1-1-2010

Oral Reading Fluency And Maze Selection For Predicting 5Th And 6Th Grade Students' Reading And Math Achievement On The Illinois Standards Achievement Test

Samuel Whitley
Eastern Illinois University

This research is a product of the graduate program in Psychology at Eastern Illinois University. Find out more about the program.

Recommended Citation
Whitley, Samuel, "Oral Reading Fluency And Maze Selection For Predicting 5Th And 6Th Grade Students' Reading And Math Achievement On The Illinois Standards Achievement Test" (2010). Masters Theses. 607.
http://thekeep.eiu.edu/theses/607
THESIS MAINTENANCE AND REPRODUCTION CERTIFICATE

TO: Graduate Degree Candidates (who have written formal theses)

SUBJECT: Permission to Reproduce Theses

The University Library is receiving a number of request from other institutions asking permission to reproduce dissertations for inclusion in their library holdings. Although no copyright laws are involved, we feel that professional courtesy demands that permission be obtained from the author before we allow these to be copied.

PLEASE SIGN ONE OF THE FOLLOWING STATEMENTS:

Booth Library of Eastern Illinois University has my permission to lend my thesis to a reputable college or university for the purpose of copying it for inclusion in that institution's library or research holdings.

__________________________  ________________
Author's Signature          Date

I respectfully request Booth Library of Eastern Illinois University NOT allow my thesis to be reproduced because:

__________________________
I plan to develop further.

__________________________  ________________
Author's Signature          Date

This form must be submitted in duplicate.
Oral Reading Fluency and MAZE Selection for Predicting 5th and 6th Grade Students' Reading and Math Achievement on the Illinois Standards Achievement Test

BY

Samuel Whitley

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Specialist in School Psychology

IN THE GRADUATE SCHOOL, EASTERN ILLINOIS UNIVERSITY
CHARLESTON, ILLINOIS

2010 YEAR

I HEREBY RECOMMEND THIS THESIS BE ACCEPTED AS FULFILLING THIS PART OF THE GRADUATE DEGREE CITED ABOVE

THESIS COMMITTEE CHAIR DATE

DEPARTMENT/SCHOOL CHAIR OR CHAIR'S DESIGNEE DATE

THESIS COMMITTEE MEMBER DATE

THESIS COMMITTEE MEMBER DATE
Table of Contents

I. List of Tables ....................................................................................... 3

II. List of Figures .................................................................................. 4

III. Dedication ...................................................................................... 5

IV. Acknowledgements ......................................................................... 6

V. Abstract .......................................................................................... 7

VI. Literature Review
   1) Curriculum-based Measurement ................................................. 12
   2) Curriculum-based Measurement and Benchmarking ................. 15
   3) The Relationship Between Curriculum-based Measurement and
      Reading Performance on State Achievement Tests ..................... 18
   4) Relationship Between Reading and Mathematical Abilities .......... 24
   5) Curriculum-based Measurement and Mathematics Performance on
      State Achievement Tests ............................................................... 28
   6) Purpose and Hypotheses ............................................................... 30

VII. Method
   1) Participants ................................................................................ 32
   2) Materials ................................................................................... 32
   3) Procedure .................................................................................. 40
   4) Design and Data Analysis ........................................................... 40

VIII. Results ....................................................................................... 41

IX. Discussion .................................................................................... 51
   1) Conclusions and Implications ...................................................... 55

X. References ..................................................................................... 57
List of Tables and Figures

Table 1. ISAT Reading and Mathematics Scale Score Ranges and Performance Definitions by Grade .................................................. 12

Table 2. Diagnostic Efficiency Statistics Template ................................................. 19

Table 3. ISAT Minimum Reading and Math Scale Scores Required to Meet Standards .............................................................................. 39

Table 4. Descriptive Statistics for CBM Scores and ISAT Subscales for Fifth Grade .............................................................................. 42

Table 5. Descriptive Statistics for CBM Scores and ISAT Subscales For Sixth Grade .............................................................................. 42

Table 6. Correlation Matrix between CBM Scores and ISAT Subscales for Fifth Grade .............................................................................. 44

Table 7. Correlation Matrix between CBM Scores and ISAT Subscales for Sixth Grade .............................................................................. 45

Table 8. Summary of Simple Linear Regression Analysis for 5th Grade CBM Predicting ISAT Reading Performance ..................................... 46

Table 9. Summary of Simple Linear Regression Analysis for 5th Grade CBM Predicting ISAT Mathematics Performance ............................. 46

Table 10. Summary of Simple Linear Regression Analysis for 6th Grade CBM Predicting ISAT Reading Performance ..................................... 47

Table 11. Summary of Simple Linear Regression Analysis for 6th Grade CBM Predicting ISAT Mathematics Performance ............................. 47

Table 12. 5th Grade Cut Scores ........................................................................... 48

Table 13. 6th Grade Cut Scores ........................................................................... 48

Table 14. Diagnostic utility statistics for the 5th grade sample ............................ 50

Table 15. Diagnostic utility statistics for the 6th grade sample ............................ 51
List of Figures

Figure 1. Scatterplot Utilized to Determine Diagnostic Utility Statistics..........49
Dedication

This project is dedicated those individuals who have supported me throughout the course of my life, namely my parents and my wife. There are no means through which I could ever express my gratitude or reconcile my debt for their endless support or for providing me with skills, resources, and desire to succeed. Furthermore, to my wife, Amanda, I could not have completed this process without your understanding or sympathetic ear.
Acknowledgments

There are numerous individuals who deserve thanks for without their assistance this project would not have been completed. These individuals include Tammy Claypool, Erin Meyer, Erin Conn, Debbie Schoppenhorst, and Ann Kuehnel. Finally, there are a small group of individuals who deserve special thanks; Dr. Ronan Bernas for substantially improving the content of this manuscript through his statistical expertise, Professor Cathy Schoonover for improving the structure of this manuscript through her editing recommendations, and Dr. Assege HaileMariam. Were it not for her advice, flexibility, and calm demeanor this project certainly would not have been completed and I believe I am better for having had the pleasure of having her as a professor.
Abstract

This study sought to examine the relationship between two reading curriculum based measures (oral reading fluency and MAZE) and the reading and mathematics subsections of the Illinois Standards Achievement Test (ISAT) at the 5th and 6th grades. It was predicted that MAZE would be better at predicting who would pass the ISAT than oral reading fluency. Bivariate correlations were conducted between each curriculum based measure and each subsection of the ISAT. Simple linear regression analyses were also conducted to derive cut scores for predicting ISAT performance. The ability of each curriculum based measure to predict student performance was evaluated by diagnostic utility statistics. These analyses indicate that MAZE may be equally suited to predicting student performance as oral reading fluency at both grade levels. In conclusion, reading CBM measures may be used to adequately predict student performance on the ISAT reading subsection at the upper elementary school levels.
Oral Reading Fluency and MAZE Selection for Predicting 5th and 6th Grade Students’ Reading and Math Achievement on the Illinois Standards Achievement Test

This study examined and compared the ability of two Curriculum Based Measurement (CBM) measures, Oral Reading Fluency (ORF) and MAZE, a reading comprehension measure, to predict fifth and sixth grade student performance on the reading and mathematics sections of a state-wide standardized achievement test in the state of Illinois. In Illinois, students who are in the third through eighth grades are tested yearly in the area of reading and math. The purpose of this testing is to measure students’ progress toward grade level proficiency in reading and math throughout their elementary and middle school years as mandated by the No Child Left Behind (NCLB, 2004). In 2001, President George W. Bush announced his endorsement of federal legislation meant to address the nation’s educational shortcomings and remediate those deficiencies experienced by students throughout the nation within the core academic subjects of math and reading. This legislation was christened with a bold name, The No Child Left Behind Act of 2001. The NCLB Act reauthorized the Elementary and Secondary Education Act of 1965, as well as outlined principles and strategies meant to strengthen the American educational system (U.S. Department of Education, 2004).

Among the espoused principles and strategies was a new emphasis on statewide accountability systems concerning students’ math and reading proficiency (U.S. Department of Education, 2004). These systems require that all states elucidate explicit and challenging learning standards for all students, develop yearly objectives to ensure that all students obtain grade-level proficiency in math and reading by the school year 2013-2014, and implement yearly statewide testing of all students in grades 3 through 8.
Each state must be able to demonstrate adequate yearly progress (AYP) towards the mandated goal of universal grade-level proficiency in math and reading for students in grades 3 through 8. Fuchs and Fuchs (2004) define AYP as the minimum rate of growth a school needs to make from its initial proficiency level determined at the time of the introduction of NCLB in order to obtain the universal proficiency level within the dictated timeframe. Schools that fail to make AYP are subject to several different corrective actions such as extended school days or management of the school by the state (Illinois Interactive Report Card, n.d.).

In order to determine AYP many states have utilized the implementation of end of the year statewide testing protocols, otherwise known as outcome measures, developed by their respective state boards of education, which are aligned to their state learning standards. These “high stakes” tests play a large role in not only measuring the progress of individual students and funding issues, but they have also forced some educators to place an increased emphasis on the early identification and intervention of students’ educational deficiencies. The logic behind this shift is that the earlier educational difficulties are detected, the greater the probability that these difficulties will be successfully remediated (Juel, 1988), and due to these early educational interventions the greater the number of students who may be determined to be proficient in reading and math; thus increasing a school’s likelihood of obtaining adequate yearly progress as determined through state wide outcome measurements. In order to identify students as early as possible, many school districts have turned to the use of benchmarking with an assessment method known as curriculum-based measurement (CBM). This process has not only shown considerable promise in improving educational outcomes of students, but
it has also created much controversy concerning what measures are best suited to adequately determine which students require educational interventions. Nowhere is this debate more prevalent than within the realm of benchmarking the reading abilities of upper elementary school students through the use of ORF or MAZE. This controversy, the measures under scrutiny, and the process of benchmarking are discussed in later sections of this study.

