ABSTRACT

VALUE-ADDED AND CURRICULUM-BASED MEASUREMENT TO EVALUATE STUDENT GROWTH

by Aubrey M. Micheli

The following study examines the relationship between state achievement assessments and curriculum-based measurements. Specifically, reading performance on the Ohio Achievement Test and curriculum-based measurement oral reading fluency scores were evaluated. Relationships were examined to determine possible predictors of future state achievement test performance. Trimester benchmarks in oral reading fluency were utilized in evaluating the reading performance of 575 students from a suburban Southwestern Ohio school district. The results indicated that curriculum-based measurements of oral reading fluency and Ohio Achievement Test reading have a strong predictive relationship. A model, utilizing the most recently administered Ohio Achievement Test and oral reading fluency scores demonstrated the strongest predictive relationship with future successful state test performance. Given the federal and state objectives focused on increasing student achievement, these findings provide schools and educators with practical methods of monitoring and predicting student success.
VALUE-ADDED AND CURRICULUM-BASED MEASUREMENT TO EVALUATE STUDENT GROWTH

A Thesis

Submitted to the
Faculty of Miami University
in partial fulfillment of
the requirements for the degree of
Education Specialist
Department of Educational Psychology
by
Aubrey M Micheli
Miami University
Oxford, Ohio
2010

Advisor_______________________________________
Dr. Michael F. Woodin

Advisor_______________________________________
Dr. Jason T. Abbitt

Reader_______________________________________
Dr. Tonya S. Watson

Reader_______________________________________
Dr. Jane E. Cole
Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literature Review</td>
<td>1</td>
</tr>
<tr>
<td>Method</td>
<td>16</td>
</tr>
<tr>
<td>Design and Procedures</td>
<td>17</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>18</td>
</tr>
<tr>
<td>Results</td>
<td>20</td>
</tr>
<tr>
<td>Discussion</td>
<td>26</td>
</tr>
<tr>
<td>Limitations</td>
<td>28</td>
</tr>
<tr>
<td>References</td>
<td>30</td>
</tr>
</tbody>
</table>
List of Tables

Table 1: Dependent and Independent Variables 18

Table 2: Correlations among Oral Reading Fluency and Ohio Achievement Test Scores 20

Table 3: Correlation Coefficients between Spring 2008 OAT and Spring 2009 Performance 22

Table 4: Correlation Coefficients between Winter 2009 and Spring 2009 OAT Performance 23

Table 5: Comparison of Regression Models examined Utilizing Backwards Selection 25

Table 6: Part and Partial Correlations of the Predictors with Spring 2009 OAT Performance 26
List of Figures

Figure 1  

Figure 2  

iv
The No Child Left Behind Act (NCLB) was established with the intent of providing all children with an equal and quality education. Federal law mandates that districts be held accountable for educational practices and student outcomes; critical consequences are allocated to districts that fail to meet established standards. States have developed their own compilation of Local Report Cards (LRC) to assess the academic performance and progress of students to demonstrate instructional effectiveness. (McDonald, 2002). The Ohio Department of Education annually issues Local Report Cards for every district and school within the state. A new component of the LRC, value-added (VA), was recently included to better account for student progress within an academic school year through evaluating annual student growth by means of performance on Ohio Achievement Tests (OAT). However, since the initiation of this study and completion data collection, the OAT has been renamed the Ohio Achievement Assessment. Thus, it will be referred to as the Ohio Achievement Test (OAT) throughout this paper.

This high-stakes environment has led schools to utilize tools, such as curriculum-based assessment (CBA), which was designed to evaluate students’ performance level on skills taught within the curriculum. Curriculum-based measurement (CBM), one type of CBA, utilizes standardized procedures and is administered on a trimester schedule throughout the school year. Therefore, student performance and yearly growth can be established in the areas of reading, math, written expression (Hosp, Hosp, & Howell, 2007), and spelling (Howell & Nolet, 2000). However, it should be noted that curriculum-based measurements are not actually based on any curriculum, but on instructionally relevant skills and material. These assessments are used for early identification and intervention for struggle students. This framework of early remediation results in improved academic outcomes for those students when CBM’s are used in conjunction with decision rules (Fuchs & Deno, 1991; Howell & Nolet, 2000). In addition, CBM oral reading fluency scores have been demonstrated to have a strong predictive relationship with state achievement tests (Crawford, Tindal, and Stieber,
This relationship allows school to efficiently and effectively monitor students’ progress while also predicting which students are at the highest risk for OAT failure (Schulte, Villwock, Whichard, and Stallings, 2001).

The objective of this study is to examine the utility of value-added scores in reflecting student progress; more specifically progress in reading will be evaluated. This will be done by examining the relationship between CBM scores and OAT scores. Currently, the VA model predicts student’s future test performance by evaluating past OAT scores. CBM benchmarks are collected on a trimester schedule throughout the year to document gains in achievement. Thus, CBM benchmarks will be compared to OAT scores to examine possible predictors of future OAT performance. If a predictive relationship does exist between one or more of these variables, perhaps CBM could serve as an alternative measure of student progress to the current model.

The following relationships will be examined:

Question 1: What is the relationship between Ohio Achievement Test scores and reading curriculum-based measurement scores from the previous year?

Question 2: What is the relationship between Ohio Achievement Test scores and reading curriculum-based measurement scores from the same year?

Question 3: To what degree are oral reading fluency scores able to predict Ohio Achievement Test scores for the following year?

Question 4: How does the relationship between Ohio Achievement Test scores from Fall 3rd grade to Fall 4th grade compare to oral reading fluency scores over the same time period?

Local Report Cards

Current Local Report Cards measure school performance in four core areas: State Indicators, Performance Index, Adequate Yearly Progress (AYP), and value-added growth reports. Based on performance on these four core areas, each school is awarded a
designation as an indication of the schools overall effectiveness. Each element is detailed below.

