

# Predictive Validity of Curriculum-Based Measures in the Reading Assessment of Students who are English Language Learners

Paul Muyskens, Joseph Betts, Matthew Y. Lau, & Doug Marston  
*Minneapolis Public Schools*

The inclusion of English Language Learners as a subgroup in the No Child Left Behind legislation has leant additional importance to the need for valid and efficient measures of reading for students whose first language is not English. This study examines the use of Curriculum-Based Measurement (CBM) reading fluency as a predictor of later reading performance on state accountability tests for fifth grade ELL students. The findings of this study indicate that CBM is a significant predictor of later performance on tests for accountability for ELL students as a whole, and for the individual language groups of Spanish, Hmong, and Somali. Implications for these findings are discussed.

One of the greatest contemporary opportunities and challenges in America is the education of culturally and linguistically diverse students whose first language is not English. The general term, English Language Learner (ELL), is used to describe a group of students who are non-native English speakers and who score low on a measure of English language proficiency. The No Child Left Behind Act (NCLB, 2001) refers to this group as students with limited English proficiency, and defines them as students who belong to one of the following categories:

- a) Was not born in the United States or speaks a native language other than English;
- b) Is a Native American, Alaska Native, or native resident of outlying areas and comes from an environment where language other than English has had a significant impact in the individual's level of English language proficient, or
- c) Is migratory, speaks a native language other than English, and comes from an environment where language other than English is dominant, or
- d) May be unable, because of difficulties in speaking, reading, writing, or understanding the English language, to score at the proficient level on state assessments of academic achievement, learn successfully in classrooms where the language of instruction is English, or participate fully in society.

## The Number and Achievement of ELL Students

Estimates during the 1990s indicated there was an increase of about 1 million ELL students, which resulted in about 5.5% of all students being served in public schools speaking English as a non-primary language (National Research Council, 1997). Kindler (2002) estimated that the number had climbed even higher during the 1999-2000 year with an estimated 4.4 million ELL students in public schools (about 9% of all students in public education). The US Census Bureau (2000) estimated that about 18% of children between the ages of 5 and 17 speak a language other than English as their primary language in the home.

Unfortunately, the educational achievements of these ELL students have not increased as dramatically as their numbers. While there are differences between home-language groups, studies have found that ELL students in general are lower performing on tests of academic achievement when compared to their English-speaking peers (August & Hakuta, 1997; Moss & Puma, 1995). These types of outcomes

for ELL students, which may be related to the linguistic complexity of the items included in the test, have been found in other research on both mathematics and reading (Abedi, 2002; Abedi, Lord & Hofstetter, 1998; Cocking & Chipman, 1988; Liu, Anderson, & Thurlow, 2000; Thomas & Collier, 1997). Hopstock and Stephenson (2003) found that when taking state required tests for graduation students with limited English proficiency were much more likely to fail than were the student group as a whole (50% vs. 24%). August and Hakuta (1997) also found higher dropout rates for ELL students. The need for ELL students to accelerate their academic achievement has received new emphasis with the implementation of the No Child Left Behind Act of 2001. Under this law, all children, including the specific subgroup ELL students, are expected to reach a proficient level in reading and math each and every year starting at the third grade.

### **Curriculum-Based Assessment of Reading**

Curriculum-Based Measurement (CBM) as a measure of oral reading fluency has long been shown to be an efficient and valid measure of academic progress for English-speaking general education and special education students. For example, in an examination of 11 studies looking at various measures of the reliability of CBM Marston (1989) found a mean reliability rating of .91. The validity of CBM measures has also been established through numerous studies showing a strong relationship between measures of oral reading fluency and a variety of standardized reading assessments (Fuchs & Deno, 1981; Fuchs & Fuchs, 1999; Shinn, Good, Knutson, Tilly & Collins, 1992). It has also been shown to be a good measure of reading comprehension across grades (Kranzler, Miller & Jordan, 1999; Shinn et al., 1992) and with specific subgroups (Deno, Fuchs, Marston, & Shin, 2001; Hintze, Callahan, Matthews, Williams, & Tobin, 2002).

One recent development in CBM research is to examine the relationship between oral reading fluency and student performance on state accountability tests (Deno, 2003). Several studies have reported moderate to high correlations between CBM and state assessments (e.g., Good, Simmons, & Kame'enui, 2001; McGlinchey & Hixson, 2004; Pearce & Gayle, 2009; Sibley, Biwer, & Hesch, 2001; Stage & Jacobsen, 2001). In addition, the validity of using benchmark goals or cut scores on CBM measures to predict pass and fail rates on high-stakes assessments has also been supported (Hintze & Silbergitt, 2005).

