

5-2008

# Possible Predictive Factors for the Greater Johnstown School District's Pennsylvania System of School Assessment Reading Results

Mary E. Armstrong Rauch  
*Indiana University of Pennsylvania*

Follow this and additional works at: <http://knowledge.library.iup.edu/etd>

---

## Recommended Citation

Rauch, Mary E. Armstrong, "Possible Predictive Factors for the Greater Johnstown School District's Pennsylvania System of School Assessment Reading Results" (2008). *Theses and Dissertations (All)*. 1033.  
<http://knowledge.library.iup.edu/etd/1033>

This Dissertation is brought to you for free and open access by Knowledge Repository @ IUP. It has been accepted for inclusion in Theses and Dissertations (All) by an authorized administrator of Knowledge Repository @ IUP. For more information, please contact [cclouser@iup.edu](mailto:cclouser@iup.edu), [sara.parme@iup.edu](mailto:sara.parme@iup.edu).

POSSIBLE PREDICTIVE FACTORS FOR THE  
GREATER JOHNSTOWN SCHOOL DISTRICT'S  
PENNSYLVANIA SYSTEM OF SCHOOL ASSESSMENT  
READING RESULTS

A Dissertation

Submitted to the School of Graduate Studies and Research

in Partial Fulfillment of the

Requirements for the Degree

Doctor of Education

Mary E. Armstrong Rauch

Indiana University of Pennsylvania

May 2008

© 2008 by Mary E. Armstrong Rauch

All Rights Reserved

Indiana University of Pennsylvania  
The School of Graduate Studies and Research  
Department of Educational and School Psychology

We hereby approve the dissertation of

Mary E. Armstrong Rauch

Candidate for the degree of Doctor of Education

---

Edward M. Levinson, Ed.D.  
Professor of Educational and School Psychology,  
Advisor

---

Mary Ann Rafoth, Ph.D.  
Dean, College of Education and Educational  
Technology

---

William F. Barker, Ph.D.  
Professor of Educational and School Psychology

---

Victoria B. Damiani, Ed.D.  
Professor of Educational and School Psychology

---

Michele S. Schwietz, Ph.D.  
Assistant Dean for Research  
The School of Graduate Studies and Research

Title: Possible Predictive Factors for the Greater Johnstown School District's  
Pennsylvania System of School Assessment Reading Results

Author: Mary E. Armstrong Rauch

Dissertation Chair: Dr. Edward Levinson

Dissertation Committee Members: Dr. Mary Ann Rafoth  
Dr. William Barker  
Dr. Victoria Damiani

Relative to Curriculum-Based Assessment (CBA) and the Pennsylvania System of School Assessment (PSSA), this study sought to determine whether a school district's CBA reading scores were predictors for the PSSA reading scores for a grade-level cohort of students. Those CBA reading scores were collected in fourth through seventh grades; while the PSSA measure was a scaled score earned on the eighth grade reading measure. The sample included 268 eighth grade students who were roughly 48% male, 22% registered as being an ethnic minority, and 21% categorized as having an educational disability. Pearson and Spearman's correlations between fourth grade reading grades and fourth grade CBA scores validated the district CBA reading measure as a measure of reading skill. The students received two possible reading curricula, the Johns Hopkins' *Success for All* reading program or a *Houghton Mifflin* reading basal program. Consistent with the literature, students categorized with educational disability performed more poorly on the eighth grade PSSA than their peers ( $p < .01$ ), and CBA reading ( $p < .01$ ) was a significant predictor of PSSA reading performance. Conversely, neither sex ( $p \geq .05$ ) nor curriculum ( $p \geq .05$ ) was a significant predictor of that PSSA reading performance.

## ADKNOWLEDGMENTS

Blessings and gratitude are due to my husband, family, the various church families with whom I have worked, and my department at Indiana University of Pennsylvania. Special thank yous are sent to Ruth Ann Ficco and Dr. Gerald Zahorchak, who shared their data and process at the Johnstown schools. Additionally, I benefited from the great minds who created and sustain state testing in Pennsylvania, among whom are Dr. James Masters and Richard Murray. My hope is that this study will help to tie together the body of current knowledge regarding curriculum-based assessment and state testing.

## TABLE OF CONTENTS

Chapter		Page
I	INTRODUCTION.....	1
	Statement of the Problem.....	1
	Reading Assessment.....	1
	Assessment of Students of Diverse Needs.....	2
	Curriculum-Based Assessment.....	3
	Curriculum-Based Assessment and Published Norm Referenced Achievement Tests.....	6
	Curriculum-Based Data and High Stakes Tests.....	7
	History of the Pennsylvania System of School Assessment.....	8
	Purpose of the Study.....	11
	Rationale and Importance.....	15
	Research Questions.....	16
	Definition of Terms.....	18
	Assumptions.....	18
	Limitations.....	19
	Summary.....	20
II	LITERATURE REVIEW.....	21
	Reading Assessment.....	21
	Social Goals and Supports for Reading Education.....	22
	Authentic Assessment.....	25
	Characteristics and Advantages of Curriculum-Based Assessments.....	25
	Curriculum-Based Assessment and Reading Performance.....	28
	Curriculum-Based Assessment Process.....	32
	Student Disability and Reading Performance.....	33
	Socioeconomic Status and Ethnicity Factors as Predictors of Student Reading Performance.....	34
	Reading Disability and Sex of the Student.....	36
	Familial Factors in Reading Difficulties.....	36
	Web-Based Assessment Systems.....	37
	Curriculum-Based Assessment of Early Literacy Skills.....	39
	Reading Curricular Design.....	45
	Data-Based Reading Supports.....	45
	Curriculum-Based Measurement as a Discriminant Measure.....	47
	Technical Features of Curriculum-Based Measurement.....	49
	Psychometrics of Curriculum-Based Measures.....	49
	Curriculum-Based Measurement: Expanded Practices.....	51

	Page
Curriculum-Based Measurement and Differentiated Instruction.....	52
Technological Supports for Differentiating Instruction.....	53
Curriculum-Based Measurement and Curriculum.....	54
Curriculum-Based Measurement and Students with Disability.....	56
Reading Development Across Grade Levels.....	60
Reading Level and Race.....	62
Curriculum-Based Measurement: Basic and Advanced Uses.....	63
Curriculum-Based Measurement and Ethnicity.....	63
Reading Progress for Students with Reading Difficulty.....	65
Applying Reading Research in the Curriculum.....	66
Using Alternate Curriculum in Curriculum-Based Measures.....	66
Accountability Measures and Reading Assessment.....	67
Curricular Data as a Predictor of Pennsylvania System of School Assessment Performance.....	69
Summary.....	70
III METHODS AND PROCEDURES.....	72
Method.....	72
Introduction.....	72
Research Questions and Hypotheses.....	73
Research Question One.....	73
Research Question Two.....	73
Research Question Three.....	74
Research Question Four.....	74
Design.....	75
Population.....	78
Sample.....	78
Assignment.....	79
Measurement.....	79
Curriculum-Based Assessment.....	79
Pennsylvania System of School Assessment.....	81
Curricula and Related Hypothesis.....	83
Success for All.....	83
Houghton Mifflin Basal Reading.....	84
Procedures.....	85
Data Collection.....	86
Sample Size.....	87