State Indicators

There are currently 25 different state indicators on which schools are evaluated. Two of these indicators reflect graduation and attendance rates. The remaining 23 reflect student performance on the Ohio Achievement Tests (OAT). Ohio Achievement Tests are administered to children in grades 3-8, while the Ohio Graduation Test (OGT) is given to students in grade 10 and grade 11 if necessary. The range of scores on each test comprises the following performance labels: advanced, accelerated, proficient, basic, and limited. Schools are then assessed on the percentage of total indicators met (ODE, 2005). Descriptions of each state indicator are provided below.

Proficiency test indicators – 3rd – 8th grade. In grades 3 through 8, there are 13 indicators that measure student performance. All students are evaluated in reading and math, while fourth grade students are additionally evaluated in the subject of writing. To acquire an indicator for each individual subject area, 75 percent of students must score at or above the proficient level. If less than 75 percent of students met the proficient level, the school does not acquire that particular state indicator (ODE, 2005).

Proficiency test indicators – 10th & 11th grade. In grades 10 and 11, there are ten performance indicators to be acquired. Students in both grades are assessed for achievement on the Ohio Graduation Test (OGT) in the subject areas of reading, writing, math, social studies, and science. In grade 10, consistent with the previously described standards, 75 percent of students must achieve proficiency for an indicator to be met. If less than 75 percent of students do not reach the proficient level, the state indicator has not been met. It is essential to note that the state minimum standard of students at the proficient level is increased to 85 percent on the 11th grade OGT (ODE, 2005).

Graduation rate state indicator. The graduation rate from the preceding school year determines if the state indicator is met. Only high schools containing grades 9 through 12 are evaluated on this indicator. A 90 percent graduation rate must be met to
acquire the state indicator. Thus, to determine the graduation rate for the 2007-2008 Local Report Card, the graduation rate for the 2006-2007 school year must be reviewed (ODE, 2005).

*Attendance rate state indicator.* The final state indicator evaluates the student attendance rate for an entire school, including all grade levels served. The aggregate attendance rate is calculated by comparing the number of days all students spent in attendance with the total number of students enrolled. An attendance rate of 93 percent must be met to obtain the state indicator (ODE, 2005).

The next significant feature of the Local Report Card is the Performance Index, which is subsequently described.

*Performance Index*

The Performance Index (PI) is designed to reflect the performance of each student assessed with the OAT and/or OGT. Weighted scores are awarded for different levels of student performance. Thus, any students that do not reach proficiency are taken into consideration and scored accordingly. Scores range from 0.0 to 120.0 and are assigned for each subject area in all tested grades. It is noteworthy that the OGT is only measured during the 10th grade administration (ODE, 2005).

The third key factor used to evaluate a school’s Local Report Card designation is value-added Reports, which are described below.

*Value-Added Reports*

Value-Added is a statistical measure designed to evaluate student progress from one year to the next. The intended use of VA is to evaluate instructional effectiveness and the impact schools have on students by examining yearly student growth. This is accomplished by comparing a student’s current level of learning with their level of learning in the past via OAT scores.
VA scores are compiled by gathering achievement test scores of individual students that are available within the preceding five years. Scores are subsequently used to longitudinally produce estimates of future student achievement trajectories. In order to determine the growth trajectory of a student, that student’s past achievement history is compared to the current achievement trajectory norms of similarly performing students. This trajectory is compared to the growth rate necessary to reach proficiency on the subsequent grade-level assessment, establishing a standards-based measure of adequate growth. Students with similar achievement histories will, consequently, yield similar growth trajectories, regardless of demographic attributes (ODE, 2006).

The value-added model is grounded on the assumption that prior achievement is an accurate predictor of future test achievement. Test scores that surpass projected values suggest exceedingly effective instruction. Conversely, scores that fall short of projected values indicate inadequate instruction (Hershberg, Simon, & Lea-Kruger, 2004).

Value-Added designations. Beginning with fourth grade and continuing through eighth grade, growth trajectories are developed for all students. Thus, scores on third grade OAT tests are used to establish fourth grade trajectories. Similarly, eighth grade trajectories are intended to predict performance on the OGT. Student proficiency is expected by the next grade level in a new school. For example, in a school that serves students in grades K through 5, a student’s growth trajectory would be directed at reaching proficiency by grade 6. If trajectories fail to demonstrate adequate proficiency gains, schools are pressured to improve academic progress. Schools are awarded points for academic gains in each subject and receive designations accordingly: “Green” indicates growth that exceeds expected state-level trajectories; “Yellow” indicates growth that coincides with expected state-level trajectories; and “Red” indicates growth that fails to meet growth trajectory expectations. Schools exceeding expected yearly growth may move up one LRC designation, while schools failing to meet expected growth for two consecutive years will move down one designation. Schools that meet expected growth rates are unaffected by value-added measures (ODE, 2006).
The final element of a Local Report Card designation is Adequate Yearly Progress, which is described below.

*Adequate Yearly Progress (AYP)*

The intent of Adequate Yearly Progress (AYP) is to assess the performance of all tested students within 10 demographic subgroups. AYP goals include evaluation on participation rates, proficiency standards, and attendance rates. Results may be assessed annually or averaged with the previous year’s calculations. The 10 different subgroups evaluated include the following: All students, Economically Disadvantaged Students, Asian/Pacific Islander Students, African American Students, American Indian/Alaskan Native Students, Hispanic Students, Multi-Racial Students, White Students, Students with Disabilities, and Students with Limited English Proficiency (ODE, 2005).

