of all students in a district is taken into consideration. The cost difference between CBM and MAP must be considered in light of the finding that there were no significant differences in how well each one predicted student scores on the state-mandated assessment. If CBM scores can predict student reading performance as well as MAP, the less costly option may be important for school personnel to consider.

Given these data, a combination of CBM and MAP may be tailored to provide an assessment package that is predictive of future performance, instructionally informative, efficient, and affordable. It may be that each assessment does not need to be administered to every student, or at each benchmark period, but are utilized more strategically. For example, MAP may be a good option as a universal screening instrument at the start of each school year, as it clearly predicts student performance. CBM may present a better option for benchmark assessments, which are given three times per year, and progress monitoring, which is administered at least monthly to students at risk for reading problems. The National Center on Response to Intervention publishes technical adequacy standards for progress monitoring tools, which include evidence of reliability and validity, sensitivity to improvement, and adequate data on weekly growth rates (National Center on Response to Intervention, 2010). According to these standards, CBM meets criteria as an appropriate progress measure. Even though MAP provides student growth data when it is given three times per year, it does not possess the same technical properties as a progress monitoring instrument. For school districts that do not use MAP, these

data suggest that curriculum-based measurement in reading can be used to identify students at risk for reading problems with accuracy that matches MAP and for less time and money.

Limitations and Future Research

Although this study resulted in positive findings, some limitations should be considered. First, it was based on a convenience sample from two schools in an affluent district in the Northeast. Secondly, it was limited to third grade students. Further research could expand on these results by including students in early elementary grades, to address which instruments are most predictive of reading performance by grade level and over longer periods of time. Replicating this study with a more diverse population of readers could also help identify whether and how MAP and CBM identify specific instructional needs according to a student's level of reading performance.

Third, another limitation to this study was the very small sample of students participating in Tier 2 and Tier 3 instruction. As a result of the small numbers participating in these services, certain statistical analyses could not be conducted on this group, even when the students in Tiers 2 and 3 were combined. More research is needed to explore comparisons of the different reading measures by instructional group. Specifically, if one measure better predicts performance for lower-performing readers, this has implications for the selection of benchmark screening and progress monitoring instruments. Furthermore, the data from this analysis were not able to detect placement in one or another instructional group. With a more robust sample of Tier 2 and Tier 3 students, further studies could examine the classification accuracy of each reading measure, by comparing cut scores for basic proficiency to external indicators of risk.

Fourth, and finally, this study did not examine the extent to which MAP and CBM data are useful to teachers in targeting and differentiating instruction. The ultimate value of these

tools is in how well they lead to data-based decision making and more effective interventions for students. MAP score reports provide individual student and class-wide data in four areas of reading. CBM data are reported as raw fluency scores that are compared to grade-level national benchmarks. Future research should explore the extent to which these outcomes are differentially beneficial to teachers in guiding their instruction.

Summary

Results of this study suggest that there are multiple methods for identifying students at risk for reading problems and predicting student performance on high stakes tests. Factors such as accuracy, time, cost, and instructional utility should be considered when making assessment decisions. Findings showed that different widely used measures including CBM and MAP predicted student reading performance in highly similar ways. This result leaves schools with important decisions concerning how best to screen student reading skills so that timely interventions can be provided. Most important is how a school uses the data that has been collected, so that students are provided the instruction and supports they need to be successful in school.

References

- AIMSweb. (2008). Retrieved April 28, 2009 from: <http://www.aimsweb.com/>.
- AIMSweb. (2010a). AIMSweb Growth Table, Maze – Comprehension. Retrieved May 10, 2010 from
<https://aimsweb.pearson.com/Report.cfm?inc=rReport18.cfm&ClearSchool=1&Aggregate=1&CFID=8289266&CFTOKEN=35537513&jsessionid=b230357ac91dcfe462f737142a417a201154>
- AIMSweb. (2010b). AIMSweb Growth Table, Reading – Curriculum Based Measurement. Retrieved May 10, 2010 from
<https://aimsweb.pearson.com/Report.cfm?inc=rReport18.cfm&ClearSchool=1&Aggregate=1&CFID=8289266&CFTOKEN=35537513&jsessionid=b230357ac91dcfe462f737142a417a201154>
- Ardoin, S. P. & Christ, T. J. (2008). Evaluating curriculum-based measurement slope estimates using data from triannual universal screenings. *School Psychology Review, 37*, 109-125.
- Ardoin, S. P., Witt, J. C., Suldo, S. M., Connell, J. E., Koenig, J. L., Resetar, J. L., et al. (2004). Examining the incremental benefits of administering a maze and three versus one curriculum-based measurement reading probes when conducting universal screening. *School Psychology Review, 33*, 218-233.
- Baker, S. K., Smolkowski, K., Katz, R., Fien, H., Seeley, J. R., Kame'enui, E. J., et al. (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review, 37*, 18-37.
- Begeny, J. C. & Martens, B. K. (2006). Assisting low-performing readers with a group-based reading fluency intervention. *School Psychology Review, 35*, 91-107.