	Page
Statistical Analyses.....	90
Questions, Hypotheses, Predictors, Statistical Analyses, and Statistical Assumption for the PSSA Predictors Project.....	90
Research Questions.....	96
Assumptions.....	97
Summary.....	99
 VI	
RESULTS.....	101
 Introduction.....	101
Research Question One.....	102
Research Question Two.....	102
Research Question Three.....	102
Research Question Four.....	102
Complications.....	103
Computer Programs.....	103
Analysis: Step One.....	103
Research Question One.....	103
Results of Step One.....	105
Analysis: Step Two.....	105
Research Question Three.....	105
Results of Step Two.....	106
Analysis: Step Three.....	113
Research Question Two.....	113
Results of Step Three.....	114
Analysis: Step Four.....	121
Research Question Four.....	121
Results of Question Four.....	126
Summary.....	126
 V	
DISCUSSION.....	128
 Introduction.....	128
Purpose of the Study.....	128
A Review of the Procedures and Analyses.....	128
A Review of the Results.....	130
Research Question One.....	130
Research Question Two.....	131
Summary.....	134
Research Question Three.....	135
Educational Disability.....	136
Sex of the Students.....	138
Research Question Four.....	139

	Page
Implications of the Study.....	140
Recommendations for Future Research.....	146
Limitations of the Study.....	149
Summary.....	153
<b>REFERENCES.....</b>	<b>155</b>
<b>APPENDICES.....</b>	<b>179</b>
Appendix A – Standards and Rubrics.....	180
Appendix B – Letter of Approval.....	183
Appendix C – Curriculum-Based Assessment Paragraphs.....	185

## LIST OF TABLES

Table		Page
1	Measures which Predict Reading Difficulties at School Entry.....	43
2	NAEP Long-Term Reading Results: Male, Female, Black, Hispanic and White Students Scoring at or above Scaled Scores of 150/200/250/300.....	64
3	CBA and PSSA Assessment Project Task.....	88
4	Research Questions, Hypotheses, Predictors, Statistical Analyses, and Statistical Assumptions.....	91
5	Case Processing Summary.....	104
6	Correlation Matrix of JSSA Reading Performance Level and Reading Grade Point Average of Fourth Graders from the Greater Johnstown School District.....	104
7	Multiple Regression Predicting the Eighth Grade PSSA Reading Performance Level from Sex of the Student and Disability.....	107
8	Multiple Regression Predicting the Eighth Grade PSSA Reading Performance Level from Sex of the Student and Specific Disability.....	109
9	Multiple Regression Predicting the Eighth Grade PSSA Reading Performance Level from the Fourth- through Seventh-Grade JSSA Reading Performance Level, Sex of the Student, and Disability.....	111
10	Multiple Regression Predicting Eighth Grade PSSA Reading Performance Level from Sex of the Student, Specific Disability, and Fourth- through Seventh-Grade District Curriculum-Based Assessment (JSSA) Reading Performance Levels.....	115

Table		Page
11	Multiple Regression Predicting the Eighth-Grade PSSA Performance Level from the Fourth-Grade JSSA Performance Level, Sex of the Student, and Specific Disability.....	118
12	Multiple Regression Predicting PSSA Reading Performance Level from Sex of the Student, Disability, Curriculum-Based Assessments, and Two Reading Curricula.....	122
13	Multiple Regression Predicting PSSA Reading Performance Level from Sex of the Student, Specific Disability, Curriculum-Based Assessments, and Two Reading Curricula.....	124

## LIST OF FIGURES

Figure		Page
1	Research Design Diagram for PSSA Predictors Project.....	76
2	Research Path Diagram for PSSA Predictors Project.....	77

## CHAPTER I

### INTRODUCTION

#### Statement of the Problem

#### *Reading Assessment*

Of the core curriculum skills, reading and writing typically develop as a part of normal developmental language experience. Mastery of language skills is necessary to successful performance in elementary school. Children of all backgrounds assign meaning to language based on their own experiences (Harp, 1994). Hennings (1992) describes the components of prior knowledge, structure and organization of language, and metacognition as necessary to the interpretation of written words. Children require opportunities to predict and problem-solve using meaningful personal experiences to move through the developmental stages which produce competent readers (Harp, 1994).

Lyon and Shaywitz (2003) recently analyzed this process of becoming a competent reader, positing specific brain anatomy modules which serve specific skill functions. Alternately, Coles (2004) postulates a more complex, environmentally- and instructionally-related process which cannot yet be guided by brain anatomy and research.

As the neurological processes which undergird reading cannot be examined directly, the nation and its educators seek to determine reliable data with which to examine and monitor the reading process. Many researchers begin by examining groups of students.

Slavin (1997) describes four groups of developing readers (1997). He estimates that 40%-50% of children will read regardless of the quality of instruction. These

children he calls natural readers. Another 30%-40% of children are captured by the term teachable readers, as high quality, well-delivered instruction ensures their success (Slavin, 1997). Young at-risk readers are categorized as tutable readers as they benefit from additional ecological supports and individual tutoring in first grade (Slavin, 1997). Last, Slavin (1997) notes the approximately 1%-2% of children who qualified as true dyslexics, demonstrating difficulty reading even after all possible data-based supports have been applied. It is further noted that these 1%-2% of children with dyslexia do not equate to the approximately 5% of children nationally who have been categorized as being learning disabled in reading. These struggling readers demand concrete and meaningful instruction to learn to read.

American education, as it continues to evolve, struggles to bridge this dichotomy of increasingly diverse students and increasingly well-defined standards of academic performance. One significant attempt to meet this dichotomous union is curriculum-based assessment. Curriculum-based assessment (CBA) is a concept which fits the educational and political objective of our time, specifically that of successfully educating high-achieving, average, and low-achieving students, as well as those students with special needs.

#### Assessment of Students of Diverse Needs

Children of diverse needs require monitoring of their progress in their own reading curricula. According to Salvia and Ysseldyke (1985), criterion-referenced reading assessments are helpful in determining reading strengths and weaknesses and in designing well-fitted interventions for students. These authentic assessments can also highlight common problems in groups of children, and so help to focus instructional time.

Finally, the use of these assessments is based on the curriculum which best related to the specific skills and understandings which a student has captured and personalized (Edyburn, 1994).

Christenson, Ysseldyke, and Thurlow (1989) identify instructional factors which are essential when instructing students with mild disability. These factors include effective classroom management, positive school environment, a match between instructional level and curriculum, clear and explicit expectations for student performance, instructional performance monitoring and instructional adjustment, sufficient and efficient instructional time, high opportunity for student response, active monitoring for understanding, and frequent and appropriate performance evaluation (Christenson, et al., 1989). Instructional match, appropriate goals and expectations, instructional design, and monitoring and evaluation are all addressed directly through CBA. CBA uses a measure of actual academic skill or fluency to begin instruction at a mildly challenging level of curriculum and to monitor how well students progress with particular instructional interventions.