*Safe Harbor.* If AYP is not met through proficiency and participation rates, Safe Harbor offers an alternative way to meet standards. By demonstrating a 10 percent increase in the number of students who reach standards of proficiency, a school may satisfactorily obtain credit for AYP. Safe Harbor evaluates each subgroup individually, rewarding any subgroup that failed to meet other methods of proficiency (ODE, 2005).

*Consequences of Adequate Yearly Progress.* A school that meets AYP cannot be classified any lower than the Continuous Improvement status, regardless of scores on state indicators and performance index. However, Excellent or Effective schools that fail to meet AYP standards for three consecutive years are dropped into the Continuous Improvement category. Schools are eligible to advance out of the Continuous Improvement category when they demonstrate AYP standards for two consecutive years. In addition, a school that has otherwise failed to meet AYP or Safe Harbor standards may qualify for AYP if it demonstrates satisfactory value-added academic gains (ODE, 2005).

*Report Card Designations*

When awarding designations, Local Report Cards consider a school’s achievement on state indicators, performance index, value-added reports, and AYP status.
Classifications of Excellent with Distinction, Excellent, Effective, Continuous Improvement, Academic Watch, and Academic Emergency are awarded based upon: (1) The percentage of state indicators met or the total score earned on the Performance Index and (2) by determining if AYP standards have been met (ODE, 2005).

**Excellent with Distinction.** A school is awarded an Excellent with Distinction designation if it earned a green value-added score, exceeding expected growth rates. In addition, between 94 and 100 percent of state indicators or 100 to 120 points on the Performance Index must be earned. Finally, an Excellent with Distinction school must demonstrate Adequate Yearly Progress goals (ODE, 2005).

**Excellent designation.** A school must either obtain between 94 and 100 percent of state indicators or between 100 and 120 points on the Performance Index to be awarded an Excellent designation. In addition, although an Excellent designation typically means Adequate Yearly Progress goals were met, failure to meet AYP for three consecutive years will constitute a drop to the Continuous Improvement designation. Finally, a designation of Excellent can be achieved by obtaining a green value-added designation, thus constituting a move up on designation from Effective (ODE, 2005).

**Effective designation.** A school is classified as Effective if it obtains either between 73 and 93.9 percent of state indicators or a score of 90 to 99.9 points on the Performance Index. A school may move up a designation (from Continuous Improvement to Effective) or down a designation (from Excellent to Effective) based upon their value-added score. Finally, similar to the Excellent classification, although the majority of schools with this designation meet Adequate Yearly Progress goals, failure to meet AYP for three consecutive years will result in a drop to Continuous Improvement (ODE, 2005).

**Continuous Improvement designation.** A school may earn the Continuous Improvement designation by several different means. As previously described, a Continuous Improvement classification occurs when an Excellent or Effective school fails to meet AYP for three consecutive years. Moreover, based on value-added scores, a
school may move up from Academic Watch to Continuous Improvement. Conversely, regardless of low performance on state indicators or Performance Index, a school that meets AYP goals will be awarded a Continuous Improvement designation. Thus, any school meeting Adequate Yearly Progress will secure a designation no lower than Continuous Improvement (ODE, 2005).

In addition, a school will earn a Continuous Improvement classification by obtaining between 50 and 74.9 percent of state indicators or 80 to 89.9 points on the Performance Index. Finally, if a school designated as Academic Watch demonstrates a 10-point increase on the Performance Index it will move up on classification to Continuous Improvement. The increase in points must be earned during the previous two years; 3 of those points must be gained in the current year evaluated (ODE, 2005).

**Academic Watch designation.** The next report card classification is Academic Watch. Any school that has garnered this designation failed to meet Adequate Yearly Progress goals. A school will also fall under Academic Watch by earning between 31 and 49.9 percent of state indicators or by 70 and 79.9 points on the Performance Index. Additionally, a school may move up a designation (from Academic Emergency to Academic Watch) or down a designation (from Continuous Improvement to Academic Watch) based on their value-added score. Finally, a school classified in Academic Emergency may improve one designation to Academic Watch by demonstrating a 10-point increase on the Performance Index. As previously described, the increase in points must be earned during the previous two years and 3 of those points must be gained in the current academic year evaluated (ODE 2005).

**Academic Emergency designation.** The final report card classification is Academic Emergency. If a school is given this designation, it failed to meet AYP goals, earned fewer than 30.9 percent of state indicators, fewer than 69.9 points on the Performance Index, and received a red or yellow value-added designation (ODE, 2005).

Consequences of Local Report Card designations can be significant and severe. Continuously low classifications may result in critical changes in curriculum and
administration. While the objectives behind Local Report Card measures are intended to improve quality of instruction, considerable limitations exist in the current system.

**Limitations**

Value-Added is the Ohio Department of Education’s attempt at better reflecting the instructional effectiveness of an individual building or district and the growth it makes. Instead of focusing on achievement levels, it is a way of recognizing the growth students, teachers, schools, and districts make despite of achievement levels that may be below target. Accordingly, VA scores play a significant role in affecting a building or district’s overall Local Report Card designation. Thus it is necessary to explore the validity of VA in measuring its objectives.

*Criterion-referenced tests.* A significant weakness of the particular value-added model that the state of Ohio uses lies in the solitary reliance on student performance on criterion-referenced state assessments to measure instructional effectiveness. While some states utilize a more varied and robust model that incorporates different criterion- and norm-referenced measurements, Ohio relies solely on its own assessments, which are criterion-referenced. Criterion-referenced exams are designed to evaluate a predetermined set of academic skills, with benchmark goals established as a means of evaluating overall student performance in relation to state standards. Assessments of this nature provide a “snapshot” of student achievement at one point in the academic year, disregarding extraneous variables that may impact student performance during testing (Shapiro, 2004).