Why do educators at the local, state and national level see reading and math proficiency important for students? The concern for improving the reading achievement of American school children has a long history (National Research Council, 1998) because of its implication for the individual. One of the most serious implications of poor reading is the limited employment opportunity and resulting loss of competitive wages. It is estimated that 75% of adults who are unemployed have limited literacy skills (Family Literacy, 2006). Adults with low literacy skills are frequently unemployed and earn lower wages than individuals with high literacy skills (Kirsch et al., 2001). This, in turn, costs society not only tax revenue, but also in direct expenses to support unemployed and underemployed individuals. Students with low math skills do not seem to fare any better. Unfortunately a majority of American students are graduating high school without sufficient proficiency in mathematics, hindering higher education and future job performance (Harniss, Stein, & Carnine, 2002). On the other hand, students who have good math skills are three times more likely to attend college and earn more money as a result (NEA Today, 1998).

In addition to the critical importance of reading and math proficiency for success in life, a relationship between reading and math has been established. This relationship is
discussed later. For now, briefly, it has been indicated that reading may be an important
access skill to mathematics proficiency (Thurber et al, 2002; Helwig et al, 1999) and
reading has been shown to be correlated with mathematics performance on standardized
achievement tests (Thurber et al, 2002). Robinson, et al. (2002) suggested a similar path
for reading and math difficulties, weakness in phonological processing and encoding and
retrieval problems. Given this information, if ORF and MAZE predict students’ reading
and math scores, as hypothesized in this study, the use of ORF and MAZE in progress
monitoring and prevention of reading and math problems can be strengthened.

Before discussing the utility of ORF and MAZE for predicting reading and math
achievement, the means for assessing yearly achievement must be addressed. The Illinois
State Board of Education dictates that the percentage of children meeting or exceeding
grade level proficiency in reading and mathematics must increase each year until 100%
of Illinois’ students are proficient. For example, in 2010, 77.5% of students must be
deemed proficient, while in 2011, 85% of students must be proficient (Illinois Interactive
Report Card, n.d.).

In order to determine if schools are meeting the prescribed percentage of children
demonstrating grade level skills in reading and mathematics, students are administered
the Illinois Standards Achievement Test (ISAT) during the spring of each school year.
The ISAT measures the extent to which students are meeting the Illinois Learning
Standards (Illinois State Board of Education Division of Assessment, 2008). While
students in grades three through eight are administered the ISAT, the present study is
concerned only with the reading and mathematics sections administered to students in
grades five and six. Thus, the current study investigated if ORF and MAZE scores predict
performance level on the ISAT. Scale score range and performance definitions on the ISAT for reading and math are presented in Table 1.

Table 1.

*ISAT Reading and Mathematics Scale Score Ranges and Performance Definitions by Grade*

<table>
<thead>
<tr>
<th>Grade</th>
<th>Academic Warning</th>
<th>Below Standards</th>
<th>Meets Standards</th>
<th>Exceeds Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reading</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5&lt;sup&gt;th&lt;/sup&gt;</td>
<td>120 – 160</td>
<td>161 – 214</td>
<td>215 - 246</td>
<td>247 – 351</td>
</tr>
<tr>
<td>6&lt;sup&gt;th&lt;/sup&gt;</td>
<td>120 – 166</td>
<td>167 – 219</td>
<td>220 – 256</td>
<td>257 – 360</td>
</tr>
<tr>
<td><strong>Mathematics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5&lt;sup&gt;th&lt;/sup&gt;</td>
<td>120 – 179</td>
<td>180 – 213</td>
<td>214 – 270</td>
<td>271 – 369</td>
</tr>
<tr>
<td>6&lt;sup&gt;th&lt;/sup&gt;</td>
<td>120 – 193</td>
<td>194 – 224</td>
<td>225 – 275</td>
<td>276 – 379</td>
</tr>
</tbody>
</table>

*Curriculum Based Measurement*

ORF and MAZE are a type of CBM, a method of assessment that has garnered much attention in both the applied setting and the research setting. CBM was developed as an alternative to traditional IQ and achievement testing (Elliot & Fuchs, 1997) for use within a problem solving approach to remediate students’ educational issues primarily in the core academic areas of reading, mathematics, and writing (Wayman, Wallace, Wiley, Ticha & Espin, 2007). Within the realm of reading, the development of CBM was derived from the shift toward a developmental theory of reading.
Prior to the late 1960s the predominate approach to reading instruction was meaning based (Indrisano & Chall, 1995). The primary emphasis was to instruct child to read for meaning from the very start of their educational careers. Children were taught high frequency whole words in the very early elementary grades and word analysis skills were introduced later as their skills developed (Indrisano & Chall, 1995). Thus, there was only one approach to reading and the reading processes of competent adult readers were identical to those of developing readers. This approach was slowly abandoned for what is known as a coding emphasis on reading. This emphasis holds that beginning and mature readers utilize qualitatively different skills and processes for reading (Indrisano & Chall, 1995). Within this more developmental theory, the reader moves from basic decoding of words on one extreme of the spectrum to reading in order to obtain meaning from text (Indrisano & Chall, 1995).

Developmental theories of reading hold that students acquire requisite component skills that suffice the child’s reading demands, but in order for an individual’s level of reading proficiency to increase, so must their skills (Fuchs & Fuchs, 1993). Thus, students initially acquire phonemic awareness, the ability to hear and manipulate individual sounds. Phonemic awareness consists of its own developmental hierarchy of which a full discussion is outside the realm of this study. One note of importance is that phonemic awareness’s strong association with phonological memory (Joseph, 2006). Phonological memory is what allows individuals to hold patterns of phonemes within their working memory in order to produce whole words (Joseph, 2006). In order to move to the next developmental stage readers must be able to store individual sounds within
their working memory in order to process the codes into whole words (Joseph, 2006; Swanson, Zheng, & Jerman, 2009).

While phonological memory concerns the encoding of phonological combinations, rapid naming consists of the automatic retrieval of these codes in order to read (Joseph, 2006). The acquisition of rapid naming abilities helps readers move into the decoding stage of reading. In order to be able to decode words children must be able to selectively analyze the orthographic features of letters including shapes and common letter patterns visually (Joseph, 2006). The next vital feature of initial decoding involves an understanding of the alphabetic principle. This principle refers to the fact that readers must understand that there is a correspondence between letters or groups of letters and sounds (Joseph, 2006). Once readers understand the alphabetic principle they are able to link this with orthographic knowledge in order to decode words (Joseph, 2006). Through practice, readers are able to utilize their skills to automatically name words by sight and must no longer devote as many cognitive resources to decoding words.

Within the next stage of development children move from reading automaticity to reading fluency (Joseph, 2006). Fluency involves quickly and accurately reading text within a passage. Fluency also involves prosody or reading with expression (Joseph, 2006). Fluent readers abide by punctuation, alter their pitch, and make no or minimal mistakes. Once text can be decoded fluently they are more able to gain meaning from the presented text possibly due to a bridge effect that links working memory to comprehension (Swanson & Connor, 2009).

Fluent reading ability allows individuals to focus more on understanding the meaning of words (vocabulary) and comprehending the concepts present within the
reading passage (Joseph, 2006). Comprehension of written text allows readers to obtain new information to which they may not have been otherwise exposed. Furthermore, as students read they may activate long term memory in order to integrate newly presented information with knowledge they had attained previously. Finally, short term memory is also implicated in adequate reading comprehension. In order to make sense of text being read individuals must be able to retrieve encoded information in the correct sequence within and between sentences and words (Swanson, Zheng, & Jerman, 2009).

Curriculum-based Measurement and Benchmarking

CBM is often administered within a benchmarking process. Within the benchmarking process every student is administered the same measurement at the same time. All of these scores are then compiled in order to develop norms for each grade and class. Individual student scores are then compared to these norms in order to determine need for intervention or strengths. Within this process CBM utilizes numerous short and easily collected measurements of students’ academic skills providing general outcome measures (GOMs) of student growth toward broad basic skill acquisition (Deno, 1992). Thus, GOMs are more focused on a composite skill as a whole rather than the separate subcomponents of the skill. These “vital signs” or GOMs are collected through continuous measurement on equivalent forms of tasks across time (Deno, 1992). The data collected through this process are thought to represent student growth within the domain being measured and represent improvement on many skills subsumed by the broad domain being measured (Deno, Espin, & Fuchs, 2002). For example, if a student can fluently read text, then it can be assumed that he or she has mastered each previous reading process and the individual skills within each process.
CBM benchmarking is typically administered on a trimester basis (Fall, Winter, and Spring) to the entire student body. Students’ performances are then either compared against national norms or against some form of local norms derived by the school. If the student’s score falls within the bottom quarter of the national or local norms the student may then begin to receive remedial educational services (Shinn, Shinn, Hamilton, & Clarke, 2002).

A key characteristic of CBM that has made it popular for benchmarking is that it does not require the use of specific stimulus materials (Dena, 1992). Rather, CBM elucidates how to select, create, and administer the stimulus materials, as well as score the students’ performance. The emphasis on standardized measurement within CBM means that the stimulus materials being utilized are based on students’ curriculum (Deno, 1992) unlike the traditional approach to assessment using norm-referenced tests.