### Curriculum-Based Assessment

CBA is a method where student performance is observed and measured or rated within the student's own curriculum. CBA represents a wide range of procedures which are not statistically or procedurally similar. One type of CBA, entitled curriculum-based measurement (CBM) is validated as sensitive to small increments of growth, brief in sampling time, procedurally standardized, and reliable across disaggregate groups of students (Shinn, 1995; Shinn & Good, 1993). In its various iterations, CBA reading assessment allows educators to observe and categorize student reading skills, while CBM

allows teachers to listen to the student reading aloud for a short period while noting the student's errors, skipped and substituted words, and the rate of reading in words-per-minute. In order not to allow a single unknown word to spoil the rate measure, unknown words are read aloud to the student after a failed student attempt to decode the word.

CBA allows educators to evaluate the appropriate entry or instructional level within the curriculum as well as providing guidance to plan and monitor the academic progress of students. In 1977, Gickling presented the CBA process to Texas evaluators in the first known use of the CBA term (Rosenfield & Kuralt, 1994). The CBA process encourages accurate "instructional match" between the student's emerging understanding and skill and the difficulty or level of the specific curriculum. CBA uses the scope and sequence of academic curricula to determine accurate student placement and systematic progress in that very curriculum (Gickling & Armstrong, 1978).

Marston (1989) reviews the increased usefulness of curriculum-based assessment when compared to individually administered norm-referenced tests, such as published academic achievement tests. Individual, standardized tests have weaknesses in technical adequacy and do not provide valid, legally defensible, decision-making data (Marston, 1989). Published achievement tests often fail to match or sample the students' curricula, do not accurately sequence the skills taught and learned, or may use vocabulary better matched to another specific curriculum. Standardized tests cannot reflect local instructional practice and often are inaccurate predictors of progress in local curriculum (Shapiro & Derr, 1987). Further, instructional planning demands data concerning specific skills: error analysis, fluency or mastery, and comprehension, none of which are readily addressed with standardized measures.

Best practice in CBA demands that evaluators begin with the end in mind. In reading, that meaningful general outcome is the understanding of what is read. Of the many skills which must be mastered to achieve a broad education, reading comprehension has been the gateway through which each educated student must pass. To facilitate developing reading comprehension skill, educators attempt to define the student's instructional level with reading assessments and to move the student toward incrementally more difficult passages. "Reading comprehension, in a literate society, is the key to cognitive development" (Wick, 1987, p. 92), and must be the focus of both assessment and effective instruction.

A variety of reading strategies are propounded and employed to improve reading comprehension, including inflection, cueing at punctuation marks, self-monitoring for accuracy, monitoring of pace, and tracking with a pointer or indicator. Improvements occur over different interventions for different students (Allinder, Dunse, Brunken, & Obermiller-Krolikowski, 2001). Still, the selection and measurement of effectiveness across many strategies demands a simple and unified measure.

CBA is a data-based form of academic assessment, easily employed in the classroom and authentically linked to the students' curriculum. Deno (1987) defines CBA as a set of measurement procedures which use the teacher's observation and recording of a student's performance in the local district or building curriculum to make instructional decisions. Deno's seminal work using CBA, entitled *data-based program modification* (Deno & Mirkin, 1977), fueled intensive research of reading and assessment program alternatives that served students' diverse instructional needs. Research by Shinn (1989a,

1989b), Hintze and Shapiro (1997), and Fuchs and Fuchs (1992) all attest to the effectiveness of the CBA model in assessment and planning in a reading curriculum.

In a parallel CBA development, Rosenfield (1987) and Shapiro (1989) address use of the Informal Reading Inventory (IRI) to establish the correct placement of students in the instructional sequence. The IRI and other forms of curriculum-based assessment provide both guidance for instructional placement and help to monitor instructional progress. The IRI uses reading pace and words read correctly per minute, as well as student hesitation and errors, to determine how difficult a passage is for that student.

Curriculum-based measurement differs from curriculum-based assessment in its stronger technical character, its ability to provide a numeric value, and its standardized process.

### Curriculum-Based Assessment and Published Norm

#### Referenced Achievement Tests

Marston, in Shinn's (1989) text, reviews the increased usefulness of curriculum-based assessment (CBA) when compared to individual published, norm referenced achievement tests (PNATS). These standardized tests have weaknesses in technical adequacy and have not provided valid, legally defensible, decision-making data (Marston, 1989). Published achievement tests often fail to match or sample the students' curricula, do not accurately sequence the skills taught and learned, or use vocabulary better matched to another specific curriculum. Standardized tests cannot reflect local instructional practice, and often are inaccurate predictors of progress in local curricula (Shapiro & Derr, 1987). Further, instructional planning demands data concerning

specific skills: error analysis, fluency or mastery, and comprehension, none of which are readily addressed with standardized measures.

### Curriculum-Based Data and High Stakes Tests

Such curriculum-based data may provide additional value as a predictor of student success on the standards driven Pennsylvania System of School Assessment (PSSA) of reading. The PSSA measures student reading performance as reflected in written responses to the reading probe. No oral directions are given and all students receive a reading probe and a related question which represents grade-level reading, unrelated to the students' instructional levels. Students read the probe silently, then read the question. A draft response is commonly created during the first day, while a final clean copy is completed the next day. The Pennsylvania Department of Education (PDE) rating rubric measures the responses according to these criteria:

1. "Below Basic" responses demonstrate limited understanding of the text, include errors in text-based facts, and are disjointed, incomplete or irrelevant.
2. "Basic" responses consist of literal responses to the text, with continued factual errors and disjointed production.
3. "Proficient" responses are more coherent, with no major comprehension errors and some elaboration or extension of understanding.
4. "Advanced" responses, with a thorough understanding of textual complexity, demonstrate personal, critical and evaluative elements. The "Advanced" responses make connections to personal, cultural or background knowledge, as well.

It would be quite useful to know if student performance at particular developmental levels accurately predict an individual student's performance on the PSSA reading assessment. In light of the direct-instruction time and resources expended, it would be strategically helpful to determine whether reading assessment at specific elementary and middle school grades better predict those students' later scores on the PSSA reading assessment. Additionally, it would be valuable to examine whether those results differ by demographic characteristics of the students, such as educational disability or sex of the student. Finally, investigation of whether the reading curriculum chosen by the students' elementary school staff impacts on the success of these students is desirable.

#### History of the Pennsylvania System of School Assessment

The PSSA measures were developed in 1992 by a small committee at the PDE. Those initial assessments were broader, less focused, and included more measures of thinking skills than do current PSSA assessments. In later years, the committee expanded to 230 educators and PDE staff members who were involved in further developing the test some. Fifty to sixty participants developed each grade level measure (Kohr, 2004; Masters, 2004).

Leadership staff at PDE had been working on a procedure to set true performance goals since 1990, when TELLS testing shifted to the PSSA assessment. The Technical Advisory Committee leaders during this period included Camara, Hill, Hamilton, Lewis, and Lane, as well as Kohr, Masters, and Lock from PDE. School-based performance data and system evaluation was the initial political and district focus, precluding the production of individual student data (Murray, 2004).