Despite the extensive study of CBM, and its widespread use, published research on the use of CBM for ELL students is limited. Baker and Good (1995) investigated the reliability and validity issues of CBM in English with bilingual Hispanic students. They concluded that CBM was as reliable and valid for Hispanic bilingual students as for their English speaking peers. The convergent and discriminant data from this study provided further support for CBM as a measure of English proficiency in reading and comprehension for bilingual students.

In another relevant study that included Hispanic and Caucasian youth, Klein and Jimerson (2005) examined the potential bias of oral reading fluency as a predictor of future reading proficiency, considering gender, ethnicity, home language, and socioeconomic status. Analyses of longitudinal data from 398 students enrolled grades 1-3 revealed consistent intercept bias effects for the combination of ethnicity and home language factors at grades one, two, and three. Specifically the results indicated that, when using a common regression equation, oral reading fluency probes overpredicted the reading proficiency (as measured by the *Stanford Achievement Test – Ninth Edition (SAT-9) Total Reading*) of Hispanic students whose home language is Spanish and underpredicted the reading proficiency of Caucasian students whose home language is English. More recent research development in this area involves investigations using nonsense word fluency (NWF) to predict reading performance. Studies have found early literacy skills such as alphabetic understanding and phonological recoding ability measured by NWF have a significant predictive value on real-word reading and reading performance on standardized measures such as state accountability tests (e.g., Vanderwood, Linklater, & Healy, 2008 and Fien et al., 2008). With few studies examining CBM among ELL students, further research is warranted.

The ability to evaluate and predict reading ability of students can be particularly challenging with students whose primary language is not English. It is not uncommon for school-based professionals to

question the validity of CBM measures with ELL students by pointing out that reading fluency does not necessarily correspond to comprehension. These professionals maintain that ELL students can at times decode words without having the contextual or topical knowledge needed to understand what they are actually saying. This may result in fluency scores that indicate mastery, while the student does not really understand what they have read, and thus they are likely to suffer on comprehension based assessments. Based upon this argument, some ELL advocates promote a portfolio or theme-based assessment system. For instance Sudweeks, Glissmeyer, Morrison, Wilcox, & Tanner (2004) recommended oral retellings to assess ELL students' reading comprehension. On the other hand, in a study of 66 third grade students in Pacific Northwest, Hamilton & Shinn (2003) reported that "word callers" (students who can read fluently but do not understand what they read) scored fewer correct words per minutes and earned significantly lower scores on comprehension measures. Understanding that the participants of this study were not ELL students, it seems evident that students who read poorly and without fluency are also likely to comprehend poorly.

The complexity of the issues related to language acquisition and reading fluency and comprehension are also related to the nature and the characteristics of the home language spoken by the student. For example, Spanish and English are phonetic-based languages that share many underlying cognates, or common word origins, the teaching of which can be used as a strategy to enhance vocabulary development (Carlo, August, McLaughlin, Snow, & Dressler, 2004). Hmong, however, has distinctly different grammatical and phonemic usage than English and belongs to a group of languages, often referred to as the Miao-Yao languages, spoken in Southeast Asia and Southern China. Unlike English, Hmong is mostly a monosyllabic language. Moreover, it is a tonal language, meaning pitch variations are used to signal a difference in meaning among words. On the other hand, Somali language is a member of the Cushitic languages spoken mostly in Somalia and nearby Djibouti, Ethiopia, and Kenya. While this language has a tonal component, there is also significant overlap with the English alphabet. Both of these cultures have an emphasis on oral tradition.

### **The Present Study**

The purpose of this study was to investigate the concurrent and predictive validity of a CBM measure of oral reading fluency for ELL students. One objective was to provide validity evidence for CBM as a predictor of a state-mandated proficiency level assessment of reading for ELL students. These results would be used to validate the ability to make inferences from ELL student's CBM scores to reading in general, and also in predictions of proficiency status on a high-stakes assessment. This study focused on three distinct ELL populations represented by their home languages, which are Spanish, Hmong, and Somali.