The two CBM procedures of primary focus in this study are ORF and MAZE. ORF is individually administered and is a relatively simple method of assessment. According to Deno (1985), students are presented with a reading passage that has either been developed from reading materials utilized locally in the classroom or is a generic passage. Students read the printed text out loud for one minute. The individual’s score is determined by first obtaining the total number of correctly read words by the student and then subtracting the total number of errors made by the student (correct words per minute). An error is made when a student omits, inserts, substitutes, or mispronounces a word (Deno, 1985). Hesitations (pauses of longer than three seconds) are also considered reading errors (Deno, 1985). Scores on ORF range from 0 to approximately 250. On the other hand, MAZE has been developed as an alternative to ORF. Students are
administered the MAZE task in a class-wide setting. They are charged with silently reading a passage for three minutes in which the first sentence is fully intact. Throughout the rest of the passage every 7th word is deleted and replaced with three different options (Wayman et al., 2007). One of the three words is both grammatically and logically correct while the other two are meant to distract the student (Wayman et al., 2007). Students are asked to select one word from the group of three that makes the most sense within the context of the presented sentence. The students’ total score is determined by counting the total number of items completed and the total number of incorrect answers. The number of incorrect answers is then subtracted from the total number of items completed. The difference is the students’ score on the MAZE task (number of correct words). Scores typically range from 0 to 50.

Currently, within the applied setting there is much controversy concerning which measure is most appropriate for benchmarking purposes at the upper elementary school grade levels. Some school districts have elected to administer only ORF, while some have elected to administer only MAZE, and still others have decided to administer both tools. MAZE owes its existence primarily due to two criticisms. Many individuals have claimed that ORF focuses solely on fluency at the expense of what they feel to be equally, if not more important, comprehension. Many teachers have complained about “word callers,” students who read fluently but do not understand the text, being misidentified by ORF measures (Meisinger, Bradley, Schwanenflugel, & Kuhn, 2010). MAZE is often thought to possess more face validity than ORF as it may represent a more authentic assessment encompassing not only fluency but also comprehension (Shinn, Shinn, Hamilton, & Clarke, 2002; Wayman et al., 2007). Second, MAZE can be
group administered (Wayman et al., 2007). This is of the utmost importance, as teachers are perpetually seeking ways to maximize and save instructional time throughout the educational day. Shinn et al. (2002) estimate that it takes approximately 5 minutes a student to administer an ORF assessment. If one makes a conservative estimate of 20 students to a classroom, it would take a little over an hour and a half to benchmark an entire class. In contrast, it takes 3 minutes to administer MAZE to an entire classroom. Even if it takes an individual half an hour to score the measure, the savings in time may be substantial and invaluable.

*The Relationship between Curriculum-based Measurement and Reading Performance on State Achievement Tests*

Recently, researchers have examined the ability of CBM to predict student performance on statewide standardized testing. This has been done in order to identify students who may require academic interventions so as to meet their state wide educational goals. The results have been promising.

ORF has been on the forefront of this research. Numerous researchers from across the United States have sought to examine the ability of ORF to predict student performance on state wide standardized tests (Wayman et al., 2007). Stage and Jacobsen (2001) conducted their analysis with 173 students from a school district in the Northwestern United States. Each student was administered a passage of 250 words and instructed to read the passage out-loud. Only the number of words read correctly was recorded. The students were also administered the Washington Assessment of Student Learning (WASL). The WASL is similar to the ISAT in that it is the outcome measure by which school effectiveness is calculated in Washington. Stage and Jacobsen observed
moderate correlations ($r = .43$ to $.44$) between ORF scores and WASL standard scores at each benchmarking period (September, January, and May). The researchers attempted to derive cut-scores for ORF in order to predict performance on the WASL through the application of growth curve analysis. Once these cut-scores were developed the researchers then evaluated the diagnostic efficiency of ORF in predicting performance on the WASL. According to Kessel and Zimmerman (1993), diagnostic efficiency is primarily determined through the use of several statistics: sensitivity, specificity, positive predictive power (PPP) and negative predictive power (NPP). These statistics are discussed later in greater detail. For the calculations of these statistics the reader is referred below.

Table 2.

*Diagnostic Efficiency Statistics Template*

<table>
<thead>
<tr>
<th>Cut-Score</th>
<th>Below Standards</th>
<th>Met Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below ORF Cut-Score</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>At or Above ORF Cut-Score</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

Note: Sensitivity $= \frac{a}{a+c}$, Specificity $= \frac{d}{b+d}$, Positive Predictive Power $= \frac{a}{a+b}$, and Negative Predictive Power $= \frac{d}{c+d}$

For the September benchmark, the researchers determined the sensitivity of the ORF cut-scores was $66\%$, the specificity was $76\%$. The other two benchmark period scores yielded results within $1\%$ of these figures. The researchers conclude that ORF can provide valuable information to educators attempting to determine who will pass state wide testing.
Hintze and Silbergliitt (2005) conducted a study with similar goals as Stage and Jacobsen (2001), exploring the ability of ORF to predict student performance on the reading section of the Minnesota Comprehensive Assessment (MCA) in third grade. The MCA is similar to both the ISAT and the WASL. These researchers observed stronger correlations between benchmark period scores and criterion scores ($r = .66$ to $.69$). The three third grade benchmarks possessed sensitivity ranging from $.63$ to $.65$ and a specificity of $.87$. The researchers concluded that ORF is an effective method for predicting student performance on statewide tests in Washington.

Despite its utility and research support, there have been few studies that have attempted to use ORF to predict student performance on the ISAT. Sibley, Biwer, and Hesch (2001) conducted a study that sought to establish a relationship between ORF and the reading section of the ISAT. The researchers observed a significant correlation between the fall ORF benchmark and student performance on the ISAT at the third grade ($r = .75$).

Herrera (2005) conducted an analysis to predict third and fifth grade student performance on the ISAT through the use of a regression formula. A moderate correlation between ORF and the reading section of the ISAT was found for both third ($r = .52$) and fifth grade ($r = .50$). The author then utilized a regression formula for both grades to attempt to predict student performance. The predictive ability of ORF was then evaluated through the use of the above explained diagnostic efficiency statistics. The author reports that for third grade, correct predictions were made 69% of the time, while false predictions were made 31% of the time. At the fifth grade level, correct predictions were made 68% of the time, while false predictions were made 32% of the time. The author
concluded that ORF can help to predict ISAT performance and that her results may indicate that there is an inverse relationship concerning the correlation between ORF and performance on state wide standardized tests as a function of student grade level.

Jenkins and Jewell (1993) had previously indicated the existence of the inverse relationship between ORF and standardized achievement tests as a function of grade. The researchers administered the Gates-MacGinite Reading Tests (GMRT), the Metropolitan Achievement Test (MAT), ORF measures, and MAZE measures to students in grades 2 through 6 in two schools in the Northwest United States to examine the relationship between informal reading tests (ORF and MAZE) and standardized achievement tests (GMRT and MAT). The researchers obtained results that showed that the correlations between ORF and standardized achievement tests weakened as a function of student grade level. Correlations between ORF and the GMRT were strong for grades two \( (r = .83) \), three \( (r = .88) \) and four \( (r = .86) \), but that the correlation between ORF and GMRT decreased in the sixth grade \( (r = .67) \). The correlations between ORF and MAT also showed similar declines with the correlation at second grade equaling .87, while the correlation in the sixth grade equaled .60. MAZE, on the other hand, showed correlations that varied less as a function of grade with correlations between the measure and both achievement tests ranging from the .60s to the .70s. No clear trend of increasing or decreasing occurred. The authors state that ORF performance appears to be less related to achievement tests at the upper elementary school levels and that this may be due to an increased demand for successful achievement test performance on language comprehension skills.
Silberglitt, Burns, Madyun, and Lail (2006) conducted a study that further supports previous (Herrea, 2005; Jenkins & Jewell, 1993) conclusions concerning the inverse relationship between ORF and performance on state wide standardized tests. Silberglitt et al. (2006) included over 5,000 students in grades three through eight in Minnesota. The authors sought to examine the relationship between the reading sections of the MCA and ORF to determine if the relationship between the two assessments varied as a function of grade. Students in the eighth grade were administered the Basic Skills Test for Reading (BST-R) instead of the MCA. The minimum performance benchmark of the BST-R is the level of proficiency readers must acquire in order to graduate from high school. The researchers found significant correlations across grade level between standardized tests and CBM methods. Interestingly, as grade level increased the strength of the relationship between ORF and statewide tests decreased. In third grade, the correlation was .68, while at the eighth grade the correlation was .50. The researchers concluded that while the relationship between ORF and state-wide tests is strong and well established at the early elementary school grades, it may be unwise for educators to extrapolate these findings to the upper elementary school grades given the weakening correlations that may occur as a function of grade.

Fuchs and Fuchs (1993) discovered similar findings while attempting to establish weekly growth goals for students utilizing both ORF and MAZE. The inverse relationships between ORF growth and grade was noted, but it was also determined that growth on MAZE was steady regardless of students’ grade level. The authors’ conclusions, as with those presented above, appear to logically follow the framework of the developmental theory of reading. Since MAZE is purported to more accurately
measure reading comprehension it may be a better measure of students who are approaching the apex of the developmental progression (upper elementary or middle school students). The authors conclude their article by stating that these findings seem to coincide with both the findings of Jenkins and Jewell (1993) and a factor analysis performed by Shinn, Good, Knutson, Tilly and Collins (1992).