Beginning in 1996 as a base year, the PDE development team equated the tests, anticipating the implementation of academic standards. Two methods were used: the bookmarking and borderline groups methods. Hamilton, Lewis, and Lane began to rate and bookmark items from least to most difficult in order to choose the cut points at each performance level (Murray, 2004). In bookmarking, teachers and educators estimated what the performance scores would be for students, and the policy makers at PDE used this information in determining what the cut scores would be. The State Board then examined these PDE recommendations and approved the cut scores.

Lewis led the performance standard setting process (Murray, 2004). In Lancaster, in the winter of 1999-2000, a large group of educators, policy makers, and parents met to review the presented standards. The borderline and bookmarking methods yielded nearly identically predicted scores, once scores were averaged for some areas and grade levels (Murray, 2004).

In 1999, Chapter IV was passed and standards were first measured. Because there was limited planning time, some standards were better measured than were others (Masters, 2004). At first, the data was not released by individual names, but only released as aggregated data by Data Recognition Corporation (DRC) to PDE and other researchers (Murray, 2004). In the years following 1999, standards measurement became ever more thorough and precise, allowing individual reporting as well.

Initial response rubrics for the PSSA open-ended responses included ratings of zero-five, however a “five” was not reached by a large number of students in the original math portion of the test. Some public responses included criticism that only gifted students or very bright students could earn five ratings and not the average student, which

fueled doubt about the statistical validity of this measure. Such criticism has not been proven by any study to date. This criticism led to an effort for more defined assessment standards, and in 2004, still more specific assessment anchors. These anchors are more focused, allowing students and schools to direct their performance toward greater success rates. The total performances from past PSSA resemble a bell curve. It should be noticed that *No Child Left Behind* (NCLB) legislation does not seek information on advanced-level PSSA performances/students. The top level of student data requested by NCLB is Proficient.

Types of items and testing procedure have remained stable from the initial PSSA assessment to 2004, however open-ended items were rebalanced across content areas as needed (Murray, 2004). The cut scores were set in 2001 during the Zogby administration with an increase of 0.25 of a standard error to improve rigor. The 2001 and 2002 tests were the first opportunity to review student performance changes with these cut scores applied. Good equating was accomplished in 2002 and 2003, with increasing focus on the standards. By 2003, the PSSA was a reliable and efficient measure of the standards (Masters, 2004).

Then, in 2003 as PDE prepared to release the 2004 state test, PDE administrators under Pennsylvania Governor Rendell instead released the test to contractors. Items developed in collaboration between PDE and the contractors were used in the 2004 test. DRC was enlisted by the Department of Education to make final decisions regarding the test, with review by both PDE staff and outside consultants (Masters, 2004; Murray, 2004).

In 2005, the academic anchors were measured for the first time. Whatever limited rigor currently exists is a reflection of the anchors upon which the test is based (Masters, 2004). PDE administrators intensely desire to focus this test, but did not intend to create an easier assessment. Analyses of the depth of knowledge of the anchors were completed with intent to create items at the levels of comprehension (Bloom's Taxonomy-2) and application (Bloom's Taxonomy-3) (Masters, 2004).

As the PSSA moved toward being a measure of academic standards, it better measured more narrow academic skills while becoming a lesser measure of broad, higher-level thinking skills. In 2004, Pennsylvania moved to measuring academic anchors, in order to increase the focus of the PSSA. Depth of knowledge analyses were carried out in an attempt to raise the cognitive level of the anchors (Masters, 2004).

#### Purpose of the Study

This study is an examination of the benefits of the Johnstown System of Student Assessment (JSSA). This local district system of curriculum-based assessment provides data on instructional level and progress within the curriculum, as well as responses for qualitative error analysis. It will be identified at which particular grade levels the district's CBA developmental assessment of student reading (JSSA) best relates to particular students' eventual performance on the eighth-grade level PSSA assessment of reading. Whether this prediction differs by demographic characteristics of the student, or for the two district elementary reading curricula, John Hopkins' *Success for All* or the *Houghton Mifflin* Reading Basal Series will also be examined. If the JSSA reading scores and PSSA reading results of particular demographic groups of students (boys or girls, regular education students, or students with known disability) are better correlated when

limited to one curriculum, that curriculum might prioritize the specific reading standards measured by the PSSA.

In the current national *No Child Left Behind* and PDE educational schema where data-based instruction and intervention practices are prioritized, the link between direct instructional time and student achievement outcomes is frequently examined (Bender, 2002). Curriculum-based assessment, while time efficient, demands a regular portion of that instructional time. Curriculum-based assessments of student achievement are needed to provide accurate and useful data which contribute to educational progress. However, if the curriculum-based assessments or the state standards-based assessments (i.e., PSSA) are not each valid predictors of eventual achievement outcomes, then the curriculum-based or state standards-based assessment needed to be re-tooled to serve that primary purpose.

In developing these multipurpose JSSA CBA measures, the research regarding early literacy development and written assessment was reviewed by the school district selection team. The general body of evidence indicates that paper-and-pencil assessments are not recommended for young children (Brendekamp & Copple, 1997). In fact, best practice dictates that young children are best assessed in activity embedded in play, which is their curriculum (Swadener & Kessler, 1991). As children increase their literacy skill, more formal testing becomes a valid alternative. Through this review of CBA developmental and learning research, it is posited that, from second through fourth grade, the students' reading assessments will become increasingly accurate predictors of their later PSSA scores. For this reason, CBA data beginning at grade four was used for this study. However, some research describes a slower neurological maturity for young boys

(Gesell & Ametrude, 1954; Minke & Bear, 2000), which may predict a less reliable fourth-grade reading score for males. Still, if the JSSA4 accurately predicts the PSSA8 reading scores for most students, the JSSA in fifth through seventh grades may be redundant measures.

The two reading curricula to be examined across these grade levels are the John Hopkins' *Success for All* and the *Houghton Mifflin* basal reading series. It is hoped that an examination of each curriculum, in order to determine the surface or face curriculum as well as the underlying concepts or metaphors, may help to discriminate the better reading curriculum for this school district. Additionally, of the two reading curricula, whichever best matches student demographic characteristics and Pennsylvania instructional standards, as measured by the PSSA reading assessments, should allow the most accurate prediction of those PSSA results.

The East Side Elementary School (ES) serves more than 800 students in grades four-year-old Kindergarten through fifth grade in an urban, neighborhood school. The West Side Elementary School (WS) serves more than 1,000 urban students in the same grades in a suburban setting, outside the school district. Ethnic diversity is similar across the two elementary schools, with Caucasian students composing 85% of the ES population and 80% of the WS population. Similarly, the free and reduced lunch rates across the schools are comparable, with an 82% rate at the ES school and a 77% rate at WS school. As the John Hopkins' *Success for All* reading curriculum purports to be designed for populations of students who are diverse in background and at risk for academic failure, it is thought to be the better demographic fit for this local community.

John Hopkins' *Success for All*, the East Side Elementary School reading curriculum, was designed by John Hopkins University to present multimodal instruction within a single, defined instructional-level classroom to children with diverse needs and backgrounds. The *Success for All* curriculum uses proprietary materials and texts as well as related children's literature, within a scripted and timed presentation, in order to encourage high interest and participation across groups of diverse students. Language arts and vocabulary are also instructed within the 90 dedicated minute reading period.

