Using CBM in Reading and Math to Predict Performance on NYS ELA and Math Assessments

Emily DuFore, BA
Carly Honors, BA

State University of New York
College at Oswego

Introduction
With the passage of the No Child left Behind Act (NCLB, 2001) and its’ focus on
Using CBM in reading and math to predict performance on NYS ELA and Math Assessments

educational accountability, all states are required to implement state-wide assessments to demonstrate student academic growth. This criterion was set to create accountability in regards to student success, with a large emphasis on assessing the development of reading and math skills of students in grades three through eight (e.g., Burns, Madyun, & Lail, 2006). States are required to implement assessments to measure student performance and administer these assessments to nearly all students. The implementation of mandated test administration has lead New York State (NYS) to implement two assessments, including the NYS English Language Arts (ELA) assessment and Math tests, both of which are used to evaluate a school’s ability to prepare students for academic success. Results obtained from these test administrations are used as indicators of student proficiency and overall school performance. In line with NCLB low performing schools are at risk for closing and NYS has followed the federal trend of holding educators and schools accountable for the success of the attending students (Siberglitt, Burns, Madyun, & Lail, 2006).

A parallel movement of increased academic accountability from the realm of Special Education is the recent adoption the Response to Intervention (RTI) model for assessing and treating students displaying learning delays. RTI is a preventative model that assesses every child’s basic academic performance at least 3 times per year and includes procedures for intervening with students at risk for academic failure. Utilizing the RTI model, ‘at-risk’ students are assessed and identified in the fall of the year and prior to administration of the state-wide assessment procedures, creating prompt identification measures and providing immediate support services to children.

**Curriculum Based Measurement and State Assessments.** The RTI framework provides a model that ensures success for all students through the implementation of a tiered system of supports and intervention. This model relies heavily on Curriculum Based Measurement (CBM) tools which have demonstrated” technical adequacy”- (validity and reliability) for academic screening and progress monitoring (Deno, 2003). CBM techniques have brief administration times (generally 1 to 5 minutes) and are sensitive to change over time (5 to 8 weeks). The sensitivity, brevity, and validity of these measures allows schools to utilize CBM’s to obtain reliable measurements when screening and progress monitoring students’ skills. CBM provides multiple measurements of students’ academic skills and compares these results to the expected
skills obtained from the general education curriculum that is taught (Hintze & Silberglitt, 2005), providing an accurate measure of students’ skill obtainment.

The utility of universal screening provided by CBM’s has lead to the administration of CBM measures within many schools, whereby data is obtained on every student. Given that NYS testing results are the annual criterion test for almost all students in the State, educators should be interested in whether universal CBM assessments can be used as a predictive measure of student performance on the NYS tests. In reading, Oral Reading Fluency (ORF) has been shown to be an accurate predictor of performance on the Minnesota Comprehensive Assessments in Reading (Silberglitt et al, 2006) and the Washington Assessment of Student Learning. Many researchers (Silberglitt et al, 2006, McGlinchey, & Hixon, 2004, Stage, & Jacobsen, 2001) have demonstrated the reliability of a timed recording of correct read words as empirically valid, in addition to providing evidence of an existing relationship between ORF progress monitoring measures and state test performance.

Current development of CBM in math utilizes brief assessments that monitor student’s skills through computations probes and concepts and application probes (Fuchs, Fuchs, Compton, Bryant, Hamlett, & Seethaler, 2007). Probes generally cover mastery skills that are expected to be taught throughout the curricula (Shapiro, Keller, Lutz, Santoro, & Hintze, 2006) and focus on students’ acquisition of fluency, accuracy, and mathematical conceptual understanding. Given that assessing student skills in math concepts requires some reading the relationships between CBM-math and state assessment performance are generally more modest than those observed between ORF and reading assessments (e.g., Thurber, Shinn, & Smolkowski, 2002). Relationships between CBM-Math and standardized or annual state assessments are in the range between .26 and .67 (Christ, Scullin, Tolbize, & Jiban, 2008). Given that many districts in NYS are beginning to adopt universal screening procedures in the area of mathematics it would also be advantageous for them to identify the predictive validity of CBM math probes in relation to the annual NYS math assessments.