The findings of this study potentially add an important piece to the CBM literature regarding the use of oral fluency measures on students who speak a native language that is very different from the primarily phonetic-based English and Spanish languages. All three of these language groups, along with dozens of others, are commonly grouped together for instruction; yet, their backgrounds and instructional needs may be very divergent. It would be helpful to know if we can use a common method scaled on a common metric for monitoring their progress in reading. This would also be particularly useful if the unit of measurement could be used as a formative assessment. CBM has been used as one of the tools that can provide efficient and reliable data for this purpose with English speaking students; however, it is necessary to determine if CBM is valid in this role with ELL students.

## **METHOD**

### **Participants**

This study took place during the 2003-2004 school year in a large urban school district located in the Midwestern United States. Participants were fifth-grade students who had received an ELL status in the district and reported their home language was Spanish, Somali, or Hmong (N=1,529). The rationale for not including all non-English languages was due to the fact that the selected language groups make

up the great majority of ELL students (around 88%). The remaining ELL students comprised 72 other languages and were excluded due to the small number of students in each language group. Due to mobility and absences 1,205 students (78% of possible participants) were measured on both CBM and the Minnesota Comprehensive Assessment (MCA). Table 1 delineates the demographic percentages in the targeted population and the corresponding percentages in the sample. The sample appears to be a representative sample and it is assumed that any loss of students was the result of random processes that did not relate to systematic procedures or the students actual reading ability.

**TABLE 1:** *Demographic Information*

Demographic Information		Population (N = 1,529)	Sample (N = 1,205)
Gender	Male	52%	52%
	Female	48%	48%
Home Language	Spanish	48%	46%
	Hmong	42%	44%
	Somali	10%	10%
SES Proxy	Eligible for Free/Reduced Lunch	95%	94%
	Not Eligible for Free/Reduced Lunch	5%	6%

## Materials & Procedure

The CBM oral reading fluency measures used in this study were grade-level passages drawn from the district basal reading text, *Invitations to Literacy* which was published by Houghton Mifflin in 1999. Passages were selected so as to represent the subjects, authors and styles found within the curriculum. Readability levels and pilot studies with district students were also used to ensure that the passages were of similar difficulty. Completion of CBM data collection involved the students reading three different passages for one minute each and the calculation of the median number of words read correct for data analyses. All data were collected following the standardized procedures outlined in the district manual, "Performance Assessment of Reading in the Problem Solving Model" (Minneapolis Public Schools, 2003). These CBM administration and scoring procedures have been described by Marston and Magnusson (1985), and Shinn (1989). All participants were administered a CBM measure in September as part of their school's building-wide continue progress monitoring system.

The Minnesota Comprehensive Assessment (MCA) is a measure of reading proficiency (Minnesota Department of Education, n.d.) On the fifth grade test students read passages and answered multiple-choice and constructed response items requiring them to write short answers related to the purpose of the passage and the main idea. In addition, the student must be able to synthesize information in the story to develop conclusions and make inferences. All items were designed to be aligned with the Minnesota High Standards, and are scored. In order to minimize variance all items are scored by the test vendor.

Two types of scores are derived from the MCA, a level score and a standard score. The level scores range from 1 to 5. A student scoring at level 1 is described as having gaps in knowledge and skills. Those scoring at level 3 are described as having solid on-grade level skills, while those at level 5 are considered to have superior performance beyond grade level. Level 3 is the level where students are considered to be

proficient, and corresponds to a standard score of 1420 or a raw score of 40. The raw scores on the MCA ranged from 0 to 58, with a mean of 45.81, a standard deviation of 10.34, and a reliability coefficient of .92 (Minnesota Department of Education, n.d.). The difficulty of the MCA was determined through use of the Degrees of Reading Power (DRP). According to MDE, the 5<sup>th</sup> grade passages used on the MCA have an average DRP of 54, which is the level of a typical fifth grade textbook.

For this study, the CBM data were collected shortly after school began in the fall of 2003 as part of the building-wide progress monitoring system. The MCA was administered in late April. Therefore there was about a six-month intervening period between the CBM and MCA.