The *Houghton Mifflin* basal reading curriculum, offered throughout the WS school, provides a literature-based curriculum implemented within a mixed instructional level group of students. This curriculum stresses phonics, decoding, and vocabulary skill development within a traditional reading classroom. *Houghton Mifflin* is a literature-based program which provides unabridged versions of children's literature, with ample opportunity for associated writing prompts (Lesiak, 1998). The vocabulary depth and difficulty, as well as the typical length of passages, differ within this curriculum. Each reading passage relates to the specific skills, processes, and changes in domain which best affect the transfer of reading skills to a body of students (Marzano, et al., 1989). The WS school also dedicated 90 minutes to reading.

As the district needs a common measure to evaluate reading progress across these two curricula, prompts were selected for measuring reading progress. As the *Houghton Mifflin* curricula contained many prompts which met the chosen criteria, prompts were recommended from that traditional curriculum (Ficco, 1999).

Racial diversity was relatively similar, as each school served between 80%-85% of Caucasian students (80% for the WS school and 85% for the ES school). Each school used the same dedicated period of 90 minutes each morning, in order to present their curriculum.

Prompts used for measuring student progress were selected from the traditional, *Houghton Mifflin* reading curriculum. Each prompt contained at least 200 words of a high-interest, developmentally appropriate prose passage. Readability scores were calculated using grade-level word lists plus the publisher's information. Student directions were read aloud and any unfamiliar, unique vocabulary was introduced in the classroom setting before assessment in order to minimize the effects of cultural difference (Ficco, 1999). Students then responded in writing to a series of three questions, and the answers to those prompts were intended to measure reading comprehension (Ficco, 1999).

Scoring team members for the resulting student responses were chosen from reading specialists and other educators who had been trained by the PDE as scorers for the state reading assessment. The PDE trained District Trainers provided formal professional education presentations for those district staff who participated locally.

### Rationale and Importance

Greater Johnstown School District (GJSD), like many other urban, fiscally challenged institutions, had a need to assess children's reading skills with great strategic elegance, using the most effective available brief measure. The district met with school staff, community members, administrators, and reading specialists in order to gain consensus from these stakeholders regarding the chosen measure (Ficco, 1999).

Although the reading specialist group had made recommendations regarding the prompts, the decision to use CBA was an internal, cooperative and political process not directly influenced by external expert advice. The stakeholders were familiar with CBA measures, and the district chose to implement the process immediately, as it was evident that many students had entered school with very limited concepts of print, phonological skill, and experience with literature. Research clearly states that the years between five and nine are critical to establishing those skills (National Reading Panel, 2001; Simmons & Kameenui, 1998). Further, this highly mobile population encourages a system which will continuously assess students' ability and skills in reading, across the two elementary reading curricula. For these reasons, the district mission was centered on creating a district-wide curriculum-based system of measurement which would best maximize instructional time and student ability.

The curriculum-based assessment results which are collected three times each school year are scored by a PDE trained local staff team, using the same standards and rubrics taught at PDE trainings (See Appendix A). Instructional placement decisions, progress monitoring for individual students, and suggestions for titrating the instructional time in particular classrooms result from these assessments. This study is undertaken to examine the reading results for the disaggregated group of students, relating their curriculum-based results to their later PSSA reading scores.

#### Research Questions

1. How predictive were the students' JSSA4 through JSSA7 scores of their later PSSA scores, when limited to the John Hopkins' *Success for All* reading curriculum?

It is hypothesized that JSSA4 scores will predict PSSA8 scores, with little added value from the JSSA 6 and JSSA 7 measures, as dedicated reading instruction does not extend to the sixth and seventh grade years. Few students have made unusual gains after the “fourth grade slump,” and about 14% of all US students from grades 3 through 11 underachieve in reading (Snow, Burns & Griffin, 1998).

2. How predictive were the students’ JSSA4 through JSSA7 scores of their later PSSA scores, when limited to the *Houghton Mifflin* curriculum?

It is hypothesized that JSSA4 scores will predict PSSA8 scores, with little added value from the JSSA6 and JSSA7 measures, for the same reason.

3. Did the factor of sex of the student improve the ability to predict later PSSA scores within either curriculum?

It is hypothesized that sex of the student may be an additional predictive factor in determining later PSSA reading scores from JSSA scores. Statistics demonstrate a higher percentage of fathers than mothers who self-reported reading disability, and higher percentages of boys than girls who are diagnosed with Attention Deficit Hyperactivity Disorder and other physiological and psychological diagnoses which may have increased the predisposition of male children to demonstrating reading difficulties (National Reading Panel, 2000; Snow, Burns, & Griffin, 1998).

4. Were other measured demographic factors significantly predictive of the students’ PSSA scores; and how did these factors interrelate?

It is hypothesized that various combinations of sex and disability may predict PSSA performance. Although high economic disadvantage (>75% of students receiving free and reduced lunch) is known to relate to a higher rate of students with reading

difficulty, this rate did not differentiate the two cohorts studied in the district (Snow, Burns, & Griffin, 1998).

### Definition of Terms

JSSA reading scores will be determined as the standards-based scores from the spring assessments of reading comprehension for each grade, fourth through seventh grade. These scores are reported as: BB: Below Basic, B: Basic, P: Proficient or A: Advanced, with Proficient or Advanced scores viewed as meeting academic expectations. These scores are earned based on the PSSA system of rubric scoring.

PSSA reading scores are the numeric reading comprehension scores earned by students on their eighth grade Spring PSSA reading exam.

Reading curriculum for each student is determined as that reading curriculum, in which the student was instructed for the majority ( $\geq 75\%$ ) of his or her time in the GJSD.

Exclusions: All GJSD students in the chosen cohort who have attended the district during the measured elementary and middle school years are included in this body of data. However, any students missing the PSSA eighth grade reading assessment or those assessed while attending other districts, are eliminated from the study.

### Assumptions

The following assumptions shapes the design and completion of this study. It is first assumed that each level of district CBA results and district reading grades are well correlated measures of reading comprehension skill. That correlation will be prepared. It is further assumed that the CBA measure has structural integrity--that it does not simply reflect another central skill or characteristic of the students in each cohort. This CBA

measure is then assumed to measure reading comprehension equally well in each cohort using the traditional *Houghton Mifflin* prompts as cold read samples. Research indicates that a traditional measure can be used accurately for each cohort, both those using the traditional and those using an alternate curriculum (Hintze & Shapiro, 1997).

Finally, the study assumes that students' CBA scores are indicators of their progress in achieving Pennsylvania reading standards, as measured by the PSSA. Statistical analysis was employed to determine predictive factors and to examine the interactions among those factors.

### Limitations

Possible limitations of the data include unmeasured variables across participants, instruments, and procedure. First, student participants in one building might have a much higher absentee rate, and this attendance rate might affect student progress in sequentially based skill development. Second, the sensitivity of the individual prompts to measure increments of student growth as well as the varied readability and vocabulary levels may limit the recognition of student progress for all students or for a disaggregated subgroup of students. Third, there may be a different assessment schedule across buildings, grades or classrooms. Further, those different locations are subject to diverse ecological differences in noise level, heating or cooling systems, or assigned teachers versus substitute staff. Finally, perhaps those teachers and students who regularly use the traditional curriculum are more or less motivated by the familiar style of the prompts than the teachers using *Success for All* to present the task as a meaningful activity and an important measure.