As districts across adopt CBM techniques for universal student assessment the use of these scores to predict student performance on high stakes state assessments should become increasingly popular. If CBM assessments in reading and math were found to be adequately predictive of NYS ELA and Math performance then predictive cut scores could be established.
In the fall predictive cut-scores could be used to identify those students at risk for failing the NYS assessment later that spring. Additionally, a winter cut score could identify a performance level of math and reading the predicted student success on the NYS test. This winter benchmark could serve as the performance goal such that educators implementing academic interventions could progress monitoring student progress toward a skill level that predicts success on the NYS tests. CBM measures can provide cut scores to predict performance on high-stakes testing (Hintze, & Silberglitt, 2005). With this cut score method of identifying and intervening with students early, schools are moving away from the ‘wait to fail’ model (VanDerHeyden, 2010), and towards RTI best practices.

Current Study. A primary purpose of this study is to investigate the relationship between CBM in reading at fall and winter and the NYS ELA (English Language Arts) in spring, in addition to investigating the relationship between CBM in math at fall and winter and the NYS Math assessment in spring. In addition, this study was conducted to identify predictive cut scores in CBM reading and math that can be used to predict future success on NYS assessments. This study investigated the correlation between CBM and NYS assessments in grade 3 and 5 as well as the a) sensitivity (number of students who were predicted to fail who actually failed), b) specificity (number of students who were predicted to pass who actually passed), and c) efficiency (percent of correctly identified students based on the specific data). Further the number of students targeted for intervention were considered in addition to the sensitivity, specificity, and efficiency of CBM cut scores used to predict performance on NYS assessments.

Method

Participants

Anonymously coded, archival data was obtained for the purpose of this study. Schools throughout New York State were contacted for voluntary participation. Schools were offered a comprehensive summary of results obtained, as well as normative scores that may be used for identification of at-risk students. School locations ranged from the southern tier to northern regions of New York. Archival data was obtained for students in second through eighth grade though only grades 3 and 5 were included in the current analysis. Benchmark scores for fall and
winter were requested for measures of oral reading fluency and mathematic skills as well as New York State assessment scores. Student data that did not have corresponding New York State assessment scores were eliminated from the analysis. The total number of participants (n) for reading analysis in the 3rd grade during the fall and winter was 356. In the 5th grade, n=326. For math analysis, at the 3rd grade level fall and winter computations n= 140 and 143 respectively. For concepts and applications n= 140 and 144 respectively. At the 5th grade level for fall and winter computation n= 162 and 166 respectively. For concepts and applications n= 160 and 166 respectively.

Measures

Reading

Student scores on the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency (ORF) routinely administered by school staff were collected as benchmark measures of student reading performance in the fall and winter. The DIBELS measure was collected as it is widely used among schools districts and has research-based validity as a measure of reading fluency.

Math

Graduate students at SUNY Oswego created curriculum-based measures in math. The math probes were designed to reflect criteria in proportion to the New York State end of year math exam. Two types of math probes were utilized to obtain a general outcome measure in math. A computations probe and a concepts and applications probe were used to collected benchmark measures of students mathematical abilities in the fall and winter. CBM-math probes materials are available on-line at http://www.schooltoolz.us/

Cut Score Program

Dr. Michael Leblanc, at SUNY Oswego, created a cut score program utilizing Microsoft Excel spread sheets. Dr. Leblanc created a template in Excel that utilizes ordinary least squares regression to calculate initial starting points for cut off scores based on benchmark data and end of year assessment scores. The program reports descriptive statistics on local data that can
provide practitioners with a greater understanding of student performance at a systems-level. An 
r score was produced, which is a measure of correlation between the two scores of concern 
(CBM data and NYS assessment score). The r score identified the strength of the relationship 
between the CBM utilized and performance on a NYS assessment. The program allowed 
researchers to alter the suggested cut score to achieve optimal values of efficiency, sensitivity, 
specificity, and to restrict the range of students targeted for intervention. The cut score program 
developed for this project is also to be available at http://www.schooltoolz.us/

Design

A quasi-experimental design was utilized for the purpose of this analysis. The analysis 
was completed in three phases. Phase one consisted of data collection and organization. Phase 
two consisted of data input, using the cut score program, and phase three consisted of a review of 
data output.