## Data Analysis

Data analyses examined whether a CBM measure at the beginning of the school year could be helpful in predicting scores for ELL students toward the end of the year on a state-mandated, high-stakes, standards-based assessment in reading. Simple regression analyses were completed to address this question (Neter, Kutner, Nachtsheim, & Wasserman, 1996). Logistic regression models (Hosmer & Lemeshow, 1989) were used to assess the predictive validity of using CBM to estimate proficiency on the MCA. The MCA score of 1420 is the state-mandated cut-score for proficiency levels. Thus any score greater than or equal to 1420 was coded as 'pass' and any score less than 1420 was coded as 'fail'. It should be recognized that the student does not actually fail the test but rather fails to get a score that indicates proficiency in the area of reading. This model allows for the computation of diagnostic accuracy statistics. The basis for this is the generation of a predicted p-value from the model (Hosmer & Lemeshow, 1989; Neter, Kutner, Nachtsheim, & Wasserman, 1996). For this analysis, we utilized a p-value of 0.5 as the cut-point for classification. Thus, students with a predicted p-value of greater than 0.5 were predicted to pass, while those less than 0.5 were predicted to fail.

Furthermore, in the analyses examined the following eight measures of diagnostic accuracy (Swets, Dawes & Monahan, 2000): total correct classifications, total incorrect classifications, sensitivity, specificity, false positives, false negatives, positive predictive power and negative predictive power. The total correct classification is defined as the number of students who were predicted to pass and did actually pass plus the number of students predicted to fail who actually did fail. This sum is then divided by the number of students in the study. It is also obvious that one minus the sum gives a measure of the misclassification of the model.

Sensitivity is defined as the conditional probability of getting a logistic regression p-value of greater than 0.5 given the student actually got a passing score. The basic interpretation is that it tells us the percent of students we predict to pass out of the subset of students who actually did pass. More directly, it computes the probability that the CBM scores correctly identified a student as passing from the subset of all students who actually did pass the MCA proficiency level in reading. This expresses how sensitive the scores from the CBM are at identifying students who will make the proficiency standard.

Specificity is defined as the conditional probability of getting a logistics regression p-value of less than 0.5 given the student actually did fail to reach the 1420 cut-point for proficiency. The basic interpretation is that it tells us the percent of students we predicted to fail out of the subset of students who did actually fail. More directly, it computes the probability that a CBM score will correctly identify a student as not meeting the proficiency level out of the subset of students who actually did not meet the proficiency level. The probability expresses the ability of the CBM scores to specify those who are unable to meet the proficiency level.

There are other methods of defining sensitivity and specificity. Many researchers will reverse the definitions of sensitivity and specificity to highlight the ability of a procedure to be sensitive to deficits in some area. This would be akin to reversing our definitions and would highlight an attempt to be sensitive to reading problems in ELL students. Under these circumstances, one could simply reverse the identifications in the above definitions. This is pointed out to alleviate any potential confusion this might engender with respect to the literature on diagnostic accuracy measures of deficits or identifiable disabling conditions.

In addition, analyses examined the amount of error associated with the classification predictions, by analyzing the extent to which incorrect classifications were observed. Besides using one minus the correct classification, it is also possible to specify the number of false positive and false negative identifications. False positives are defined as those students predicted to pass ( $p$ -value  $> 0.5$ ), who did not actually pass. This is similar to a Type I error. False negatives are defined as those students predicted to fail ( $p$ -value  $< 0.5$ ), who did actually pass. This is similar to a Type II error.

The positive and negative predictive power gives us an estimate of how well the CBM scores predict passing or not passing status. The positive predictive power provides the conditional probability that given a person is predicted to pass the reading proficiency level ( $p < 0.5$ ) they actually do pass the proficiency level. This provides a relative likelihood of the student actually passing the proficiency level given the fact that they are predicted to pass based on the CBM scores. The negative predictive power is the conditional probability that a given student is predicted not to pass the MCA, how likely it is that they actually do not pass. These measures should not be confused with sensitivity and specificity which are conditional probabilities computed over different base groups.

Finally, analyses of the logistic regression were considered by using a receiver operating characteristic (ROC) curve analysis. This provides a measure of discrimination in using the CBM scores to classify later MCA proficiency (Hanley & McNeil, 1982; Hosmer & Lemeshow, 1989). Each of the three language groups were compared individually to examine whether the CBM has any differential functioning between the groups. Analyzing the area under the curve (AUC), statistic from the ROC analysis and using the predicted values from the logistic regression give us a measure of the ability of the CBM to discriminate between pairs of individuals. AUC results of greater than or equal to 0.9 are considered to provide outstanding discrimination, values between 0.8 and 0.9 are considered excellent, and values between 0.7 and 0.8 are considered acceptable (Hosmer & Lemeshow, 1989).