The usefulness and applicability of this study would be augmented by replication, expansion to other cohorts within the school district, or extension of the cohort study to 11<sup>th</sup> grade PSSA and JSSA assessment measures. The comparison of 11<sup>th</sup> grade CBA, PSSA, and available SAT assessment data would strengthen the conclusions drawn.

### Summary

Reading comprehension and fluency measures have been used to measure the academic progress of regular education, at-risk students, and students with disability as they all learn to read. School staff and parents want to follow the progress of these students and to encourage each student to achieve the best skills possible for each one.

The GJSD, an urban district with a high percentage of students from low socioeconomic status (SES) homes, sought to rapidly and efficiently increase the reading skills of these students. The educational mission required an economical and accurate system of brief assessment in order to monitor the progress of students across time, classrooms, and curricula.

The JSSA was created to meet those needs, and to accurately predict student performance on the PSSA. This study is designed to determine whether the student JSSA scores are reliable predictors of PSSA scores for several disaggregated groups of these students. An entire class cohort of these students will be followed from their fourth through their seventh grade CBA reading measures; and these assessments will then be compared to the same students' eighth-grade PSSA reading scores.

CHAPTER II  
LITERATURE REVIEW

Reading Assessment

This chapter describes the advantages of curriculum-based assessments when compared to traditional assessments. Curriculum-based assessments are evaluated for relative usefulness across subgroups and for the ability to discriminate progress in academic curricula. Then, curriculum-based assessment and measurement are related to federal accountability measures and high-stakes testing.

CBA supports emerging readers as they develop through these stages. Well organized, prepared, expert teachers explicitly direct children to attend to skills, to practice skills, and to master those skills in order to achieve fluent reading processes and comprehension (Snow, Burns, & Griffin, 1998). Those fluent readers then demonstrate accuracy, appropriate speed and pacing, and suitable expression when reading (The National Reading Panel, 2001). Further, fluent readers then demonstrate understanding of what they read by recalling the information, using and transferring the information, questioning and summarizing the understandings gained. Some students however, continue to fail to read fluently and to understand.

As had been traditional in Greek culture, oral exams were sometimes used to monitor reading skill and fluency during the early years of education in this country. Horace Mann, as the Secretary of Education for the Commonwealth of Massachusetts in 1845, recommended that written exams replace these traditional oral exams in order to provide uniformity (Joshi, 1995). E. L. Thorndike, in about 1910, began to design these standardized assessment instruments; and by 1922, at least 8-10 instruments had been

created for this purpose (Joshi, 1995). Curriculum-based assessments and measurement provide a blending of the authentic oral assessments and the more reliable written assessments. In every decade, these assessments have identified students who are challenged when reading text necessary for civic life and participation.

### Social Goals and Supports for Reading Education

As our society increasingly demands literacy skills for all citizens, the consequences for students who fail to understand what is read become ever more dire and enduring. Additionally, “reading comprehension in a literate society is the key to cognitive development” (Wick, 1987, p. 92). Reading comprehension is the gateway through which each educated student must pass. We attempt to define the student’s instructional level with reading assessments, and to move the student toward incrementally more difficult passages, in order to facilitate reading comprehension. A variety of reading strategies have been employed to support and improve reading comprehension, including self-monitoring for accuracy, direct and explicit instruction, phonics- and phonemics-based strategies, monitoring of pace, and tracking with a finger, with greater relative improvements occurring with different interventions for different students (Allinder, et al., 2001). Still, reading difficulty has been the overwhelming reason for children to be provided special education or to continue at grade level, particularly for children from Hispanic and African-American homes (Commission on Reading, 1985; Learning First Alliance, 1998).

Our national goal then becomes producing students who read and can comprehend at a basic level or better, with increasingly challenging content. Instruction and assessment for the building of reading fluency and comprehension skills is then and

now the focus of reading and other academic classes (Campbell, Donahue, Reese, & Phillips, 1996; Learning First Alliance, 1998).

Corrective feedback during these reading instructions and assessments has further improved students' comprehension for students with reading disabilities (Pany & McCoy, 1988). However, "many experiences contribute to reading development without being prerequisite to it; and although there are many prerequisites, none by itself is considered sufficient" (Snow, et al., 1998, p. 3). A need exists to determine which strategies and experiences yield success with individuals and sub-groups of students. Consequently, the selection of strategies, and the measurement of effectiveness across many strategies demands a simple and unified measure of reading fluency and comprehension.

Lyon, who coordinated the National Institute of Child Health and Human Development (NICHD) longitudinal reading research, followed more than 10,000 children with reading disabilities. The 15 research centers and approximately 100 researchers involved in this study recommend that an integrated curriculum, using explicit and direct instruction, should be supplemented with extensive skills practice (Mercer, Campbell, Miller, Mercer, & Lane, 2000). Specifically, this study recommends phonics instruction, sight-recognition word instruction, and repeated story readings, combined with integrated assessments, in order to build reading fluency and comprehension.

However, like silent reading, reading comprehension cannot be assessed directly using criterion measures. In a complete, traditional assessment of reading comprehension, the measured student response includes literal comprehension, inferential comprehension, and critical comprehension. As each student's personal schemata and experience filters

the skills used to comprehend text, the skills used are variable (Cooper, 1986). Various curricula included vocabulary, grammar, dictionary skills, narrative detail, cause and effect, sequencing, and critical reading within those skills measured through CBA (Fuchs, Deno, & Miller, 1982; Fuchs & Deno, 1991a; Greenwood & Carter, 1987; Greenwood, Terry, Marquis, & Walker, 1994).

The various types of curriculum-based assessment of reading comprehension measure a variety of skills focused on measures of text recall, and move with students through a developmental progression of fluency indicators. Initially, readers are assessed at a literal level, with personal experience being assayed by specific questioning. In responding to a reading prompt, the students incorporate personal experience and make judgments about the passage, as well as discerning the literal meaning of that passage. Emergent readers respond at the levels of knowledge and comprehension. Later, more experienced readers can be assessed at an inferential and critical level, at which the student's experience, reasoning, cognition, and literary skill contributed. These responses from experienced readers require analysis, synthesis and evaluation, and are not pure measures of reading comprehension. Finally, according to Salvia and Hughes (1989), while the effective assessment of reading comprehension can minimize the contribution of associated skills such as the understanding of language structure, memory, reasoning and literary skill, the level of reading comprehension cannot surpass the student's general language competence.

## Authentic Assessment

One must learn by doing the thing; for though you

Think you know it, you have no certainty until you try.

-Sophocles

Both Atwell (1991) and Harp (1994) related this quote within their remarks on instruction and evaluation, for the instructional activity itself is often its very measure. The need to assess student learning within instruction is not a controversial topic, however, the types and methods of such assessment have been passionately defended and radically challenged throughout the 20<sup>th</sup> century (Harp, 1994; Nevin & Paolucci-Whitcomb, 1998).