Moreover, VA measures attempt to link these test outcomes with expected growth trajectories, which may generate erroneous estimates of growth. Value-Added measures involve comparing scaled scores, assuming that each increase in scores represents an equal increase in curriculum difficulty. Unfortunately, it is impossible to judge difficulty objectively (Evergreen Freedom Foundation, 2001). Inevitably, further research is needed to explore concerns regarding the accuracy of a value-added measure in demonstrating academic gains (Amrein-Beardsley, 2008).
A similar drawback in employing criterion-referenced assessments for measuring student achievement is the issue of content-validity. It is critical that assessments encompass items that accurately reflect essential skill requirements. However, criterion-referenced assessments typically rely on evaluating skills with items derived from published curricula. Consequently, many criterion-referenced test scores represent a mastery of logical, rather than empirical knowledge (Shapiro, 2004).

A further limitation of criterion-referenced assessments is a general negligence in considering the influence of environmental variables. Amongst schools, disparities exist in demographics and composition of student populations. Test bias is a recurrent criticism of these measures. Bias exists in relation to students of varied socio-economic statuses and ethnicities, generating subjective testing content. The effects of these factors on student achievement should not be discounted. In his study, Heck (2000) pays highlights the significance of such factors, finding that characteristics such as socioeconomic status and ratio of special education students have a significant impact on school achievement outcomes. He reasoned that schools should not be penalized for external factors beyond their control and accountability measures must consider demographic discrepancies.

Thus, because criterion-referenced assessments are based on subjective contextual content, they may not be a valid measure of school performance. Finally, one criticism we make is that in relying solely on one measure may not best reflect student growth and that incorporating additional assessments may provide a better depiction of student progress.

OAT estimates of instruction. Marzano (2003) states that standardized tests are not sufficiently sensitive to actual student learning, which is essential in productively making data-based decisions about effective instruction and student growth. Within the elementary reading curriculum, standardized tests generally do not isolate and directly reflect the independent skills that are actively being taught within the classroom, such as reading fluency or comprehension. Thus, such evaluations tend to be insensitive to student progress.
Research conducted by the Delaware Statewide Academic Growth Assessment Pilot of 2007 (Rodel Foundation of Delaware, 2007) further supported Marzano’s assertions by revealing that grade-level assessments are not sensitive enough to demonstrate student growth. Moreover, when attempting to utilize a value-added model to demonstrate student progress, growth was nearly undetectable for students performing significantly above or below grade-level expectations. In fact, the further away from the grade level expectation a student scored on the state assessment, the lower the reliability of the assessment in detecting academic growth. These results expanded on Shapiro’s 2004 study which concluded that a considerable flaw of criterion-referenced tests is unreliability in detecting small changes in student progress.

In his study, Kupermintz (2003) examined the means by which the value-added model identified effective teachers. His study asserts that it is presumptuous to assume that scores on state achievement tests are exclusively the result of exceptional instruction. He concluded that the VA model greatly oversimplified the concept of effective instruction. Moreover, a common obstacle for schools is the failure to translate standardized test data into formative measures useful in modifying instruction. Therefore, the utility of a school’s value-added scores are greatly diminished because schools are unable to extract meaning from the data (Raudenbush, 2004).

Curriculum-Based Measurement (CBM)

Curriculum-Based Measurements (CBM) represent a prevalent and well-documented alternative to evaluating academic performance. Comprised of brief, ongoing formative assessment, CBM has successfully been used to monitor academic progress. Marzano (2003) maintains that prompt, detailed feedback is vital to student achievement. Formative assessment was found by Black and William (1998) to improve student achievement by 70 percent of a standard deviation. However, this type or use of formative assessment may not include CBM’s, as they are not administered daily to each student within a classroom. Recent research has also indicated that formative assessment is a valuable tool for informing instruction. Utilizing formative assessment within the
classroom is an effective way to provide in-depth detail in a timely manner to both students and teachers (Marzano, 2003; Stiggins & Chappuis, 2008).

Madelaine and Wheldall’s (2004) review of the literature, found that CBM has historically proven to be a sensitive, objective tool effective in evaluating the impact of instruction and informing academic modification. CBM has successfully been used to formatively monitor reading, math, and spelling progress, with results strongly relating to overall student performance. Measurements are generally collected on a trimester schedule during the school year to evaluate student progress within a given subject. Benchmark scores are evaluated to establish student performance in relation to peer norms, with rate of improvement similarly evaluated amongst benchmark rates to compare student progress (Cusumano, 2007).

CBM Growth. Teachers that utilize CBM data and feedback when evaluating responsiveness to instruction are better able to accurately develop ambitious progress goals and modify instruction as necessary (Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993, Haager, Klingner, & Vaughn, 2007). Fuchs et al. (1993) established standard norms for weekly CBM growth to aid in developing acceptable progress rates. Growth rates were calculated by grade level, revealing the basis for realistic and ambitious growth trajectories. Deno, Fuchs, Marston, and Shinn (2001) later established growth standards for special education students, distinguishing between typical and effective instructional practices; thus establishing reliable improvement rates for a multitude of students.

Good et al. (2002) constructed a report regarding the technical properties of Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Benchmark Assessment, a highly regarded CBM reading system. Results of the study yielded measures of reading progress that systematically increase with subsequent benchmarks throughout the school year. The exception to this trend is found only in between grades, from the spring benchmark of one year to the fall benchmark of the following year. This decline in progress is generally attributed to summer regression, which is instigated by a lack of academic instruction throughout summer break. These results are consistent with the
meta-analysis that Cooper et al. (1996) conducted regarding summer break and academic losses.