## RESULTS

The results of descriptive statistics indicated that the median number of words read correctly on CBM was about 80 with a standard deviation of about 33. The average MCA score was about 1313 with a standard deviation of about 170. Given the proficiency level cut-score of 1420 on the MCA, this indicated that approximately 74% of the students did not reach the proficiency level.

The results of the regression analysis indicate that the use of the fall CBM measure appears to be a significant predictor of the MCA reading score in the spring of the year,  $F(1,1203) = 749.79$ ;  $p < 0.001$ ;  $r^2 = 0.39$ . This significant test result along with resulting in a large effect size (Cohen, 1988) indicated fairly strong validity evidence, suggesting a meaningful index of validity.

In addition, we observe that both the intercept ( $\beta_0 = 1064.02$ ) and slope parameters ( $\beta_1 = 3.22$ ) from the model were significant ( $p < 0.001$ ). The results indicate that for every single word increase in CBM scores in the fall, there is an expected increase in MCA reading scores of about three points. Therefore, from this result one would expect a score of 1420 on the MCA with a fall median CBM score of about 111 words read correct per minute.

The CBM score predictions of student classification with respect to the passage of the state-mandated test were also investigated. Passage on the test indicates proficiency levels in the area of reading and is also utilized for Adequate Yearly Progress (AYP) calculations for schools. Therefore, it is helpful to know how well it predicts passing on the mandatory, high-stakes test. This investigation was completed through two steps. The first step was to analyze the logistics regression of MCA status (where 1=pass and 0=not pass) on CBM scores. The second step was to analyze these results by a ROC procedure to identify how well CBM discriminated between those two groups.

The logistics regression results indicated that the CBM measure was a significant predictor of proficiency status on the MCA reading ( $\chi^2 = 285.833$ ;  $p < 0.001$ ) and accounted for about 30% (Nagelkerke's  $r^2 = 0.297$ ) of the maximal variance. Based on the results from the logistic regression, the diagnostic accuracy indices were tabulated (see Table 2).

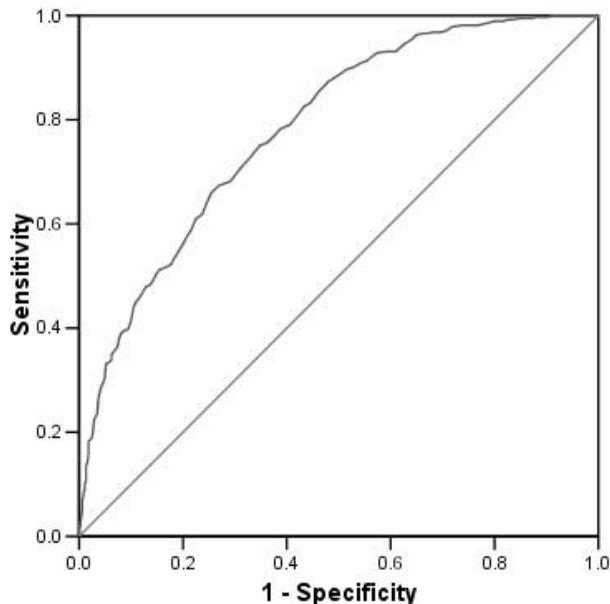
**TABLE 2.** *Classification Matrix Based on the Logistic Regression*

		Predicted		Marginal Percent Correct
		Pass	Not Pass	
Actual	Pass	166	211	44.0
	Not Pass	87	741	89.5
Marginal Percent Correct		65.6	77.8	75.3

By using the CBM scores, one can achieve an approximate correct classification percentage of about 75%. This indicates that about three out of four ELL students can be correctly classified on later proficiency status based on their CBM scores from the beginning of the school year. Therefore about one out of four will be incorrectly classified. Overall, the CBM appears to lend validity to inferences about later reading proficiency outcomes as measured by a state-mandated, high-stakes test.

To get a better understanding of the diagnostic accuracy associated with classification, the other measures were examined. The sensitivity of CBM to predict passing, given a student actually passes, was about 44%. The Specificity was about 90%. False positives were about 34% and false negatives were about 22%. The positive predictive power was about 66% and the negative predictive power was about 78%. It should be noted that any changes in the cut-point for classification can and will change these proportions. For our analysis we used the p-value of greater than or equal to 0.5 based on the logistic regression computations. From the regression analysis the predicted cut-point for classification was about 111 words read correct per minute. If this cut-point is changed in any manner, the associated diagnostic accuracy indices will also be changed. To identify the potential trade-offs in these indices the receiver operating characteristic (ROC) curve can be used (see Figure 1). This figure visually represents the associated trade-offs between sensitivity and 1-specificity related to different cut-points over the range of CBM scores.