Shinn et al (1992) conducted a factor analysis to examine the relationship between decoding and comprehension within the reading process with third and fifth grade students using ORF, nonsense word decoding, written retell, and cloze tasks. Nonsense word decoding involves presenting students with nonsense words and asking students to correctly decode them. Written retell requires students to read a passage and, within given time parameters, recount the story in writing using their own words. The cloze task required students to read a completely intact passage. At the end of a specified amount of time, the students were then given the same packet with every seventh word deleted. The students were charged with correctly writing in an exact match of the missing word. The researchers then conducted a confirmatory factor analysis to determine the best fit of the data to four differing theoretical models of reading.

For the third grade sample the authors determined a one factor model of reading to be the most parsimonious with all measures loading significantly on what the authors termed Reading Competence. For the fifth grade sample, the authors observed a two factor model of reading which they determined consisted of Reading Comprehension and Reading Decoding. The factor loadings on the Comprehension factor ranged from .61 for written retell to .86 for cloze. The loadings on Decoding ranged from .66 for nonsense words to .90 for ORF. These results may indicate that while ORF most directly measures
reading proficiency in the lower elementary grades, it may not be the most effective measure at the upper elementary school level as it may neglect to reflect students’ comprehension abilities. This possibility is one of the factors that may have lead to the adoption of MAZE as an alternative to ORF when assessing students.

Given this evidence, it may be more appropriate to utilize MAZE in the prediction of statewide reading performance than ORF at higher grade levels. Wiley and Deno (2005) conducted a study to compare the predictive ability of MAZE and ORF with both English language learners (ELLs) and non-English language learners (non-ELLs) in third and fifth grade. For the non-ELLs in the third and fifth grades, the researchers found correlations between MAZE and MCA of .73 and .71, respectively. The correlations between ORF and the MCA for the same group were .71 and .57. The researchers determined that MAZE appears to access student reading abilities that are not otherwise measured by ORF. Furthermore, based on the results from this study it appears that MAZE may be a superior tool in predicting upper elementary school students’ performance on statewide standardized testing in comparison to ORF.

Relationship Between Reading and Mathematical Abilities

Several researchers have extended the use of CBM reading measures to predict student mathematics performance on statewide tests. This application of reading CBM has been utilized due to the fact that students, especially as grade level increases, are often required to read in order to access content information across numerous subjects. This is true not only of subjects such as science, but also of mathematics. Mathematics tests and text books generally contain items that require numerical operations, as well as, word problems (Thurber, Shinn, and Smolkowski, 2002). Furthermore, the steps to
completing and understanding mathematical operations are presented through written language. Thus, reading abilities may be required for successful mathematics abilities. Although it is beyond the scope of this study to present an exhaustive discussion on the relationship between reading and math, several studies highlight the importance of considering reading abilities when evaluating mathematical ability.

Thurber, Shinn, and Smolkowski (2002) conducted a confirmatory factor analysis with the data derived from fourth grade students in order to determine whether math computation or general math achievement was measured by a math curriculum based measurement (M-CBM). The authors also examined the role of reading in mathematics achievement. Students were administered math fact probes (M-CBM), the Stanford Diagnostic Mathematics Test (SDMT), the Computation and Mathematics Concepts and Applications subtest of the California Achievement Tests (CAT), mathematical applications items from the National Assessment of Educational Progress (NAEP), and MAZE. The authors tested three models of math assessment and a two factor model was supported by their results with Computation and Applications emerging as the two distinct constructs. In all three models, reading, as measured by MAZE, was moderately correlated with the mathematics constructs. The correlations between Reading and Computation ranged from .69 to .79, while the correlations between Reading and Applications ranged from .76 to .77. The authors state that their results may support the conclusion that reading is necessary for overall mathematics proficiency.

Helwig, Rozek-Tedesco, Tindal, and Heath (1999) sought to examine if mathematics tests that utilize multiple-choice items measure solely mathematics proficiency. The authors conducted their study with 325 sixth grade students. These
students were administered an ORF probe, a 21 question basic math skills test which contained 19 computation problems and two simple word problems, and 60 problems form the Standardized Mathematics Achievement Test (SMAT) which contained word problems with no simple computation items. Half of the study’s word problems were administered through paper and pencil means, while the other half were administered through a video version of the test. The video version was administered in order to reduce the reading demands of the word problems. The scores in the video condition were higher than those in the standard condition, but this result was not statistically significant. The authors also indicated that as the complexity of the test items increased, the scores in the video condition tended to be better than the scores on the standard condition items. The authors further concluded that reading is a critical access skill to general mathematics competency.

Helwig et al (1999), provide a possible explanation of cognitive processing that may underlie the relationship between reading and mathematics performance on tests involving word problems. Students with established reading fluency are able to more automatically decode words. This increased processing allows for more cognitive resources to be devoted to higher order operations such as comprehension and synthesis of what is being read. Thus, individuals with weaker word decoding abilities may devote too much effort in determining what a word is that they are less likely to comprehend what is being read due to an inability to identify a word or inaccurate reading (Helwig et al, 1999). Restrictions of working memory are also implicated (Helwig et al, 1999). As reading input is brought into working memory, it is connected to new information input in order to integrate and synthesize new information with older information, but for poor
readers, as new information is brought in, old information may decay and be lost due to a significant time delay resulting from laborious text decoding. Poor decoding may also retard the ability of the reader to adequately control his or her processing in order to attend to pertinent information while simultaneously ignoring irrelevant or erroneous information (Swanson, Zheng, & Jerman, 2009). Short term memory may also be negatively affected. Some students’ fluency may be so weak that as they decode words towards the end of the problem, they may have forgotten vital information from the beginning of the text that is required for successful completion of the item (Helwig et al., 1999). If text cannot be decoded it cannot be encoded into memory, therefore, it cannot be retrieved or built upon when individuals are asked to complete a mathematical task or learn a new mathematical skill.

The role of memory retrieval in the relationship between reading and mathematical ability has been further supported. Koponen, Aunola, Ahonen, & Nurmi (2007) state that, like reading, mathematical calculation is a multi-component developmental skill. In order to solve multi-digit written problems (procedural calculation) children must first be able to automatically and accurately retrieve single-digit calculations as they are often intermediary steps (Koponen et al., 2007). Furthermore, mathematical calculation fluency and reading fluency have shown moderately strong correlations (Koponen et al., 2007).

This relationship may be related to phonological skills. Central to attaining adequate calculation, fluency is the ability to retrieve verbal information from long-term memory (Koponen et al., 2007). Poor phonological skills may severely hinder individuals’ abilities to encode this information. This concept has been supported by
research indicating that phonological awareness, rate of retrieval of phonological codes from long term memory, and phonological memory possibly explain nearly all of the covariance between calculation and reading (Koponen et al., 2007).

Curriculum-based Measurement and Mathematics Performance on State Achievement Tests.

In line with the above evidence, Crawford, Tindal and Stieber (2001) conducted a study in the Northwestern United States to not only predict third grade student reading performance on a statewide achievement test through the use of ORF, but to additionally predict student performance on the test’s mathematics portion also through the use of ORF. The rationale behind this approach was that very often mathematics subtests consist of numerous multiple-choice questions. Therefore, students must possess a minimum level of reading ability in order to understand and answer the items (Crawford et al., 2001). The authors report a moderate correlation of .64 between student’s math and reading performance on the statewide assessment. Moderate correlations between CBM score and both reading ($r = .60$) and mathematics ($r = .46$) performance were also observed. The researchers conclude that a certain level of reading proficiency is necessary to accurately complete mathematics sections of standardized tests that utilize word based multiple choice questions.

Once again, MAZE has emerged as an alternative to ORF in predicting student performance. Jiban and Deno (2007), sought to utilize CBM in reading to predict student performance on a statewide mathematics assessment in Minnesota. Based in part on the conclusions drawn by Crawford et al. (2001), the authors state that reading proficiency is a necessary component of adequate performance on the mathematics subtests of many
standardized statewide assessments. The authors sought to determine if MAZE explains a unique amount of variance concerning third and fifth grade student performance on the MCA’s mathematics subtest and to observe if MAZE or two different math CBM methods would better predict MCA math performance. For the third grade, the authors observed a moderate correlation of .44 between MAZE and the MCA’s mathematics section. This correlation was stronger than all but one of the math CBM methods, which also correlated at .44 with the MCA. For the fifth grade, MAZE correlated with the MCA at .63, greater than all the math CBM methods. The authors conclude that at fifth grade MAZE contributed uniquely to predicting student performance. Finally, the authors state that schools may use reading CBM scores to predict student performance on mathematics tests that depend on the use of reading skills, such as the ISAT.

Correlations between ISAT reading and mathematics scale scores were determined by both the Illinois State Board of Education (2008) and by Wick (2002). According to the ISBE, for both fifth and sixth grade the correlation between the reading and mathematics sections was .76 (Illinois State Board of Education Division of Assessment, 2008). Wick (2002) reports similar findings for fifth grade, with the correlation being .79. This high degree of intercorrelation may indicate that despite the distinctness of the two scale names, reading and mathematics, the latter may be not only measuring students’ mathematics reasoning, but, in large part, also their reading abilities. This has important implications when one attempts to predict student performance on the ISAT through the use of a screening measure such as curriculum based measurement (CBM).
In summary, the federal legislation, NCLB, has made it clear that all students must reach proficiency in reading and math, and that schools will be held responsible for failing to do so. As a result, each state administers an outcome based achievement test each academic year to assess students’ academic proficiency. In order to achieve higher proficiency, schools identify benchmarks for monitoring student progress using ORF and MAZE in reading and math. Reading and math are two critical skills students need for higher education as well as living skills. Further, an applied and a theoretical relationship between reading and math has been established. Thus, this study attempted to identify whether or not ORF and MAZE predict math and reading level of proficiency of 5th and 6th grade students on the ISAT.