- Brown-Chidsey, R., & Steege, M. W. (2005). *Response to intervention: Principles and strategies for effective practice*. New York: Guilford.
- Cronin, J. (2004). Aligning the NWEA RIT Scale with the Maine Educational Assessments (MEA). Retrieved April 1, 2009 from <http://www.nwea.org/assets/research/state/Maine%20Alignment%20Report%204.18.07.pdf>
- Delong, S. E. H. (2007). A study of the relationship between the utilization of NWEA MAP testing and student achievement. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 68 (10-A), 42-68.
- Deno, S. L. (2005). Problem-solving assessment. In R. Brown-Chidsey (Ed.), *Assessment for intervention: A problem solving approach* (pp. 10-40). New York, NY: The Guilford Press.
- Deno, S. L., Reschly, A. L, Lembke, E. S., Magnusson, D., Callender, S. A., Windram, H., et al. (2009). Developing a school-wide progress-monitoring system. *Psychology in the Schools*, 46, 44-55.
- Good, R. H., Gruba, J., & Kaminski, R. A. (2002). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an outcome-driven model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 699-720). Bethesda, MD: The National Association of School Psychologists.
- Good, R. H., Wallin, J., Simmons, D. C., Kame'enui, E. J., & Kaminski, R. A. (2002). System-wide Percentile Ranks for DIBELS Benchmark Assessment (Technical Report 9). Eugene, OR: University of Oregon.
- Graney, S. B., Missall, K. N., Martinez, R. S., & Bergstrom, M. (2009). A preliminary investigation of within-year growth patterns in reading and mathematics curriculum-based measures. *Journal of School Psychology*, 47, 121-142.

- Hall-Michalcewiz, J. (2008). Predicting Reading and Mathematics Delaware State Testing Program performance using the measures of academic progress. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 69 (4-A), 13-39.
- Hintze, J. M. & Silbergitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high-stakes testing. *School Psychology Review*, 34, 372-386.
- Individuals with Disabilities Education Improvement Act. (2004). Public Law 108-446 (20 U.S.C. 1400 *et seq.*).
- Keller-Margulis, M. A., Shapiro, E. S., & Hintze, J. M. (2008). Long-term diagnostic accuracy of curriculum-based measures in reading and mathematics. *School Psychology Review*, 37, 374-390.
- Kingsbury, G.G. (2002). An empirical comparison of achievement level estimates from adaptive tests and paper-and-pencil tests. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Linan-Thompson, S. & Hickman-Davis, P. (2002). Supplemental reading instruction for students at risk for reading disabilities: Improve reading 30 minutes at a time. *Learning Disabilities Research & Practice*, 17, 242-251.
- McGlinchey, M. T. & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review*, 33, 193-203.
- Measured Progress. (2009). New England Common Assessment Program 2008-2009 Technical Manual. Retrieved May 11, 2010 from <http://www.maine.gov/education/necap/documents/2008-09NECAPTechReportMASTERDOCUMENT.pdf>

- National Center on Response to Intervention. (2010). Progress Monitoring Tools Chart: Reading and Math. Retrieved May 14, 2010 from http://www.rti4success.org/chart/progressMonitoring/PMTToolsChart_04-20-10a.pdf
- New England Common Assessment Program. (2010). Fall 2009 NECAP Tests State Summary. Retrieved May 10, 2010 from <http://www.maine.gov/education/necap/0910necapscores/0910statewide.pdf>
- Northwest Evaluation Association. (2004). Reliability and validity estimates: NWEA Achievement Level Tests and Measures of Academic Progress. Retrieved April 1, 2009 from <http://www.nwea.org/assets/research/NWEA%20Reliability%20&%20Validity.pdf>
- Northwest Evaluation Association. (2008). RIT Scale Norms For Use With Measures of Academic Progress. Lake Oswego, OR: Northwest Evaluation Association.
- Northwest Evaluation Association. (2009). Technical Manual for Measures of Academic Progress and Measures of Academic Progress for Primary Grades. Lake Oswego, OR: Northwest Evaluation Association.
- Northwest Evaluation Association. (n.d.). RIT Charts. Retrieved May 8, 2009 from http://www.nwea.org/assets/documentlibrary/Reading_single.pdf
- Shinn, M. R. (2002). Best practices in using curriculum-based measurement in a problem-solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 671-697). Bethesda, MD: The National Association of School Psychologists.
- Silbergliitt, B., Burns, M. K., Madyun, N. H., & Lail, K. E. (2006). Relationship of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. *Psychology in the Schools, 43*, 527-535.

- Stage, S. A., & Jacobsen, M. D. (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review, 30*, 407-419.
- Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42*, 795-819.
- Stecker, P. M., Lembke, E. S., & Foegen, A. (2008). Using progress-monitoring data to improve instructional decision making. *Preventing School Failure, 52*, 48-58.
- U.S. Department of Education, Office of Elementary and Secondary Education. (2002). *No Child Left Behind: A desktop reference*. Washington, DC: Author.
- Wiley, H. I. & Deno, S. L. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education, 26*, 207-214.

Biography of the Author

Kristina Andren was born in Rockport, Maine, and received her high school diploma in Camden, Maine. She graduated with her B.A. in Sociology/Anthropology from Carleton College in 1997, and received her M.A. in Educational Psychology from Boston College in 2000.

Kristina earned her M.S. in School Psychology from the University of Southern Maine in 2005.

She is a Nationally Certified School Psychologist, and has worked in several public schools. She is a candidate for the Psy.D. degree in School Psychology from the University of Southern Maine in August, 2010.