**FIGURE 1.** *ROC Curve Indicating Tradeoff in Diagnostic Accuracy.*



Diagonal segments are produced by ties.

The second part of the analysis evaluated the discrimination ability of the CBM measure. For this a ROC analysis was completed. This method allows us to evaluate the effectiveness of the model to discriminate between the two different groups (those that passed and those that did not) based on the CBM scores, in general, and then within the different language groups (see Table 3). For this analysis the p-value computed from the logistic regression was utilized. The area under the curve (AUC) provides a useful metric about how well the model using the CBM score discriminates between students who later pass or do not pass the MCA reading proficiency. The closer the value is to unity, the better job it does at discriminating.

The ROC analysis indicated significant results and acceptable discrimination for the overall group and for each group individually. These results suggest that in about 78% of all possible pairs of cases (AUC = 0.78) in which one student passed and another student failed, using only the CBM score from the Fall semester, the logistic model assigns a higher probability of passing to those who actually passed. This suggests that the CBM scores are a valid indicator of later reading passing status on a state mandated proficiency tests.

Next, the analysis was run on the different language groups. This will allow for the analysis of whether the CBM discrimination index is different between the groups. The results of this analysis indicate that the different language groups have overlapping confidence intervals with respect to the area under the curve (see Table 3). This implies that the model works just as well for each group and that the CBM is equally discriminating within all the language groups.

**TABLE 3.** *Results of the ROC Analyses*

Variables	Area under the curve	Standard Error	Asymptotic Significance	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
All Groups	0.784	0.014	p < 0.001	0.758	0.811
Spanish	0.796	0.023	p < 0.001	0.751	0.840
Hmong	0.779	0.021	p < 0.001	0.737	0.821
Somali	0.778	0.047	p < 0.001	0.686	0.870

## DISCUSSION

A basic tenant of data-based procedures is the necessity of using well-grounded and psychometrically sound measures to accumulate information on student standing and progress. While CBM has a long history as a valid and reliable measure of reading (Deno, 2003), the provisions of NCLB have brought renewed focus to the importance of establishing the generalization of CBM measures to students of all backgrounds. This study provides additional information regarding the validity of using CBM oral reading fluency measures in English with English Language Learners. These measures, when used at the beginning of the fifth grade school year, provided a good predictor of later reading status on a state-mandated proficiency level test.

These results are particularly relevant because of the rapid growth in the ELL student population. ELL students can pose unique challenges for school staff, as they are challenged to learn about varied cultures, language backgrounds, levels of vocabulary development, etc. In addition, little information may be available on each student's specific background and development in reading. Efficient and reliable methods of early assessment are necessary for teachers and support staff to be able to direct interventions toward those students at-risk of failing to meet state proficiency standards. This study suggests that the long-standing findings related to the validity of CBM measures are also applicable to ELL students. The findings of this study also support the notion that these results are applicable across three

very divergent language groups (Spanish, Hmong, and Somali). This is particularly interesting because the limited previous research data supporting the use of CBM with ELL students has primarily focused upon Spanish-speaking students, but no study, as we aware, has been conducted for Hmong and Somali students.

The establishment of CBM as a valid tool for the purposes of screening and progress monitoring with ELL students can provide a practical framework for implementation of a Problem-Solving Model or RTI type approach to intervention and decision making. This approach seems particularly well suited to ELL students, for whom the use of norm-referenced assessment measures for special education eligibility decisions has long been in dispute.

However, despite this strong overall predictive ability, CBM provided better classification information for students who did not pass the proficiency level test, while providing weak classification information about the later status of students who actually did pass the proficiency level. These results suggest that the CBM has a high level of specificity and, thus, is a good indicator of later status as failing to meet the proficiency level in reading on the state-mandated, proficiency test. While the ability to have correctly predicted the classification of those students who do not pass proficiency exams seems more important than correct prediction of those who eventually pass the test, it may be that the addition of additional variables into the process would enhance overall sensitivity.