Phase One: Data Collection and Organization

Participating schools sent anonymously coded data via e-mail. Researchers utilized 
Microsoft Excel spreadsheets to organize collected data. For measures of reading, data was 
organized by grade and time of assessment (fall, winter, or spring). For measure of math, data 
was organized by grade, type of probe (computation and concepts and application), and 
assessment time (fall or winter). Using Microsoft excel data was arranged into corresponding 
columns of student performance across all measures (i.e. benchmark data and NYS assessment 
scores). Researchers calculated mean, standard deviation, and percentile ranking (10th, 25th, 75th, 
and 90th percentile) of the data. The sorting tool was utilized to identify any missing data. For 
example, a missing NYS assessment score was identified to eliminate corresponding benchmark 
data from the analysis, as calculations were dependent on a perceived relationship between the 
two measures. Columns consisting of individual benchmark measures with corresponding NYS 
assessment scores were used to organize the remaining data. For example, a fall ORF score was 
paired with the students corresponding NYS ELA score.

Phase Two: Data Input
The cut score program utilized, allowed data inputs consisting of one CBM measure (DIBELS or CBM-Math) with its corresponding high stakes assessment score (NYS ELA or NYS Math) for one assessment time period (i.e. fall ORF probes). Previously organized data was entered into the program to determine a suggested cut score. NYS performance scores for 2010-2011 were used for the purpose of this analysis.

**Phase Three: Review of Data Output**

The cut score program outputs a suggested cut score tailored to the specific data entered. The reported cut scores were then in some cases altered for various reasons including limited resources for intervening with a high projected number of students identified as at risk or to establish an increase in the cut score across time (form fall to winter). Researchers looked at four descriptive statistics to alter the suggested cut score: efficiency, sensitivity, specificity and number of students targeted for intervention were considered. Ideally, researchers were looking to obtain a cut score the yielded a high level of sensitivity while keeping the efficiency high (attempted to reach 80% efficiency). The researchers also altered data to keep the number of students targeted for intervention within a reasonable range.

After the three phases were completed, data was arranged into grade level tables including correlation, mean, standard deviation, percentile rankings, sensitivity, specificity, efficiency, and the altered cut score. The tables were intended to provide an easily interpreted source of information for districts to utilize.

**Results**

The results of this analysis provide a greater understanding of the predictive relationships between CBM measures in reading and math and NYS final exams. The relationship between the two identified measures provides educators with the ability to identify students at risk of failing a high stakes test early in the year in order to provide necessary interventions to identified students. By utilizing the cut score program, suggested benchmarks were obtained. A review of the data output reveled a need for alterations in both math and reading measures, as the suggested scores did not always follow a logical progression of skill development over time.
Table 1, represents a collection of data in reading for grades 3 and 5. Table 2, represents a collection of data in math for grades 3 and 5. There are five statistical measures reported within the tables, they include the cut score, sensitivity, specificity, efficiency, and correlation. The reported cut score is the minimum suggested score that students would need to achieve in order to be successful on the corresponding NYS exam. If a student scores below the suggested cut score, they are at risk of failing the NYS exam at the end of the year. Sensitivity refers to the suggested cut scores ability to correctly identify students who will fail the NYS exam. For example, in Table 1 83% of the students predicted to fail the NYS exam based on a suggested ORF cut score of 77 actually failed at the end of the year. Specificity refers to the suggested cut scores ability to correctly identify students who will pass the NYS exam. Efficiency refers to the cut scores ability to correctly identify students overall. For example, in Table 1 a cut score of 71 during fall in the 3rd grade correctly identified 80% of the students.

Reading

An investigation of the relationship between CBM in reading at fall and winter produced the following results. At the third grade level, a moderately strong positive relationship was found in the fall (r=.63), and a similar relationship was found in the winter (r=.61). At the fifth grade level, a moderate positive relationship was found in the fall (r=.53) and winter (r=.50).

Researchers reviewed the data and identified the following cut scores for use in identifying students at risk of failing the NYS ELA exam at the fall and winter. At the third grade level, the suggested cut score for fall and winter are 71 and 79 respectively. At the fifth grade level, the suggested cut score for fall is 81 and for winter is 87. The scores represent a cut off point in which any student who falls below these scores is at risk of failing the NYS ELA exam.