**CBM in Conjunction with Achievement Tests.** It is reasonable to consider CBM in relation to state achievement test outcomes based on the functional and empirical evidence supporting CBM use as a valuable tool in progress monitoring. A review of the literature reveals that CBM has the capacity to predict future academic success. When CBM scores of oral reading were compared to state reading achievement test scores, ORF measures were found to predict performance on state achievement tests to a moderate and strong degree (Crawford, Tindal, and Stieber, 2001; Hintze and Silberglitt 2005; McGlinchey & Hixson, 2004; Shapiro, Keller, Lutz, Santoro, & Hintze, 2006; Stage & Jacobsen, 2001). Determining specifically which ORF scores, fall, winter, or spring, best predict student performance on the OAT would provide schools with an efficient and effective method of monitoring student progress and predicting the likelihood of proficient performance on state reading achievement tests.

As a result, this study will examine the relationship between fall, winter, and spring ORF scores and OAT performance. Specifically, fall, winter, and spring ORF benchmarks from third grade, along with fall and winter ORF scores from fourth grade will be evaluated as possible predictors of future OAT performance. These will also be compared to the current value-added system, which evaluates student progress by longitudinally evaluating OAT performance to the students’ performance from previous years. It is important to evaluate both methods of predicting future student performance to determine the method most effective in accurately predicting OAT performance. The serious consequences connected with ineffective instruction and poor performance on state achievement tests further support the need to determine the most effective method for predicting future success or failure.

Previous studies examining CBM scores and state achievement tests have produced mixed results on which CBM score functions as the strongest predictor of future performance. Stage and Jacobsen (2001) contend that fall ORF scores are the strongest predictor, while Shapiro et al. (2006) concluded that in fact winter and spring
scores were the best predictors. In addition, Springer (2008) found that “the further a failing student is from [the] state’s performance threshold, the greater their predicted fall-to-spring test score gain…” This indicates, regarding low achieving students, that the greater their deficit compared to state expectations, the greater the predicted CBM score gains throughout the school year. Thus, the most recent CBM scores obtained best reflects failing students’ performance level prior to the state achievement test administration.

Finally, Schulte, Villwock, Whichard, and Stallings (2001) suggest that CBM could provide a powerful addition to current student performance assessments. In 2002, the U.S. Department of Education conducted an evaluation to ensure that CBM in reading closely aligned with Adequate Yearly Progress outcomes. Thus, the findings supported the assumption that reading CBMs may effectively be employed to monitor Adequate Yearly Progress (U.S. Department of Education, 2002). It may additionally be reasonable to project that CBM could be successfully utilized in developing more accurate and functional growth trajectories when individual student data is aggregated for use in, or as a substitute for, value-added measures.

Curriculum-based measurement systems. Currently, districts generally employ school-wide curriculum-based measurement (CBM) assessments with the intention of identifying and monitoring struggling students, students “at risk” of academic difficulties, and students clearly far below average, warranting intensive remediation or special education services. Subsequently, curriculum-based measures are given to all students as a screening procedure (Deno, Fuchs, Marston, & Shin, 2001). Standardized assessments have since been published for use by buildings to interpret and manage the immense data collected.

One such system is AIMSweb, a web-based progress monitoring system, which incorporates standardized curriculum-based measures of early literacy skills, reading fluency, and comprehension, among other academic areas. The primary objective of this system is to identify educational needs early so that they may be addressed within the classroom as necessary (Shinn & Shinn, 2002). Additionally, AIMSweb assessments are
designed for progress monitoring, with the rate of improvement being calculated amid fall, winter, and spring benchmarks. Thus the extent to which students are benefiting from instruction can be determined (Daly, Glover, & McCurdy, 2006). Districts that purchase AIMSweb are provided with training and technical support, imparting a greater tendency for assessments to be implemented with integrity (Doll & Cummings, 2008). Although this is an advantage over other management systems, purchasing the AIMSweb program can be fiscally demanding.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS) is an outcomes-driven model of assessment used to identify students in need of additional instructional support. This system is also designed to monitor the progress of student performance to further inform instruction within the classroom. Measures of the following skills are evaluated: Initial Sound Fluency (ISF), Phoneme Segmentation Fluency (PSF), Nonsense Word Fluency (NWF), Letter Naming Fluency (LNF), and Oral Reading Fluency (ORF). These skills are recognized by the National Reading Panel as foundational skills for successful reading outcomes (National Reading Panel, 2000). In addition, trimester benchmarks are established to evaluate the minimal performance required for proficient outcomes.

At the systems level, DIBELS can be used to evaluate the overall effectiveness of instruction and curriculum in educating all students. An outcomes report is generated that informs districts of the percentage of students that have achieved necessary literacy outcomes (Good, Gruba, Kaminski, 2001). To access this information, districts must purchase and subscribe to the DIBELS online data management system. However, because DIBELS assessment probes are available online for free public use, weaknesses subsist in issues of treatment integrity. For research purposes, DIBELS data was utilized despite this disadvantage. Its wide availability to school districts, regardless of ability to purchase the data management system, lends itself to use in this study.
Method

Participants

One school district in suburban Southwestern Ohio participated in this study, providing data from 746 students. During the 2008-2009 school year, the district served approximately 9520 students of various races, ethnicities, and backgrounds: 74.6% White, non-Hispanic, 20.0% Economically Disadvantaged, 13.5% Students with Disabilities, 12.2% Black, non-Hispanic, 4.8% Multi-Racial 4.5% Hispanic, 3.8% Limited English Proficient, 3.7% Asian or Pacific Islander, 0.2% American Indian or Alaska Native, and no migrant students (ODE, 2009). Two elementary schools serving student grades 1st through 4th agreed to release tests scores for male and female students in 4th grade during the 2008-3009 school year. All data was archival, thus neither informed consent nor direct participation was necessary by students or parents/guardians. The Family Educational Rights and Privacy Act (FERPA) allows for disclosure of unidentified student growth summaries without consent. No identifying student information was collected or disseminated as a result of school participation in this study. Both building principals, the district curriculum director, and district assistant superintendent approved and agreed upon participation in this study.