**Purposes and Hypotheses**

The first purpose of the current study is to extend the use of two reading CBM tools in the prediction of upper elementary students’ performance on the reading section of the ISAT. The two measures that were utilized as predictors were ORF and MAZE scores. ORF was used because it has an extensive research base supporting its use in such procedures (Herrera, 2005; Hintze & Silberglitt, 2005; Stage & Jacobsen, 2001). MAZE was utilized as an alternative to ORF for several reasons. First, there is an established inverse correlation between ORF and statewide assessment performance (Fuchs & Fuchs, 1993; Herrera, 2005; Jenkins & Jewell, 1993; Silgerglitt, Burns, Madyun, & Lail, 2006). Second, the confirmatory factor analysis performed by Shinn et al. (1992) supports that comprehension (which MAZE is purported to measure) is a separate factor that explains student reading abilities at the upper elementary school levels. This conclusion is further supported by results obtained by Wiley and Deno (2005). Therefore, it is hypothesized
that MAZE is a better predictor of student performance on the reading section of ISAT than ORF.

The second purpose of this study is to establish the ability of these two CBM methods to predict fifth and sixth grade student performance on the mathematics sections of the ISAT. Once again, the predictor measures were ORF and MAZE and the criterion measure was the students’ performance on the ISAT’s mathematics section. The establishment of this relationship is sought for several reasons. First, it has been established that reading may be an important access skill to mathematics proficiency (Helwig et al., 1999; Thurber et al., 2002) and reading has been shown to correlate with mathematics performance on standardized achievement tests (Thurber et al., 2002). Furthermore, researchers have established a relationship between math performances on statewide multiple choice mathematics assessments that require extensive use of reading skills and ORF (Crawford, Tindal, & Stieber, 2001), as well as MAZE (Jiban & Deno, 2007). Given the extensive use of multiple choice questions on the ISAT (see ISAT mathematics subsection description above) and the interrelatedness of the reading and mathematics scale scores (Wick, 2002), this seems to be a reasonable pursuit. It is hypothesized that both CBM methods would predict student performance on the math section of the ISAT. Furthermore, since MAZE is hypothesized to be a better predictor of student reading abilities at the fifth and sixth grade level, it is predicted that MAZE is a superior method of calculating math performance than ORF.

This study is important for several reasons. First, it may further establish the relationship between CBM and statewide standardized achievement tests, specifically within the state of Illinois where there is a dearth of research. Second, comparing the
ability of ORF and MAZE to predict student reading achievement may help determine which CBM method is most appropriate and efficient for reading performance benchmarking at the upper elementary school levels. Finally, this research may provide insight into the utility of using reading CBMs to predict student performance on statewide standardized mathematics tests that demands prerequisite levels of reading skills.

Method

Participants

Participants of the study consisted of 197 students in the fifth (n = 102) and sixth (n = 92) grades from a rural elementary school and two urban schools in Illinois. The school-wide student population of the rural elementary school was 95% Caucasian and the school-wide teacher population was 100% Caucasian. The school-wide student population from one of the urban schools was composed of 68% Caucasians, while the teacher population was also predominately Caucasian, 90%. The second urban school was composed of a student population that is 44% Caucasian and 48% African American. The teacher population was 91% Caucasian.

Materials

The CBM materials utilized were the Oral Reading Fluency subsection of the Dynamic Indicators of Basic Early Literacy Skills, 6th edition or DIBELS (Good & Kaminski, 2002), and MAZE selection task from the AIMSweb assessment system (Shinn & Shinn, 2002). The CBM data were collected through universal benchmarking conducted during September of the 2008-2009 school year. The fall benchmark data were utilized because they provide educators with the greatest amount of time to implement any required interventions. The high stakes mathematics and reading data were collected
during the Spring administration of the ISAT during the same school year. A discussion of the technical properties of each measure follows.

**Oral Reading Fluency**

ORF has been shown to possess technical adequacy and a strong relationship to overall reading proficiency (Wayman et al., 2007). Numerous researchers have sought to establish the psychometrics of ORF. It has shown to possess adequate levels of reliability. Tindal, Marston, and Deno (1983) evaluated ORF’s reliability through three different indexes including test-retest reliability, alternate forms reliability, and inter-judge agreement. All coefficients were shown to be adequate as the test-retest coefficient was .92, the alternate forms coefficient was .89, and the inter-judge agreement was .99.

There is a plethora of research concerning the validity of ORF. ORF scores from across grades and types of students have been analyzed in relation to student scores on numerous criterion measures. Fuchs, Fuchs, & Maxwell (1988) examined the relationship between ORF and the Reading Comprehension and Word Skills subtests from the Stanford Achievement Test. These researchers indicate that ORF may possess adequate criterion validity. The number of words read correctly correlated .91 with the Reading Comprehension subtest scores and .80 with the Word Skills subtest scores. Markell and Deno (1997) obtained scores from forty-two general education third graders. These researchers found that as passage difficulty increased, student performance decreased. For further reliability and validity information, the reader is referred to Shinn and Shinn (2002) and Wayman et al (2007).
Much like ORF, MAZE has been shown to be a technically adequate tool. Shin, Deno, and Espin (2000) conducted analyses in order to assess the technical adequacy of MAZE with forty-three second graders. The researchers computed Pearson product correlations in order to determine alternate forms reliability for MAZE. The results indicated that MAZE possesses adequate alternate forms reliability. The correlations between forms administered between one and nine months ranged from .69 to .91 with the mean equaling .81. Coefficients for forms administered within one month ranged from .75 to .90 with a slightly higher mean of .83. The correlation range for the two-month interval was .75 to .87 while the range for the three-month interval was .69 to .91, with .80 being the mean for both interval conditions.

Numerous studies have been conducted to examine the validity of MAZE. Espin, Deno, Maruyama, and Cohen (1989) and Jenkins and Jewell (1993) examined the criterion validity of MAZE by examining its correlation with ORF. Both researcher teams observed strong correlations between correctly read words and correctly selected words. Espin et al (1989) obtained a coefficient between correctly read words and correctly selected words of .86, while Jenkins and Jewell (1993) observed coefficients ranging between .80 and .89. For further information concerning the validity of MAZE readers are referred to Fuchs and Fuchs (1992), Parker, Hasbrouck, & Tindal (1992), and Wayman et al (2007). In summation, both ORF’s correct words per minute and MAZE’s number of correct words have shown to be technically adequate tools.
In order to determine if schools are meeting the prescribed percentage of children displaying grade level skills in reading and mathematics, students are administered the Illinois Standards Achievement Test (ISAT) during the spring of each school year. The ISAT measures the extent to which students are meeting the Illinois Learning Standards (Illinois State Board of Education Division of Assessment, 2008).

During the initial testing session, both the fifth and sixth grade students are administered the ISAT in three 45 minute sessions with 30 norm-referenced multiple choice items adapted from the reading section of the Stanford Achievement Test 10 (SAT 10) (Palmer, 2009). These students are administered an additional 21 questions in the following two sessions, with one of those questions being an extended response question. The extended response items are scored using a scale ranging from zero to four (Illinois State Board of Education Division of Assessment, 2008). A score of zero indicates that there is not enough information to assess the student’s achievement, scores of one or two indicate the students level of proficiency is developing, a score of three indicates the student’s level of proficiency is adequately developed, and a score of four indicates the student’s level of proficiency is well-developed (Illinois State Board of Education Division of Assessment, 2008). The second and third sessions’ questions are meant to specifically measure students’ performance within the Illinois Reading Assessment Framework with one of the two sessions being a pilot (Palmer, 2009). The combination of the first session’s multiple choice questions and the multiple choice questions based on short passages, and the extended response item from the non-pilot session are used to derive each student’s scale score and ultimately her or his performance level within the reading section of the ISAT (Illinois State Board of Education, 2008). The items within the ISAT reading section are meant to evaluate students’ ability to
read, understand, and analyze text (Illinois State Board of Education Division of Assessment, 2008); in other words, it appears to measure global reading abilities.

Concerning the mathematics sections, both grades are administered questions in three 45 minute sessions (Illinois State Board of Education, 2009). Within the first session students are administered 40 questions with the first 30 being derived from an abbreviated version of the Mathematics Problem Solving section of the Stanford Achievement Test 10 (Illinois State Board of Education, 2009). The second session contains 30 multiple choice items with 5 being pilot items and three short response items with one being a pilot item. The third session contains two extended response items with one being a pilot (Illinois State Board of Education, 2009). The short-response questions require students to solve mathematical problems without being presented with or given the correct answer (Illinois State Board of Education Division of Assessment, 2008). These responses are scored using a scoring rubric that ranges from zero to two. A score of zero indicates that the student provided no answer or an incorrect response, a score of one represents a partially correct response, and a score of two signifies a completely correct response (Illinois State Board of Education, 2009). The extended response items require that students develop a plan of action to execute the problem, carry out the plan, and reach a conclusion in relation to the initial problem (Illinois State Board of Education Division of Assessment, 2008). The scoring utilized for grading students’ extended responses falls along a zero to four scale (zero being the worst and four being the best). The combination of the first session’s 40 multiple choice questions, session two’s 25 multiple choice questions and two short response items, and session three’s extended response item determine each student’s scale score and performance level (Illinois State
Board of Education, 2008). These items are meant to assess student knowledge within three dimensions of the Illinois Learning Standards: Mathematical Knowledge, Strategic Knowledge, and Explanation (Illinois State Board of Education Division of Assessment, 2008). It appears probable that the items on the mathematics section of the ISAT tap not only students’ calculation skills, but also, given the timed multiple choice structure of many of the items, students’ reading abilities.