Researchers reviewed levels of specificity, sensitivity, and efficiency when developing the above scores. At the third grade, the following levels were obtained for the suggested cut scores in the fall, .79(specificity), .83(sensitivity), and 80%(efficiency). In the winter, .70(specificity), .88(sensitivity), and 75%(efficiency). At the fifth grade, the following levels were obtained for
the suggested cut scores in the fall, .81(specificity), .91(sensitivity), and 82%(efficiency). In the winter, .73(specificity), .81(sensitivity), and 75%(efficiency)

**Math**

An investigation of the relationship between CBM in math at fall and winter produced the following results. At the third grade level, a weak positive relationship was found in the fall for computations probes \( (r=.35) \), as well as for the concepts and applications probes \( (r=.42) \). In the winter, a similar relationship was found for computations \( (r=.42) \) and concepts and applications \( (r=.42) \). At the fifth grade level, a moderate positive relationship was found in the fall for computations probes \( (r=.55) \). A moderately strong positive relationship was found in the fall for concepts and application probes \( (r=.65) \). In the winter, similar results were found for computations \( (r=.56) \) and concepts and applications \( (r=.66) \).

Researchers reviewed the data and identified the following cut scores for use in identifying students at risk of failing the NYS Math exam at the fall and winter. At the third grade, the suggested cut score for fall and winter computations are 9 and 19 respectively. For concepts and applications the suggested cut scores are 10 and 13 respectively. At the fifth grade level, the suggested cut score for fall and winter computations are 25 and 36 respectively. For concepts and applications the suggested cut scores are 10 and 12 respectively. The scores represent a cut off point in which any student who falls below these scores is at risk of failing the NYS ELA exam.

Researchers reviewed levels of specificity, sensitivity, and efficiency when developing the above scores. At the third grade, the following levels were obtained for the suggested cut scores for fall computations, .68(specificity), .63(sensitivity), and 67%(efficiency). In the winter, .68(specificity), .64(sensitivity), and 67%(efficiency). Fall concepts and applications are as follows, .55(specificity), .70(sensitivity), and 59%(efficiency). Winter concepts and applications are as follows, .57(specificity), .71(sensitivity), and 62%(efficiency). At the fifth grade, the following levels were obtained for the suggested cut scores for fall computations, .70(specificity), .69(sensitivity), and 70%(efficiency). In the winter, .73(specificity), .81(sensitivity), and 75%(efficiency). Fall concepts and applications are as follows,
Using CBM in reading and math to predict performance on NYS ELA and Math Assessments

.70(specificity), .74(sensitivity), and 72%(efficiency). Winter concepts and applications are as follows, .75(specificity), .72(sensitivity), and 73%(efficiency).

Table 1

<table>
<thead>
<tr>
<th></th>
<th>Grade 3</th>
<th></th>
<th>Grade 5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fall ORF</td>
<td>Winter ORF</td>
<td>Fall ORF</td>
<td>Winter ORF</td>
</tr>
<tr>
<td>Cut Score</td>
<td>71(n=356)</td>
<td>79(n=356)</td>
<td>81(n=326)</td>
<td>87(n=326)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>.83</td>
<td>.88</td>
<td>.91</td>
<td>.81</td>
</tr>
<tr>
<td>Specificity</td>
<td>.79</td>
<td>.70</td>
<td>.81</td>
<td>.73</td>
</tr>
<tr>
<td>Efficiency</td>
<td>80%</td>
<td>75%</td>
<td>82%</td>
<td>75%</td>
</tr>
<tr>
<td>Correlation (r)</td>
<td>.63</td>
<td>.61</td>
<td>.53</td>
<td>.50</td>
</tr>
</tbody>
</table>

Note: ORF = Oral Reading Fluency; NYS ELA = New York State English-Language Arts Exam.

Table 2

<table>
<thead>
<tr>
<th>Grade 3</th>
<th>Fall Computations</th>
<th>Fall Concepts and Applications</th>
<th>Winter Computations</th>
<th>Winter Concepts and Applications</th>
</tr>
</thead>
</table>
Using CBM in reading and math to predict performance on NYS ELA and Math Assessments

Using CBM in reading and math to predict performance on NYS ELA and Math Assessments