### **LIMITATIONS AND IMPLICATIONS FOR FUTURE RESEARCH**

While this study attempted to limit the variability of the sample somewhat by restricting the language groups studied to three languages, the background of the sample remained very diverse. The ELL sample in this study varied on many important factors, including the number of years they have lived in the U.S., their level of acculturation, formal and informal educational background, and type of English language support services received. Such variability may actually increase the generalizability of these findings, but point out the need for caution in applying these findings to any individual student. Another limitation of this study was that the CBM data used were collected by school staff members as part of their annual fall screening activities, and not as a part of a designed research study. Although the district has implemented a standardized data collection procedure with published reading probes and an instruction manual based on best practices, there is no specific procedure in place to monitor the fidelity of data collection.

In terms of implications for future research, it would be helpful to compare the cut score points for the ELL sample to that of their English speaking peers and to further break down the sample by ethnicity. We are currently working on a larger study of differential prediction which will look at issues of both race and language history. This study can be extended downward to determine if the same methodology works for younger students who are farther from the required high-stakes assessments. It is also possible that this type of research will provide a method for administrators to make differential decisions based on costs and benefits related to different cut-off levels for classification. This study used a basic probability outcome level of 0.5 as the cut-off point. However, the utilization of different cut-off points and their effects on classification can be assessed by the ROC analysis. As noted by Hintze and Silbergliitt (2005) it may be helpful to set separate cut scores for different decision purposes (e.g., screening, classification, and entitlement). The ROC graph provides a flexible and visual method to assess the relationship between the changes in sensitivity and the false negative error rate ( $1 - \text{specificity}$ ).

## REFERENCES

- Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometric issues. *Educational Assessment*, 8(3), 231-257.
- Abedi, J., Lord, C., & Hofstetter, C. (1998). *Impact of selected background variables on students' NAEP math performance* (CSE Technical Report #478). Los Angeles: University of California, National Center for Research on Evaluation, Standards and Student Testing.
- August, D., & Hakuta, K. (Eds.). (1997). *Improving schooling for language-minority students: A research agenda*. [Electronic version]. Committee on Developing a Research Agenda on the Education of Limited-English-Proficient and Bilingual Students. National Research Council (U.S.). Washington, DC: National Academy Press.
- Baker, S.K., & Good, R.H. (1995). Curriculum-based measurement of English reading with bilingual Hispanic students: A validation study with second-grade students. *School Psychology Review*, 24, 561-578.
- Carlo, M.S., August, D., McLaughlin, B., Snow, C.E., & Dressler, C. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly*, 30(2), 188-215.
- Cocking, R., & Chipman, S. (1988). Conceptual issues related to mathematics achievement of language minority children. In R. Cocking & J. Mestre (Eds.), *Linguistic and cultural influences on learning mathematics* (pp. 17-46). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2/e. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Deno, S.L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education*, 37(3), 184-192.
- Deno, S., Fuchs, L., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review*, 30(4), 507-524.
- Fien, H., Baker, S.K., Smolkowski, K., Mercier Smith, J.L., Kame'enui, E.J., & Beck, C.T. (2008). Using nonsense word fluency to predict reading proficiency in kindergarten through second grade for English learners and native English speakers. *School Psychology Review*, 37(3), 391-408.
- Fuchs, L., & Deno, S. (1981). *The relationship between curriculum-based mastery measures and standardized achievement tests in reading* (Research Report No. 57). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities. (ERIC Document Reproduction ED212662).
- Fuchs, L., & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: A review of three forms of classroom-based assessment. *School Psychology Review*, 28(4), 659-671.
- Good, R.H., Simmons, D.C., & Kame'enui, E.J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading*, 5(3), 257-288.
- Hamilton, C.R., & Shinn, M.R. (2003). Characteristics of word callers: An investigation of the accuracy of teachers' judgements of reading comprehension and oral reading skills. *School Psychology Review*, 32(2), 228-240.
- Hanley, J., & McNeil, B. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29-36.
- Hintze, J.M., Callahan, J.E., Matthews, W.J., Williams, S.A.S., & Tobin, K.G. (2002). Oral Reading Fluency and Prediction of Reading Comprehension in African American and Caucasian Elementary School Children. *School Psychology Review*, 31(4), 540-553.
- Hintze, J.M., & Silbergliitt, B. (2005). A Longitudinal Examination of the Diagnostic Accuracy and Predictive Validity of R-CBM and High-Stakes Testing. *School Psychology Review*, 34(3), 372-386.
- Hopstock, P.J., & Stephenson, T.G. (2003). Special Topic Report #2: Analysis of Office for Civil Rights (OCR) Data Related to LEP Students. National Center on Education Outcomes, University of Minnesota.
- Hosmer, D., & Lemeshow, S. (1989). *Applied logistic regression*. New York: John Wiley & Sons.
- Kindler, A. (2002). *Survey of the states' Limited English Proficient students & available educational programs and services 1999-2000 Summary Report*. Washington, DC: US Department of Education & National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs.
- Klein, J. & Jimerson, S. (2005). Examining ethnic, gender, language, and socioeconomic bias in oral reading fluency scores among Caucasian and Hispanic students. *School Psychology Quarterly*, 20 (1), 23-50.
- Kranzler, J., Miller, M., & Jordan, L. (1999). An examination of racial/ethnic and gender bias on curriculum-based measurement of reading. *School Psychology Quarterly*, 14(3), 327-342.
- Liu, K., Anderson, M., & Thurlow, M. (2000). *1999 Report on the participation and performance of limited English proficient students on Minnesota's Basic Standards Tests* (Minnesota Report No. 30). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved June 4, 2003, from <http://education.umn.edu/NCEO/OnlinePubs/MnReport30.html>
- Marston, D. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. Shinn (Ed.), *Curriculum-Based Measurement: Assessing Special Children* (pp. 18-78). New York: Guilford Press.