The ISAT’s 2008 Technical Manual presents information concerning the reliability and validity of scores derived from the use of the ISAT. Internal consistencies using the coefficient alpha for both sections were shown to be adequate (Illinois State Board of Education Division of Assessment, 2008). For the fifth and sixth grade reading and mathematics assessments the reliability estimates ranged .89 to .93. Inter-rater agreement for the extended reading responses as determined through the combination of the percentage of exact agreements plus the percentage of adjacent agreements for fifth and sixth grade mathematics and reading items ranged from 86% to 100% (Illinois State Board of Education Division of Assessment, 2008). Item reliabilities were independently evaluated by Wick (2002) utilizing a point-biserial correlation coefficient. It was concluded that the items included within the ISAT should be considered reliable as they all attained the .15 criterion (Wick, 2002).

The ISAT was evaluated for concurrent validity with the SAT 10. The fifth grade reading and mathematics sections of the ISAT, excluding those questions adapted from the SAT 10, showed strong correlations (Illinois State Board of Education Division of Assessment, 2008). The correlations for sixth grade reading and mathematics were also strong. It was concluded that the ISAT possesses adequate validity. The correlation
between the two measures for fifth grade reading was .78. For the fifth grade math section the correlation was .85. The correlation for sixth grade reading was .77 and the correlation for sixth grade mathematics was .86 (Illinois State Board of Education Division of Assessment, 2008).

For both the reading and mathematics sections scale scores range from 120 to 400 and above. For each grade and educational dimension, students’ cut scores fall within four performance categories; Academic Warning (W), Below Standards (B), Meets Standards (M), and Exceeds Standards (E). Curriculum experts from Illinois were recruited to determine cut scores for the ISAT tests (Illinois State Board of Education Division of Assessment, 2008). These experts were selected due to their knowledge of student performance at each grade level being assessed, as well as their experience evaluating students at those grade levels. The final panel consisted of 170 educators (Illinois State Board of Education Division of Assessment, 2008). The panel utilized an established method of determining cut off scores for multiple choice tests based on professional judgment known as Agnoff’s cut score procedure (Illinois State Board of Education Division of Assessment, 2008).

For the purposes of this study two reading performance levels were used, Below Standards (BSR) and Meets Standards (MSR). Scores below 214 for fifth graders and 219 for sixth graders (both inclusive) fall within the BSR category and scores of 215 and above for fifth graders and 220 and above for sixth graders are considered to fall within the MSR category. Two mathematics performance levels were also used. Scores from 213 and below for fifth graders and 224 and below for sixth graders will be considered to be Below Standards (BSM). Scores of 214 and above for fifth graders and scores of 225
or above for sixth graders will be considered to Meet Standards (MSM). These scores are summarized in Table 3.

Table 3.

**ISAT Minimum Reading and Math Scale Scores Required to Meet Standards**

<table>
<thead>
<tr>
<th>Grade</th>
<th>Meets Standards Reading</th>
<th>Meets Standards Mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td>5th</td>
<td>&gt;214</td>
<td>&gt;213</td>
</tr>
<tr>
<td>6th</td>
<td>&gt;219</td>
<td>&gt;224</td>
</tr>
</tbody>
</table>

**Diagnostic Utility**

Due to the relative obscurity of these statistics to the lay public and their relative importance to this study, this section was devoted to facilitate the reader’s understanding. Sensitivity provides a measure of the proportion of students passing the ISAT as identified by meeting the cut score of the predictor. In other words it answers the question, among those judged as passing, how many were true success? Sensitivity is also referred to as the true positive rate (Watkins, 2010). Typically, measures with a high degree of sensitivity also possess a low rate of Type II error or false negatives. It is determined by dividing the number of true positives by the number of true positives plus false positives. Specificity refers to the proportion of students not meeting standards correctly identified as not meeting standards by the predictor CBM, or the true negative rate (Watkins, 2010). It answers the question, among those judged as failing, how many were true failures? In contrast to tests with high degrees of sensitivity, those with high rates of specificity typically also display low rates of Type I error. It is calculated by
dividing the number of true negatives by the number of true negatives plus the number of false positives. Two other measures of diagnostic utility advocated some authors are positive predictive power (PPP) and negative predictive power (NPP) (Landau, Milich, & Widiger, 1991). Positive predictive power is the probability that the student will pass the ISAT given a score that meets the cut point, or among those who actually passed, how many were identified as passing? Negative predictive power refers to the probability that a student will not pass the ISAT when the student’s benchmark performance indicates that he or she will not. The question answered by NPP is, among those who actually passed, how many were identified as failing?

**Procedure**

Once the Institution Review Board approved the study, the primary researcher collected each student’s ORF, MAZE, and ISAT scores after trained school personnel obtained them following the procedures described above. ORF and MAZE scores were collected in September as part of the school wide benchmarking. Individual parental consent for the utilization of these data was not required, as these data were collected as part of the school’s universal benchmarking in which every student participated.

**Design and Data Analysis**

Means and standard deviations were computed for all measures at both grade levels. Pearson-product moment correlations were also computed between all measures. This was done to determine the relationship between the two CBM methods and the two sections of the ISAT. Regression analyses were conducted utilizing one dependent variable where both predictors were entered. This was done to obtain standardized beta values in order to determine which CBM is the superior predictor. Individual regression
formulas for each CBM were then derived from a simple linear regression analysis. These formulas were utilized to compute cut-off scores in order to predict student performance on the ISAT through the use of the two CBM measures. ORF and MAZE scores to obtain the minimum MSR and MSM for both grades were entered into the regression formulas until the minimum score for each CBM benchmark was determined. The results obtained from the regression formulas were then entered into a Microsoft Excel diagnostic utility program (Canivez & Watkins, 1996) to determine diagnostic efficiency statistics such as those explicated by Kessel and Zimmerman (1993) between each CBM and each ISAT scale. Final determinations as to overall effectiveness of the CBMs to predict student performance were determined based on each CBM's diagnostic utility.

**Results**

*Means and Standard Deviations*

The fifth grade sample showed a mean ORF score of 122.42. This indicated that this specific sample of students was displaying adequate fluency performance, as a commonly held heuristic states that an ORF at or above 100 typically indicates mastery of that material. This was also true concerning the sixth grade sample as they showed a mean ORF performance of 126.46. The fifth grade mean performance of 17.03 and the sixth grade mean performance of 19.38 on MAZE also indicated that the students possessed adequate reading comprehension abilities as these scores are within the expected benchmark performance range for that time of the school.

In regards to the ISAT reading subsection, it appeared that most of the students within the samples of this study performed adequately. Within both grade levels the mean performance was higher than the minimum score needed to meet standards on the
subsection. The fifth graders required a minimum score of 215 to meet standards while the mean performance was 234.61. The sixth grade sample required a minimum score of 220 to meet standards, but the mean score was 231.99. Within fifth grade math, the minimum score to meet standards was 214, but the mean performance of this sample surpassed this goal as the average score was 237.80. The sixth grade minimum score needed was 225, while the mean performance of this sample was 238.41. This information is summarized in Tables 4 and 5.

Table 4.

*Descriptive Statistics for CBM Scores and ISAT Subscales for Fifth Grade (N = 102)*

<table>
<thead>
<tr>
<th>Measure</th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF</td>
<td>102</td>
<td>122.42</td>
<td>35.92</td>
</tr>
<tr>
<td>MAZE</td>
<td>102</td>
<td>17.03</td>
<td>7.10</td>
</tr>
<tr>
<td>ISAT-R</td>
<td>102</td>
<td>234.61</td>
<td>21.92</td>
</tr>
<tr>
<td>ISAT-M</td>
<td>102</td>
<td>237.80</td>
<td>26.83</td>
</tr>
</tbody>
</table>

Note. ORF = Oral Reading Fluency. ISAT-R = ISAT Reading. ISAT-M = ISAT Mathematics. Test.

Table 5.

*Descriptive Statistics for CBM Scores and ISAT Subscales for Sixth Grade (N = 95)*

<table>
<thead>
<tr>
<th>Measure</th>
<th>N</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF</td>
<td>95</td>
<td>126.46</td>
<td>37.90</td>
</tr>
<tr>
<td>MAZE</td>
<td>95</td>
<td>19.38</td>
<td>9.61</td>
</tr>
<tr>
<td>ISAT-R</td>
<td>95</td>
<td>231.99</td>
<td>26.98</td>
</tr>
<tr>
<td>ISAT-M</td>
<td>95</td>
<td>238.41</td>
<td>27.74</td>
</tr>
</tbody>
</table>

Note. ORF = Oral Reading Fluency. ISAT-R = ISAT Reading. ISAT-M = ISAT Mathematics.
CBM and ISAT subsection correlations for fifth grade

A Pearson’s $r$ was conducted on all measures within the study for both grade levels. Concerning the relationship between the two CBM measures for fifth grade, a significant and positive relationship was obtained $r(100) = .72$, $p < .01$ (one-tailed). This was also the case relative to the relationship between the two subsections of ISAT for fifth grade, $r(100) = .74$, $p < .01$ (one-tailed).

As stated above, a Pearson’s $r$ was computed between MAZE and ISAT reading, as well as ORF and ISAT reading. While utilizing an alpha level of .05, all correlations between each measure were significant, $p < .01$, (one-tailed). The correlation between MAZE and the ISAT reading subsection was moderately strong, $r(100) = .67$. ORF displayed an identical correlation with ISAT reading.