Fourth grade was chosen for this study on the basis of two rationales: (1) Value-Added growth trajectories are first established for fourth grade students based on OAT performance in third grade; thus, fourth grade represents the first value-added score allotted to students/buildings; (2) According to Silberglitt, Burns, Maydun, and Lail (2006), ORF performance is most highly correlated with achievement test scores in lower grades. Because reading growth begins to plateau as students progress through school, it is reasonable to assume that reading fluency is more closely tied to OAT performance in fourth grade, as opposed to later grades.

Measures

Dynamic Indicators of Basic Early Literacy Skills (DIBELS) assessments, specifically Oral Reading Fluency (ORF), and Ohio Achievement Tests (OAT) were
utilized. Both assessments were administered by qualified and trained personnel in each elementary school participating. Grade appropriate ORF probes were administered. Thus, 3rd grade students completed testing with 3rd grade materials and 4th grade students completed testing with 4th grade materials for each assessment. Oral Reading Fluency probes require that students read three passages, each for one minute. Incorrect responses, as outlined in the administration and scoring manual, are noted, resulting in a final score measured by words read correctly per minute (wcm). The median wcm and error scores are then recorded as a student’s final score. DIBELS testing materials, including standardized instructions and scoring procedures, were obtained by school personnel from the following website: https://dibels.uoregon.edu/.

The Ohio Achievement Test is a criterion-referenced test developed and distributed by the State of Ohio Department of Education. Each student is required to complete the test in the fall and spring of 3rd grade and then the fall of each school year through 10th grade. This test is administered by qualified and trained school personnel and then scored by an outside agency. Schools then receive the results, sending home each student’s results for parental/guardian review.

Procedure

The School Psychologist at two elementary schools in a suburban school district in Southwestern Ohio was contacted regarding participation in this study. The building principals, district director of curriculum, and district assistant superintendent were each contacted for review and approval. An application to conduct research was also required, completed, and approved. The director of data and testing complied all necessary ORF and OAT scores from 3rd and 4th grades for the students in the 2009-2010 5th grade class. Data on a total of 746 students was obtained, although data was utilized only if all nine data points were available. Thus, a total of 575 participants were used in this study.

All assessments were administered by trained and qualified school personnel. Trimester ORF benchmark scores during 3rd and 4th grades were obtained during Fall, Winter, and Spring. The 3rd grade OAT was administered in Fall, before the Fall ORF
testing, and in the Spring, shortly after the Spring ORF testing. Although typically only teachers, principals, and assistant principals administered the OAT, DIBLES probes were administered by a team of teachers, school psychologists, speech and language pathologists, and education assistants. This team also scored all DIBELS assessments, which were entered into a database by the district director of data and testing. Although OAT assessments were scored outside of the school, scores were also entered into a district database.

The director of data and testing compiled all ORF and OAT data in a Microsoft Excel document. All identifying student information used in the compilation of the document was deleted prior to distribution to the researcher. Any document rows that did not contain all nine data points were then eliminated by the researcher prior to analysis. Participants were not assigned to groups, thus there is no control group in this study. All data was collected on a single group of students from two participating schools within the same district. Several analyses were conducted: Pearson Product-Moment Correlation, Multiple Regression, and Linear Regression. There were six independent variables in this study: Fall 3rd grade ORF, Winter 3rd grade ORF, Spring 3rd grade ORF, Spring 3rd grade OAT, Fall 4th grade ORF, Winter 4th grade ORF. There was one dependent variable in this study: Spring 4th grade OAT. All variables are displayed in Table 1.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Dependent and independent variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2007 Fall</td>
</tr>
<tr>
<td>Grade 3 ORF</td>
<td>✓</td>
</tr>
<tr>
<td>Grade 3 OAT</td>
<td>✓</td>
</tr>
<tr>
<td>Grade 4 ORF</td>
<td></td>
</tr>
<tr>
<td>Grade 4 OAT</td>
<td></td>
</tr>
</tbody>
</table>

A Pearson product moment correlation was used to evaluate the null hypothesis that no relationship exists between Oral Reading Fluency scores from current years and
Ohio Achievement Test scores or Oral Reading Fluency scores from previous years and future performance on the Ohio Achievement Test. Additional statistical analyses were conducted based on the results of the correlational analysis. The following research questions were examined:

Research Question 1: What is the predictive relationship between grade 3 Fall ORF scores and grade 4 OAT performance?

Null Hypothesis 1: No relationship exists between grade 3 Fall ORF scores and grade 4 OAT performance;

Research Question 2: What is the predictive relationship between grade 3 Winter ORF scores and grade 4 OAT performance?

Null Hypothesis 2: No relationship exists between grade 3 Winter ORF scores and grade 4 OAT performance;

Research Question 3: What is the predictive relationship between grade 3 Spring ORF scores and grade 4 OAT performance?

Null Hypothesis 3: No relationship exists between grade 3 Spring ORF scores and grade 4 OAT performance;

Research Question 4: What is the predictive relationship between grade 4 Fall ORF scores and grade 4 OAT performance?

Null Hypothesis 4: No relationship exists between grade 4 Fall ORF scores and grade 4 OAT performance;

Research Question 5: What is the predictive relationship between grade 4 Winter ORF scores and grade 4 OAT performance?

Null Hypothesis 5: No relationship exists between grade 4 Winter ORF scores and grade 4 OAT performance;
Research Question 6: What is the predictive relationship between OAT scores in Spring of grade 3 and grade 4 OAT performance?

Null Hypothesis 6: No relationship exists between grade 3 Spring OAT performance and grade 4 OAT performances;

Results

Several separate statistical analyses were conducted to evaluate the relationship between the five Oral Reading Fluency (ORF) scores and two Ohio Achievement Test (OAT) scores. Specifically, each variable was evaluated for predictors of performance on the 2009 OAT. Correlation coefficients were computed using a Person product-moment correlation and a required $p$ value of .01 for significance. The results of the correlational analysis presented in Table 2 show that all variables have a moderate predictive relationship with the 2009 OAT.