## Characteristics and Advantages of Curriculum-Based Assessments

Salvia and Ysseldyke (1985) discuss the continuum of educational assessment, including screening, eligibility, program planning, progress monitoring, and program evaluation. Subjective grading policies and methods, lack of inter-rater reliability, poor validity, and limited usefulness are among the usual criticisms of early teacher-made tests. In reaction, standardized testing, perhaps in response to the work of Thorndike and his students, grew in availability through the early and middle 1900s until more than 60 million standardized tests were available in 1944 (Harp, 1994). These standardized tests and the teacher-made tests differ in statistical reliability and validity, but share a common weakness. Because no standard metric, no shared instrument or system, was used across this assessment continuum, the decision-making data are rendered uneven in units of measurement and diverse in applicability. Standardized assessment, whether related to the student's performance within a particular group of students (norm-referenced tests) or

measuring the student's performance in a narrowly defined area of learning or knowledge (criterion-referenced tests), never exactly measures the progress of that student through the student's own curriculum. Additionally, misaligned instruction and assessment confuse instructional focus and minimize teacher effectiveness (Cohen, 1987).

In response to this flawed assessment system, Marston, in Shinn's 1989 text, reviews the increased usefulness of curriculum-based assessment, when compared to individual, published tests. Published achievement tests often fail to match or sample the students' curricula, may not accurately sequence the skills taught and learned, or may use vocabulary better matched to another specific curriculum. Individual, standardized tests have weaknesses in technical adequacy; and have not provided valid, legally defensible decision-making data (Marston, 1989). Further, instructional planning demands data concerning specific skills: error analysis, fluency or mastery, and comprehension, none of which are readily addressed with standardized measures. Another difficulty which encourages curriculum-based methods is the scores typically produced with standardized measures. Standard scores, age- and grade-equivalents, and percentiles are commonly employed to describe student performance on these tests. The age- and grade-equivalent measures simply are not equal measures: their statistical and technical properties are weak, their relationship to a particular curriculum is at best indirect, their meaning is misunderstood, and their usefulness in instructional decision-making is often negligible. Percentile ranks and standardized scores, while technically more reliable, only provide information regarding relative ranking within a normative group. According to Marston (1989) neither of these measures relates the specific skills which a particular student has

or has not mastered. So, how must a serious educator elicit research-based data, which is useful for instructional planning?

Curriculum-based assessments have proven valuable in measuring student progress. Also, they measure these ongoing processes and developing skills at a depth not tapped by the traditional, teacher-created and standardized assessments (Neill & Medina, 1989; Shapiro, 1990). Harvard Project Zero, a long-term investigation of modes of assessment, examines this need for curricular progress monitoring in urban schools during a collaborative project between the Pittsburgh public schools and the Educational Testing Service. Project Zero, which initially focused upon student performance in nontraditional subject areas (i.e., visual arts, music, writing tasks), results in recommendations for project-based learning and supports both curricular performance and portfolio type assessments (Perrone, 1991). The use of these assessments encourages expanded opportunities for students to initiate, create, and participate in their own learning activities and projects.

Project Zero illustrates the importance of student participation and choice in their educational activities and processes. As the goal of public education is to produce active, responsible learners and consumers, we needed to capture and encourage the students' structured thoughtful processes as well as evaluate the products produced (Perrone, 1991). These curricular, process-based assessments provide a means to log the development of the students' metacognitive abilities and understandings, as well as the students' developing strengths and educational needs (Perrone, 1991).

In authentic testing, each assessment measures the student's readiness, progress, and mastery of that particular curriculum, focusing the placement of the student and the

provision of instruction on that student's actual progress. Within this type of assessment, students are encouraged to link their background experience to the curriculum, and to solve problems, express ideas, and to focus on meaning rather than conventions (Chittenden, 1991).

Corroborating evidence from Salvia and Hughes (1990) suggests that student involvement in examining and measuring their own learning, and in developing an understanding of their individual learning process, increases both the students' metacognitive knowledge and level of personal responsibility. From planning to performance, to reflection and the taking of perspective, the authentic assessment literature provides theory and recommends practice to support active student learning (Fuchs, et al., 1982, 1984; Fuchs & Fuchs, 1986). However, standardized assessment, whether related to the student's performance within a particular group of students (norm-referenced tests) or measuring the student's performance in a narrowly defined area of learning or knowledge (criterion-referenced tests), never intentionally measures the progress of that student through the student's own curriculum.

#### Curriculum-Based Assessment and Reading Performance

Perrone (1991) discusses three purposes for the assessment of reading: keeping track, checking up, and finding out (Perrone, 1991). Student instruction and the assessment of that student's progress requires a regular record of skills, a less frequent assessment of what has been previously learned, as well as analysis of the student's particular skills and deficits. A curriculum which responded to local children in a defined environment requires assessment to be intimately linked to the instruction process and academic ecology (Shapiro, 1988, 1989, 1990). Assessment which is linked to and

congruent with the instructional process will be ongoing, will demonstrate transfer of both information and skills, and will serve these three instructional purposes. Further, CBA nurtures complex understandings about both the curriculum and the student as a learner. Reading students will both concretely understand the skills needed and see their own growing abilities charted in CBA (Enggren & Kovaleski, 1996). This complex understanding of a subject and the increased metacognition about the student's own learning can develop and document students' evolving habits of mind (Perrone, 1991).

Curriculum-based assessment when employed in the classroom is defined by Deno as any set of measurement procedures which uses “. . . direct observation and recording of a student's performance in the local curriculum as a basis for gathering information to make instructional decisions” (Deno, 1987, p. 41). Deno's seminal research using CBA which is entitled data-based program modification (Deno & Mirkin, 1977), fuels intensive research into program alternatives and serves diverse instructional needs. Research by Shinn (1989), Hintze and Shapiro (1997), Shapiro (1992) and Fuchs and Fuchs (1992), all attest to the effectiveness of the CBA model in assessment and planning within a reading curriculum. Additionally, Snow, et al. (1998) recommends the development of curricular assessments with the sensitivity to identify young children at risk for reading failure, using day-to-day progress monitoring tasks.

Fuchs and Deno (1994) examine the relative merits and difficulties which result from the use of prompts drawn from the actual curriculum. Face validity and familiarity with the curricular materials often contribute to teacher comfort and to a sense of participation in the whole assessment and instructional process. However, a high variability in instructional or readability level within any curricular text, combined with

the difficulty of obtaining unfamiliar, “cold read” materials may limit the usefulness of the actual curriculum as a fluency measure (Fuchs & Deno, 1994). Further, the vocabulary and difficulty limits of a particular curriculum necessarily limit the ability to generalize decisions about student instruction and performance to other curricula and skill areas.

Curriculum-based measurement resolves these difficulties while allowing educational teams to determine what quality of performance, at what frequency or rate, will meet the appropriate standard for a particular group of students. The educational teams consider the previous performance of these students in the curriculum, the students’ performance on key skills on a readability-leveled probe, and the progress made by the student’s peer group over time, as well as expert opinion on expected student growth (Deno, 1992; Deno, Mirkin, & Chiang, 1982; Deno, Marston, & Tindal, 1985; Deno, Marston, Mirkin, Lowry, Sindelar, & Jenkins, 1982). Standards of performance which reflect a chain of skills can be evaluated against a low criterion; as the students are instructed on these chained skills during a long-term instructional plan (Salvia & Hughes, 1990).