A Pearson’s $r$ was also conducted on each CBM measure and ISAT mathematics. At an alpha level of .05, all correlations were once again significant, $p < .01$, (one-tailed). ORF showed a moderately strong positive correlation with ISAT mathematics, $r(100) = .55$. MAZE also showed a somewhat attenuated, although still moderately strong, relationship with ISAT mathematics, $r(100) = .53$. The difference in correlations was not significant ($t(99) = .33$, $p = .75$). Concerning both ISAT subsections, these results suggest that as scores on each CBM increase so do scores on each ISAT subsection. The correlations are summarized below.
Table 6.

Correlation Matrix between CBM Scores and ISAT Subscales for Fifth Grade ($N = 102$)

<table>
<thead>
<tr>
<th></th>
<th>ORF</th>
<th>MAZE</th>
<th>ISAT-R</th>
<th>ISAT-M</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF</td>
<td></td>
<td></td>
<td>.67**</td>
<td>.55**</td>
</tr>
<tr>
<td>MAZE</td>
<td>.72**</td>
<td></td>
<td>.67**</td>
<td>.53**</td>
</tr>
<tr>
<td>ISAT-R</td>
<td>.67**</td>
<td>.67**</td>
<td></td>
<td>.74**</td>
</tr>
<tr>
<td>ISAT-M</td>
<td>.55**</td>
<td>.53**</td>
<td>.74**</td>
<td></td>
</tr>
</tbody>
</table>

Note. ORF = Oral Reading Fluency. ISAT-R = ISAT Reading. ISAT-M = ISAT Mathematics. ** = Correlation is significant at 0.01 (1-tailed).

CBM and ISAT Subsection Correlations for Sixth Grade

Pearson’s $r$ analyses were also performed on each CBM measure and each ISAT subsection at the sixth grade. Once again, ORF and MAZE showed a significant relationship, $r(93) = .65, p < .01$ (one-tailed). The results of the correlation between the two ISAT subsections also revealed a significant relationship, $r(93) = .78, p < .01$ (one-tailed).

Utilizing an alpha level of .05, significant ($p < .01$, one-tailed) and positive correlations were obtained across each CBM and each ISAT subsection. Within the realm of reading, ORF showed a higher degree of correlation with the ISAT, $r(93) = .64$ than did MAZE, $r(93) = .62$, but this difference was not significant ($t(92) = .31, p = .75$). In relation to mathematics, this did not occur. MAZE showed a higher degree of correlation with the ISAT, $r(93) = .53$, than did ORF, $r(93) = .49$, but once again this difference was not significant ($t(92) = .55, p = .58$). These results mirror those of the fifth grade sample.
in that the results indicate that as CBM scores increase so do ISAT subsection scores. The results of the sixth grade correlations are summarized in Table 7 below.

Table 7.

*Correlation Matrix between CBM Scores and ISAT Subscales for Sixth Grade (N = 95)*

<table>
<thead>
<tr>
<th></th>
<th>ORF</th>
<th>MAZE</th>
<th>ISAT-R</th>
<th>ISAT-M</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF</td>
<td>.65**</td>
<td></td>
<td>.64**</td>
<td>.49**</td>
</tr>
<tr>
<td>MAZE</td>
<td></td>
<td>.65**</td>
<td>.62**</td>
<td>.53**</td>
</tr>
<tr>
<td>ISAT-R</td>
<td>.64**</td>
<td>.62**</td>
<td></td>
<td>.78**</td>
</tr>
<tr>
<td>ISAT-M</td>
<td>.49**</td>
<td>.53**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. ORF = Oral Reading Fluency. ISAT-R = ISAT Reading. ISAT-M = ISAT Mathematics.
** = Correlation is significant at 0.01 (1-tailed).

**Simple Regression Analyses**

Simple regression analyses were conducted at each grade level to examine how the previously discussed benchmarking tools measure predicted performance on each subsection of the ISAT. In order to determine if there is a significant difference between the amount of variance in each ISAT subsection accounted for by the respective CBM, the difference between the squared standardized beta for each CBM must equal or surpass 5%. For fifth grade reading performance, results show that ORF accounted for 15% of the variance in ISAT performance, *t*(100) = 3.91, *p* = .00, while MAZE accounted for 14% of the variance in ISAT performance, *t*(100) = 3.83, *p* = .00. The difference between the two CBMs was not significant. The results for fifth grade math showed that ORF accounted for 12% of the variance in ISAT performance, *t*(100) = 3.00,
CBM and ISAT 46

$p = .00$. MAZE accounted for slightly less variance, 8%, $t(100) = 2.34, p = .02$, than did ORF. Once again the difference was not significant. These results are summarized below.

Table 8.

*Summary of Simple Linear Regression Analysis for 5th Grade CBM Predicting ISAT Reading Performance (N = 102)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>$B$</th>
<th>$SE B$</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF</td>
<td>0.24</td>
<td>0.06</td>
<td>0.39</td>
<td>3.91</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>MAZE</td>
<td>1.18</td>
<td>0.31</td>
<td>0.38</td>
<td>3.83</td>
<td>&lt;.01</td>
</tr>
</tbody>
</table>

$R^2 = .52, p = .00$

Table 9.

*Summary of Simple Linear Regression Analysis for 5th Grade CBM Predicting ISAT Mathematics Performance (N = 102)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>$B$</th>
<th>$SE B$</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF</td>
<td>0.26</td>
<td>0.09</td>
<td>0.35</td>
<td>3.00</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>MAZE</td>
<td>1.04</td>
<td>0.44</td>
<td>0.28</td>
<td>2.34</td>
<td>.02</td>
</tr>
</tbody>
</table>

$R^2 = .34, p = .00$

At the sixth grade level, ORF accounted for more variance, 17%, $t (93) = 4.11, p = .00$, than did MAZE, 13%, $t (93) = 3.60, p = .00$, within ISAT reading, but this difference was not significant. The opposite was true in regard to sixth grade ISAT mathematics. MAZE accounted for 13% of the variance in ISAT scores, $t (93) = 3.18, p = .00$, whereas ORF accounted for 7% of the variance in ISAT mathematics performance, $t$.
(93) = 2.26, \( p = .03 \). The above difference was significant. These results are summarized in Tables 10 and 11.

Table 10.  

*Summary of Simple Linear Regression Analysis for 6th Grade CBM Predicting ISAT Reading Performance (N = 95)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>( B )</th>
<th>( SE )</th>
<th>( \beta )</th>
<th>( t )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF</td>
<td>0.29</td>
<td>0.07</td>
<td>0.41</td>
<td>4.11</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>MAZE</td>
<td>1.00</td>
<td>0.28</td>
<td>0.36</td>
<td>3.60</td>
<td>&lt; .01</td>
</tr>
</tbody>
</table>

\( R^2 = .48, p = .00 \)

Table 11.  

*Summary of Simple Linear Regression Analysis for 6th Grade CBM Predicting ISAT Mathematics Performance (N = 95)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>( B )</th>
<th>( SE )</th>
<th>( \beta )</th>
<th>( t )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF</td>
<td>0.19</td>
<td>0.07</td>
<td>0.26</td>
<td>2.26</td>
<td>0.03</td>
</tr>
<tr>
<td>MAZE</td>
<td>1.04</td>
<td>0.25</td>
<td>0.36</td>
<td>3.20</td>
<td>&lt; .01</td>
</tr>
</tbody>
</table>

\( R^2 = .32, p = .00 \)

Diagnostic Utility

After the simple regression analyses were conducted, a regression formula between each CBM measure and each ISAT subsection at each grade level was constructed utilizing the unstandardized beta coefficients. ORF and MAZE scores were then entered into each formula until the minimum score required to meet standards on each subsection of the ISAT was determined. This minimum score was then plotted on a
scatter plot of each CBM and each ISAT subsection. The minimum CBM scores required to meet standards on each ISAT subsections at each grade level are summarized below (Tables 12 and 13).

Table 12.

5th Grade Cut Scores

<table>
<thead>
<tr>
<th>CBM Measure</th>
<th>ISAT Reading</th>
<th>ISAT Mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF</td>
<td>125</td>
<td>105</td>
</tr>
<tr>
<td>MAZE</td>
<td>25</td>
<td>26</td>
</tr>
</tbody>
</table>

Table 13.

6th Grade Cut Scores

<table>
<thead>
<tr>
<th>CBM Measure</th>
<th>ISAT Reading</th>
<th>ISAT Mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF</td>
<td>152</td>
<td>161</td>
</tr>
<tr>
<td>MAZE</td>
<td>45</td>
<td>29</td>
</tr>
</tbody>
</table>

The minimum score to meet standards on each subsection at each grade level was also plotted on these graphs. These graphs were then visually inspected in order to tally the total number of true positive, false positive, true negative, and false negative identifications made. Those data points that fell within the upper left hand quadrant created by the minimum score lines for both CBM measures and ISAT cut-offs were determined to be false negatives, those students expected to fail the ISAT based on their CBM performance, but who did not. The data points that fell within the lower left hand quadrant were deemed true negatives; those students who were expected to fail the ISAT
based on CBM performance and did in fact fail to meet standards. If a data point occurred within the upper right hand quadrant of a graph, it was termed a true positive; a student predicted to meet standards based on CBM performance and who did in fact succeed in meeting standards. Finally, students who fell within the lower right hand quadrant were false positives. These were students who were predicted to meet standards on the ISAT subsection, but who did not. An illustration of this process is provided below, Figure 1.

**ORF and ISAT Mathematics**

![Scatterplot Utilized to Determine Diagnostic Utility Statistics](image)

*Figure 1. Scatterplot Utilized to Determine Diagnostic Utility Statistics*

This analysis was performed in order to calculate the diagnostic utility of each CBM when making decisions concerning the need for intervention services in order to meet standards on the ISAT subsections. Once the number of students falling into each category was counted, these figures were then entered into an electronic diagnostic utility program to calculate the specificity, sensitivity, positive predictive power (PPP) and
negative predictive power (NPP) of each CBM for each subsection at both grade levels. A generally agreed upon criterion for acceptable levels of diagnostic utility is 90%.