<table>
<thead>
<tr>
<th>Variable</th>
<th>OAT May 09</th>
<th>ORF Fall 07</th>
<th>ORF Winter 08</th>
<th>ORF Spring 08</th>
<th>ORF Fall 08</th>
<th>ORF Winter 09</th>
<th>OAT May 08</th>
</tr>
</thead>
<tbody>
<tr>
<td>OAT May 09</td>
<td></td>
<td>.627**</td>
<td>.626**</td>
<td>.616**</td>
<td>.632**</td>
<td>.652**</td>
<td>.639**</td>
</tr>
<tr>
<td>ORF Fall 07</td>
<td></td>
<td></td>
<td>.924**</td>
<td>.903**</td>
<td>.901**</td>
<td>.897**</td>
<td>.576**</td>
</tr>
<tr>
<td>ORF Winter 08</td>
<td></td>
<td></td>
<td></td>
<td>.915**</td>
<td>.907**</td>
<td>.901**</td>
<td>.565**</td>
</tr>
<tr>
<td>ORF Spring 08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.917**</td>
<td>.907**</td>
<td>.565**</td>
</tr>
<tr>
<td>ORF Fall 08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.925**</td>
<td>.570**</td>
</tr>
<tr>
<td>ORF Winter 09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.575**</td>
</tr>
</tbody>
</table>

**$p < .01$  
*p < .05*
Upon closer examination it was noted that the two strongest correlations were Winter 2009 ORF (r= .652) and Spring 2008 OAT (r=.639) scores. Although these two variables have a stronger predictive relationship, it should be noted that there is not a significant difference between the predictive strength of these two variables and the other four variables when predicting performance on the Spring 2009 OAT.

A bivariate linear regression analysis was also conducted to evaluate the predication of Spring 2009 OAT performance from Spring 2008 OAT performance and Spring 2009 OAT performance from Winter 2009 ORF scores. First, the results of the analysis between the 2008 and 2009 OAT scores were examined. The scatterplot of these two variables, shown in Figure 1, indicates that the two variables are linearly related. The 95% confidence interval for the slope ranges from .702 to .856 (Table 3), which does not contain the value of zero. Thus, Spring 2008 OAT performance is significantly related to performance on the Spring 2009 OAT assessment.

Figure 1. Scatterplot depicting relationship between Spring 2008 OAT performance and Spring 2009 OAT performance.
Table 3

*Correlation coefficients between Spring 2008 OAT and Spring 2009 OAT performance*

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE</th>
<th>β</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>108.359</td>
<td>16.442</td>
<td>--</td>
<td>6.59</td>
<td>.000</td>
</tr>
<tr>
<td>OAT May 08</td>
<td>.779</td>
<td>.039</td>
<td>.639</td>
<td>19.880</td>
<td>.000</td>
</tr>
</tbody>
</table>

*Note: R Square = .408. Adjusted R Square = .407.*

**p < .01

The regression equation for predicting future OAT performance with previous OAT performance is

\[
\text{Predicted OAT performance} = .779 \times \text{Spring 2008 OAT} + 108.35
\]

The correlation between Spring 2008 OAT and Spring 2009 OAT was .63, thus accuracy in predicting Spring 2009 OAT performance was strong. Approximately 41% of the variance was accounted for by the Spring 2009 OAT performance and its relationship with the previous school years Spring OAT performance. Thus, as hypothesized, prior OAT performance can be utilized to predict future OAT performance.

The analysis results evaluating predication of Spring 2009 OAT performance from Winter 2009 ORF scores produced the scatterplot shown in Figure 2, illustrating that these two variables are also linearly related. The 95% confidence interval for the slope, .458 to .554 (Table 4), does not contain the value of zero. Therefore, the Winter 2009 ORF performance score is significantly related to Spring 2009 OAT performance.
The regression equation for predicting future OAT performance with the most recent ORF scores is

**Table 4**

*Correlation coefficients between Winter 2009 ORF and Spring 2009 OAT performance*

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE</th>
<th>β</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>372.742</td>
<td>3.131</td>
<td>--</td>
<td>119.039</td>
<td>.000</td>
</tr>
<tr>
<td>ORF Winter 09</td>
<td>.506</td>
<td>.025</td>
<td>.652</td>
<td>20.561</td>
<td>.000</td>
</tr>
</tbody>
</table>

*Note: R Square = .483. Adjusted R Square = .474*

**p < .01**

The regression equation for predicting future OAT performance with the most recent ORF scores is
Future OAT performance = 0.506 (x) + 372.74

The correlation between Winter 2009 ORF and Spring 2009 OAT was .65, thus accuracy in predicting Spring 2009 OAT performance was strong. Approximately 42% of the variance of the Spring 2009 OAT performance was accounted for by its linear relationship with the most recent ORF performance scores. Thus, as hypothesized, the most recent ORF performance available can be utilized to predict performance on an upcoming OAT assessment.