- Marston, D., & Magnusson, D. (1985). Implementation of curriculum-based measurement in special and regular education settings. *Exceptional Children*, 52(3), 266-276.
- McGlinchey, M.T., & Hixson, M.D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review*, 33(2), 193-203.
- Minneapolis Public Schools (2003). *Performance Assessment of Reading in the Problem Solving Model*. Minneapolis, MN: Author.
- Minnesota Department of Education (n.d.). *Minnesota Comprehensive Assessments in Grades 3, 5 and 7 Reading and Mathematics and Grade 5 Writing. Technical Manual 2004 Administration*. Retrieved March 16, 2006, from <http://children.state.mn.us/mde/static/006764.pdf>.
- Moss, M., & Puma, M. (1995). *Prospects: The congressionally mandated study of educational opportunity and growth: Language minority and limited English proficient students*. Washington, DC: US Department of Education.
- National Research Council. (1997). *Improving Schooling for Language-Minority Children: A Research Agenda*. August, D. & Hakuta, K. (Eds.). Washington, D.C. National Academy Press.
- Neter, J., Kutner, M., Nachtsheim, C., & Wasserman, W. (1996). *Applied linear statistical models* (4<sup>th</sup> ed.). New York: McGraw-Hill.
- No Child Left Behind Act of 2001*. Pub. L. No. 107-110.
- Pearce, L.R., & Gayle, R. (2009). Oral reading fluency as a predictor of reading comprehension with American Indian and White elementary students. *School Psychology Review*, 38(3), 419-427.
- Shinn, M., Good, R., Knutson, N., Tilly, W., & Collins, V. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review*, 21(3), 459-479.
- Shinn, M. R. (Ed.). (1989). *Curriculum-based measurement: Assessing special children*. New York: Guilford Press.
- Sibley, D., Biber, D., & Hesch, A. (2001, March). *Establishing curriculum-based measurement oral reading fluency performance standards to predict success on local and state tests of reading achievement*. Paper presented at the Annual Meeting of the National Association of School Psychologists. Washington, DC.
- 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(3), 407-419.
- Sudweeks, R.R., Glissmeyer, C.B., Morrison, T.G., Wilcox, B.R., & Tanner, M.W. (2004). Establishing reliable procedures for rating ELL students' reading comprehension using oral retellings, *Reading Research and Instruction*, 43(2), 65-86.
- Swets, J., Dawes, R., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, 1(1), 1-26.
- Thomas, W., & Collier, V. (1997). *School Effectiveness for Language Minority Students*. Washington, DC: National Clearinghouse for Bilingual Education.
- US Bureau of the Census. (2000). *Census 2000*. Washington, DC: US Government Printing Office.
- Vanderwood, M. L., Linklater, D., & Healy, K. (2008). Predictive accuracy of nonsense word fluency for English language learners. *School Psychology Review*, 37(1), 5-17.