<table>
<thead>
<tr>
<th>Cut Score</th>
<th>9 (n=140)</th>
<th>10 (n=140)</th>
<th>19 (n=143)</th>
<th>13 (n=144)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>.63</td>
<td>.70</td>
<td>.64</td>
<td>.71</td>
</tr>
<tr>
<td>Specificity</td>
<td>.68</td>
<td>.55</td>
<td>.68</td>
<td>.57</td>
</tr>
<tr>
<td>Efficiency</td>
<td>67%</td>
<td>59%</td>
<td>67%</td>
<td>62%</td>
</tr>
<tr>
<td>Correlation (r)</td>
<td>.35</td>
<td>.42</td>
<td>.42</td>
<td>.42</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade 5</th>
<th>Fall Computations</th>
<th>Fall Concepts and Applications</th>
<th>Winter Computations</th>
<th>Winter Concepts and Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cut Score</td>
<td>25 (n=162)</td>
<td>10 (n=160)</td>
<td>36 (n= 166)</td>
<td>12 (n=166)</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>.69</td>
<td>.74</td>
<td>.67</td>
<td>.72</td>
</tr>
<tr>
<td>Specificity</td>
<td>.70</td>
<td>.70</td>
<td>.69</td>
<td>.75</td>
</tr>
<tr>
<td>Efficiency</td>
<td>70%</td>
<td>72%</td>
<td>68%</td>
<td>73%</td>
</tr>
<tr>
<td>Correlation (r)</td>
<td>.55</td>
<td>.65</td>
<td>.56</td>
<td>.66</td>
</tr>
</tbody>
</table>

Note: CBM= Curriculum-Based Measurement; NYS Math= New York State Math Exam.

Discussion

The purpose of this study was to a) investigate the relationship between curriculum-based measurements (CBM) in reading and math and New York State end-of-year assessments, b) identify predictive cut-scores for CBM in reading and math that could be used to predict success on NYS assessments, and c) to identify specificity, sensitivity, and efficiency of identified cut scores. Archival benchmark assessment and NYS testing data was collected from local schools to demonstrate the utility of these scores for identification of students at-risk of failing an end-of-
year NYS assessment. This approach is inline with the response to intervention model as it allows educators to provide interventions early in the year and does not wait for student to fail at the end of the year.

In reading a moderate to strong correlations were found in both the fall and spring across grades 3 and 5. These results suggest that educators could utilize DIBELS ORF probes as a reliable measure for predicting performance on the NYS ELA exam. Established cut scores were able to correctly identify students in a majority of the cases (75%-82%). These findings lend support to the utility of CBM’s in identifying students at risk of failing high stakes end of year assessments found in previous studies. Higher levels of correlation were found in grade 3 when compared to grade 5. These findings are again inline with previous studies that have suggested the oral reading fluency is a stronger performance measure in grade 3 as opposed to grade 5 (e.g., Keller-Margulis, Shapiro, & Hintze, 2008).

In math lower levels of correlation were observed, especially in the 3rd grade (.35-.42). This suggests that educators would demonstrate caution utilizing the CBM measure used (computations and concepts and applications probes) to predict success on the NYS Math exam at the 3rd grade level. There was a slightly stronger correlation for the 5th grade (.55-.65). The difference may be attributed to several factors. One possibility is the math probes may be more reliable at the higher levels. A second factor may be related to the need for more accurate administration and scoring of the math materials.

Limitations. The current study has a number of limitations that could limit the generalizability of the obtained results. First, all the CBM assessments were conducted by school staff. The accuracy of these scores was not able to be ensured. Second, the sample of student scores was admittedly one of convenience rather than a broad based sample more representative of the general population. Third, the CBM-math measures were developed by students in the School Psychology program at SUNY Oswego in collaboration with classroom teachers from NYS. No psychometric properties of these materials were obtained prior to the current study. This may have limited the results obtained in math.

Summary. Overall, the findings suggest a range of strengths in the relationship between the CBMs used in schools and the corresponding NYS exam. A majority of the findings indicate a moderate to strong relationship, suggesting that CBM scores are a good predictor of success on
a NYS exam especially with respect to reading and ELA. The results also lend support to the utility of the CBM’s identified as reliable measures for early identification of students in academic jeopardy. The format of the cut score program used would allow schools to tailor their academic interventions based on a variety of factors including sensitivity, specificity, and efficiency. Another benefit of the findings is support for using local data to calculate cut scores. Cut scores derived from local data provide a means of comparison that is based on the needs and environmental conditions of individual districts. The establishment and use of these predictive cut scores to intervene with struggling students can be yet another tool that schools and school psychologists use to efficiently respond to increased educational accountability.

References


Using CBM in reading and math to predict performance on NYS ELA and Math Assessments

Psychoeducational Assessment, 24(1), 19-35.