Finally, a multiple linear regression analysis was conducted to examine the degree to which each ORF and OAT score predicted future OAT performance. The predictors were the five ORF scores and one Spring 2008 OAT performance, while the criterion was the Spring 2009 OAT performance. A Backwards model selection technique was used and all models produced a correlation coefficient of 0.729 to 0.727 (Table 5). This indicates that approximately 53% of the variance of the Spring 2009 OAT performance can be accounted for by the six variables.
Although, considering all or any combination of the six variables to accurately predict future OAT performance appears possible, the fifth predictive model may prove the most practical. This model utilizes the most recent previous OAT and ORF performances to predict future OAT performance and yielded the highest partial and part correlations respectively: OAT (.426, .323) and ORF (.452, .347), as shown in Table 6. Therefore, future OAT performance can be predicted with a high degree of accuracy utilizing only two, instead of six variables.
Table 6

*Part and partial Correlations of the Predictors with Spring 2009 OAT Performance*

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE</th>
<th>( \beta )</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Constant)</td>
<td>196.645</td>
<td>16.250</td>
<td>--</td>
<td>11.855</td>
<td>.000</td>
</tr>
<tr>
<td>OAT May 08</td>
<td>.481</td>
<td>.043</td>
<td>.395</td>
<td>11.255</td>
<td>.000</td>
</tr>
<tr>
<td>ORF Winter 09</td>
<td>.330</td>
<td>.027</td>
<td>.425</td>
<td>12.107</td>
<td>.000</td>
</tr>
</tbody>
</table>

**p < .01

Discussion

As high-stakes testing continues to carry significant consequences for educators, schools will continue to monitor student growth within the curriculum using various measurement methods. Curriculum-based measurements (CBM) are designed specifically for this purpose. As their use increases across the country, so will the importance of understanding relationships between these brief measurement scores and state achievement test results.

The purpose of this study was to evaluate the utility of CBM oral reading fluency (ORF) results to predict future performance on the Ohio Achievement Test (OAT). One objective was to examine the relationship between ORF scores from the same and previous years and OAT scores. The results indicated that any ORF scores recorded prior to OAT administration had a moderate predictive relationship (range: \( r = .57 \) to \( .65 \)). This was also true for ORF scores recorded after OAT administration. However, there is no practical application for this finding.

When looking at the Spring 2009 OAT, oral reading fluency scores recorded from the same school year (2008-2009) had a slightly stronger correlation than ORF scores recorded during the previous school year (2007-2008). Additionally, when examining the Spring 2008 OAT, oral reading fluency scores recorded during the next school year (2008-2009) were also shown to have a moderate predictive relationship. Although this
last relationship demonstrates the strength and accuracy of the predicative relationship, this finding is not applicable or useful to practice in schools due to the assessment administration timeline.

Another objective of this study was to examine the degree to which ORF scores were able to predict OAT scores. As discussed, all ORF scores had a moderate predictive relationship with the state assessments. Thus, all variables could theoretically be utilized to predict OAT performance in the future. It is important to note that the Winter 2009 ORF score had a slightly stronger correlation with Spring 2009 OAT performance than the previously recorded ORF scores. When examining models to predict Spring 2009 OAT performance, one model comprising the most recent ORF score recorded (Winter 2009) and the most recent OAT score (Spring 2008) had a stronger correlation with future performance on the state assessment than other models. This model also demonstrated lower standard error of measurement (SEM). However, the difference in correlation coefficients and SEM between this predictive model and the four others examined was very small. Thus, utilizing the final model would likely prove the most efficient for school personnel.

The final objective of this study was to evaluate the relationship between OAT scores from 3rd grade Spring to 4th grade Spring compared to ORF scores over the same time period. When examining the predictive model utilizing the most recently administered ORF and OAT scores it should be noted that the assessments account for 34.7% and 32.3% of the variance respectively. A smaller degree of unique variance for each measure would indicate a higher degree of overlap between the tests and that one assessment could be selected to predict future performance instead of utilizing both. However, due to the moderate degree of variance accounted for by each assessment, one should not be used independently to predict future OAT performance.

Therefore, this study indicated that any past ORF or OAT scores could be used to assist educators when evaluating which students are likely to pass state achievement tests in the future and which are not. This could be done by examining which students have performed at the proficient or higher level, indicating a greater likelihood of passing in
the future. Likewise, examining which students have failed these assessments in the past may be an indication of a greater likelihood of not passing in the future. However, using all available data can be a daunting task as students continue to be promoted to the next grade level. The results of this study indicate that educators could use the most recently recorded ORF and OAT scores as a way to effectively and efficiently predict the likelihood of future success on state achievement tests.

There are several implications for future research as a result of this study. Examining additional measures, such as 1st grade Oral Reading Fluency, Nonsense Word Fluency, or Phoneme Segmentation Fluency may provide earlier indicators of students at-risk for failure on state achievement tests. Replicating this study with urban and rural populations could add to the research base and extend findings to similar regions of the country. Finally, including a reading comprehension measure, such as the MAZE offered by AIMSweb, could allow researchers to evaluate the predictive relationship between oral reading fluency and reading comprehension scores with state achievement tests.

**Limitations**

There are several limitations to this study. First, data was collected from only one school district in suburban Southwestern Ohio. Although this school district provides educational services to a diverse population, it may not be representative of the United States population or that of other regions across the country. Additionally, it may not represent urban or rural districts, limiting the utility of the research results.

Second, the DIBELS curriculum-based measurement system was utilized by the participating school district to collect ORF data. A similar system, AIMSweb, is available for purchase. Upon purchase, districts are provided training and technical support, imparting a greater tendency for assessments to be implemented with integrity (Doll & Cummings, 2008). Conversely, DIBELS is available online, at no cost and can be obtained and utilized without receiving proper training, thus creating issues concerning integrity in administration and scoring of assessment probes. Third, this study did not
examine fluency data collected in 1st and 2nd grades, which could provide an earlier indication and better profile of which students are likely to pass future state assessments.

**Conclusion**

In summary, the present study adds to a growing body of research examining the relationship between curriculum-based measurements and state achievement tests. The indication that these brief assessment tools can be used to predict performance on state tests is encouraging for educators searching for an efficient and effective way to monitor the progress of their students. In addition, because CBM results inform teachers of student growth and performance, educators are given the opportunity to supplement and change instruction as necessary to meet the ever changing needs within their schools.
References