In regards to fifth grade reading, ORF showed PPP value of 96%, while MAZE showed a PPP value of 100% and specificity of 100%. Not only did MAZE surpass the 90% criterion, but it also showed a PPP percentage and specificity level higher than that of ORF. In regards to fifth grade math, ORF was the only measure to display adequate levels of diagnostic utility. Its specificity was 95%, while its PPP was 93% These values surpassed the criterion and were higher than MAZE. The diagnostic utility statistics for each CBM are summarized below.

Table 14.

<table>
<thead>
<tr>
<th>CBM Measure</th>
<th>ISAT Section</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPP</th>
<th>NPP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF</td>
<td>Reading</td>
<td>61%</td>
<td>89%</td>
<td>96%</td>
<td>33%</td>
</tr>
<tr>
<td>MAZE</td>
<td>Reading</td>
<td>19%</td>
<td>100%</td>
<td>100%</td>
<td>20%</td>
</tr>
<tr>
<td>ORF</td>
<td>Math</td>
<td>30%</td>
<td>100%</td>
<td>100%</td>
<td>34%</td>
</tr>
<tr>
<td>MAZE</td>
<td>Math</td>
<td>1%</td>
<td>100%</td>
<td>100%</td>
<td>26%</td>
</tr>
</tbody>
</table>

At the sixth grade level the two CBMs showed similar percentages of diagnostic utility across category. For the reading section of the ISAT, ORF and MAZE both showed greater percentages of specificity (100%) and PPP (100%) than the other diagnostic utility statistics. The percentages in relation to predicting mathematics performance were within one percent of each other for both measures. MAZE showed a
slightly lower percentage of specificity (96%) and PPP (93%) than did ORF (97% and 94%).

Table 15.

*Diagnostic Utility Statistics for the 6th Grade Sample*

<table>
<thead>
<tr>
<th>CBM Measure</th>
<th>ISAT Section</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPP</th>
<th>NPP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF Reading</td>
<td>30%</td>
<td>100%</td>
<td>100%</td>
<td>34%</td>
<td></td>
</tr>
<tr>
<td>MAZE Reading</td>
<td>1%</td>
<td>100%</td>
<td>100%</td>
<td>26%</td>
<td></td>
</tr>
<tr>
<td>ORF Math</td>
<td>26%</td>
<td>97%</td>
<td>94%</td>
<td>38%</td>
<td></td>
</tr>
<tr>
<td>MAZE Math</td>
<td>21%</td>
<td>96%</td>
<td>93%</td>
<td>34%</td>
<td></td>
</tr>
</tbody>
</table>

**Discussion**

The purpose of this study was to examine the ability of two CBMs, ORF and MAZE, to predict 5th and 6th grade students' performances on the reading and mathematics subsections of the ISAT. In relation to both grade levels, the two CBMs showed similar correlations to the subsections of the ISAT. The strength of these relationships are similar to those obtained within previous studies (Hintze & Silberglitt, 2005; Jenkins & Jewell, 1993). Furthermore, the correlations within this study are higher than those obtained by Herrera (2005), who also examined ORF and ISAT reading at the 5th grade. The linear regression results suggest the two measures are essentially equivalent when predicting ISAT performance in reading and math. It was not until the 6th grade that MAZE accounted for significantly more variance in ISAT performance.
than did ORF, but this was only true for the mathematics subsection whereas ORF was a significantly better predictor of mathematics at the 5th grade level. This may be a product of the ISAT mathematics subsection. At the 5th grade level, reading may be required to simply access the necessary information to complete each item whereas at the 6th grade level students may require greater comprehension abilities to understand more complex task, but until this question is further investigated the above statement will remain speculation.

The results of this study at both grade levels also indicate the need to consider students' reading abilities when instructing them in seemingly disparate subjects such as math and science. Reading may possibly be a necessary access skill to obtaining mathematical instruction and knowledge (Helwig et al., 1999; Koponen et al., 2007), but it must also be noted that the relationship between academic achievement in reading and mathematics may be explained by the existence of an overarching complex cognitive construct or set of processes (Watkins, Lei, & Canivez, 2006). Finally, the relationship between reading and mathematics may also be of vital importance to perform adequately on statewide-standardized tests that use timed multiple-choice questions (Jiban & Deno, 2007; Silberglitt et al., 2006). This is especially important when one considers that the above-mentioned tests are those most likely to be used for holding schools accountable to the general public.

The ORF cut-score obtained within this study (5th grade = 125) for successfully meeting standards on the ISAT reading section at the 5th grade was higher than that observed by other research concerning the ISAT (Herrara, 2005). There has been no research attempting to establish an ORF cut-score for the 6th grade within Illinois, but the
value obtained in this study (WRC = 152) indicates that to do well on statewide tests student reading performance may need to increase across grades. The minimal ORF scores needed for adequate performance on mathematics also increased as grade level increased from 105 WRC to 161 WRC. Previously established ORF scores in relation to mathematics performance as well as in relation to MAZE and standardized test performance are lacking. The obtained cut-scores showed similar increases within each subsection at each grade level. This information possibly indicates that as performance requirements on statewide standardized tests increase with grade, so must students’ performance on benchmark assessments.

Concerning both measures, the diagnostic utility showed a similar pattern across grade levels. Both CBM measures showed higher percentages of specificity and PPP regardless of grade or subject. These results indicate that the cut-scores derived from both CBMs may add to educators’ diagnostic efficiency when attempting to determine who will meet standards on the ISAT, but not when attempting to determine who will not meet (Landau, Milich, & Widiger, 1991). This could possibly be the result of students receiving testing accommodations during the administration of the ISAT due to a legal mandate (i.e. IEP) that they did not receive during the administration of the CBMs.

The diagnostic utility statistics obtained within this study are markedly different than those obtained within the other study concerning upper elementary school students’ performance on the ISAT. Both CBMs showed greater levels of specificity than the previous study, but lower levels of sensitivity (Herrera, 2005).

While sensitivity and specificity may provide useful information towards diagnostic decision making, it may be more advantageous for individuals making those
decisions and administering interventions and resources to consider using the predictive power methods described previously (Landau et al., 1991). The CBM cut-scores for both grades tended to yield higher levels of PPP than of NPP. Thus, within both subsections of the ISAT the cut-scores for both CBMs tended to show a greater ability to show who would pass the ISAT than who would not pass the ISAT. MAZE was a superior measure compared to ORF in relation to its PPP at the fifth grade level. Despite this fact, it should be cautioned that the use of MAZE must be tempered due to a very low level of sensitivity. At the 6th grade level was nearly identical to ORF in its ability to predict who would pass the ISAT. These results support the previous hypotheses that MAZE would be more efficient in predicting ISAT performance than ORF. It appears that it is not until the very last part of elementary school that reading comprehension performance as measured by MAZE can discriminate who will or will not meet standards on the ISAT in a more efficient manner than ORF. This is especially important concerning the constant time restrictions placed on educators. MAZE may represent a more time efficient method to benchmark students at the upper elementary school grade levels.

Despite the above discussion, this study contains several limitations that should temper any conclusions based on its results. First, the sample sizes within this study are rather small and demographically limited. This is not fully representative of all the students who are administered the ISAT. Second, the lower sample sizes may have also reduced the variability of the scores utilized within this study. This may have influenced the statistics calculated, and thus, possibly negate the conclusions drawn from its results. Furthermore, MAZE scores were derived from simple correct identifications as opposed to percent correct selections. This may have possibly restricted the range of scores
derived from the MAZE CBM. Finally, the CBM data were collected by teachers and other school personnel not typically involved in standardized test administration. No formal administration of integrity surveys was conducted. CBM scores could have possibly varied by not only individual student performance, but also by the individual administering the assessment.

The above discussion helps to inform future research. Future research concerning the ability of CBM measures to predict student performance on the ISAT should consider several areas. First, future research should utilize a larger and more representative sample size. This may help to increase the amount of variability within the CBM and ISAT scores. For the same reasons, it may be advantageous to utilize percent correct selections instead of correct selections when examining MAZE’s predictive abilities. Also, it would be prudent to implement some sort of integrity check concerning the administration of the CBM assessments. This could not be done within the current study due to the archival nature of the data. Finally, due to the shared cognitive processes between reading and math and the amount of variance shared by the reading CBMs and ISAT mathematics, it may be appropriate for future researchers to examine the predictive abilities of these measures when coupled with a math calculation benchmark measure. This may help to create a more efficient diagnostic pattern (Landau et al., 1991; Sibley et al., 2000). This was not performed within the present study due to the lack of universal mathematics benchmarking at the schools from which the data were obtained.

Conclusions and Implication
In conclusion, this study indicates that reading CBM measures may be used to adequately predict student performance on the ISAT reading subsection at the upper
elementary school levels. Furthermore, the CBM measures were also moderately related to student performance on the ISAT mathematics section. The hypotheses that predicted that MAZE would be a better predictor than ORF for both ISAT subsections and for both grade levels were partially supported. Only at the 5th grade did MAZE appear to be a more efficient measure in determining who would pass the ISAT subsections. The diagnostic utility statistics for the 6th grade subsections were similar, but ORF still appeared to be a more efficient diagnostic measure. The implication of this study is that given the extensive literature base supporting the use of ORF to predict student outcomes and the dearth of similar research concerning MAZE, it may be more appropriate to utilize ORF for predicting student performance on the ISAT, until the utility of MAZE for doing the same is well established.
References