Those educators who apply academic standards to assessment created a comprehensive system to support effective teaching and learning (Reeves, 2000). The clearly defined educational objectives and strategic action planning toward achieving them are implemented by various theorists, using differing educational models. Shapiro (1990, 1992), in examining various integrated CBA models, focuses on classroom environment, instructional-level placement within the curriculum, and student progress within the curriculum. Instructional environment, instructional (level) match, and

instructional accommodations are examined for their effect on the central criteria of student academic progress. All three factors appear to be significant determinants of CBA effectiveness for planning, decision-making, screening, and prediction activities.

In a review extending reading assessment research to students with disability, Christenson, Ysseldyke, and Thurlow (1984) and Ysseldyke and Christenson (1987a, 1987b) identify similar and additional instructional factors which are essential when instructing students with mild disability. These factors include: effective classroom management; positive school environment; a match between instructional level and curriculum; clear and explicit expectations for student performance; instructional performance monitoring and instructional adjustment; sufficient and efficient instructional time; high opportunity for student response; active monitoring for understanding; and, frequent and appropriate performance evaluation (Christenson, et al., 1989). Instructional match, appropriate goals and expectations, instructional support, and monitoring and evaluation are all addressed directly through the use of CBA.

According to Salvia and Ysseldyke (1985), criterion-referenced reading assessments are helpful in determining reading strengths and weaknesses and in designing well-fitted interventions for particular students. However, authentic assessments can also highlight common problems in groups of children; and so help to focus and best capitalize on classroom instructional time. Finally, the use of assessments based on the curriculum best relates to the specific skills and understandings which a student has captured and personalized.

Reeves in 2000 and in his July 2003 leadership presentation, extends this description, telling teachers to “weed the garden,” eliminating any classroom events,

activities, or curricula which do not yield instructional gain. He recommends focused, dedicated instructional time which closes the knowing-and-doing gap for children.

### Curriculum-Based Assessment Process

As early as 1984, Fuchs, Tindal, and Deno examine the effects of consistent procedures when using repeated measurements for educational decision making. The technical adequacy of the data is suspect when procedures are variable or poorly defined during teacher training. This curriculum-based teacher training needs to be delivered in staged staffing groups in after-school or other teacher training sessions, or one-to-one as part of teacher instructional support. Shapiro (1990) suggests that an integrated CBA model would begin by evaluating the instructional environment, to determine strengths and needs of the particular learning community or ecology (Shapiro & Lentz, 1985, 1986); while Gickling would begin with a student skill assessment, including a measured ratio of knowns-to-unknowns and rates of acquisition and retention (Burns, 2001; Burns, Tucker, Frame, Foley, & Hauser, 2001). Secondly, a grade-based instructional level would be determined for the student within the general curriculum (Deno & Mirkin, 1977; Shapiro & Lentz, 1985, 1986). Baseline data typically includes a ratio of knowns-to-unknowns, rates of acquisition and retention (Burns, 2001; Burns, Tucker, Frame, Foley, & Hauser, 2001) and information regarding the student's prior knowledge of concepts, content, and vocabulary (Rosenfield & Kuralt, 1994). At this point, Curriculum-Based Assessment-ID (CBA-ID), as described by Gickling and associates, is employed (Gickling & Havertape, 1981; Gickling & Thompson, 1985). Gickling's CBA procedure has demonstrated high test-retest reliability ( $r = .90-.96$ ) as a measure of student instructional level (Burns, et al., 2001). Lastly, the CBM model can be used to

evaluate instructional progress on an ongoing basis, including the use of comparison, disaggregated data and local aggregated measure norming, as well as individual long-term goal setting (Deno, 1985, 1986; Shinn, 1988, 1989a, 1998).

Enhancing the uses of CBM, Fuchs, et al. (1991) provides post-training support, in the teachers' classrooms, for 20 to 40 minutes every week or two. During that collaborative examination of graphs, goals, and software use, the staff assists teachers to problem solve and to use CBM rules. It should be noted that previous studies clearly defined the significant change factor as CBM, rather than simply the staff supportive visit (Fuchs, et al., 1989b).

#### Student Disability and Reading Performance

The continuously evolving American education system struggles to bridge an escalating dichotomy between more diversely prepared students and increasingly well-defined standards of academic performance. One significant attempt to meet this challenging dichotomous union is curriculum-based assessment. Curriculum-based assessment addresses one of the educational and political objectives of our time, that of successfully including high-, average-, and low-achieving students in regular education, as well as accommodating that regular education to those students with special needs and disabilities.

Fuchs, et al. (1993) examines the expected progress in reading for regular and special education students. When measured using oral reading fluency, the regular education students demonstrate a negatively decelerating curve. They make their highest progress in first grade (about two words per week), decreasing to just below one word per week (0.85 words per week) in fourth grade. By grades five and six, the students progress

about 0.5 words per minute per week. This decelerating curve is not produced, however, when a maze task is implemented at the same grade levels. For special education students, who typically begin at a lower baseline fluency measure, steeper aim slopes are recommended, indicating learning at or above the rate of regular education peers. It is suggested that the annual goal for grade level peers be examined; and that a steeper aim line or increased positive slope be chosen, in order to decrease the gap in performance for students over a year of special education.

Further, those students with disabilities surpass the mean growth of 6/10 of their peers in regular education when instructional recommendations are provided to their regular educators, as part of the CBM process, although their progress without the instructional supports is reliably poorer than that of their peers (Fuchs, Fuchs, Hamlett, Phillips, & Bentz, 1994). Germann and Tindal (1985a, 1985b) suggest that both the absolute improvement data and the relative (to regular education peers) improvement data be used for each special education student in the development of their annual Individual Educational Plan.

#### Socioeconomic Status and Ethnicity Factors as Predictors of Student Reading Performance

In examining progress across economic disadvantage and ethnic minority status, the Milwaukee Public Schools (MPS) used curriculum-based data to achieve membership in the 90/90/90 schools. In these schools, more than 90% of students received free and reduced lunches; more than 90% of students were members of ethnic minorities; and more than 90% of students read at or above grade level. Standardized test results were combined with proficiency data in order to show both student and school building or

district strengths and weaknesses. Goals were set based on current instructional levels and comprehensive analysis of student data. Collaborative teamwork encouraged meaningful mission statements and empowered individual teachers. The schools remained open for evening tutoring, grouped students strategically, and offered summer school and other supportive services to families within the neighborhood schools (Schmoker, 2001).

When educators must provide instruction and monitor progress for widely varying students, an analysis of students' instructional levels, frequent probes and instructional decisions based on data allows the educators to accurately measure and plan instruction and to involve students in their own progress monitoring (King, 1994; King-Sears, Burgess, & Lawson, 1999; King, 1994). Applying curriculum-based assessment and measurement to a variety of students allows evaluators and educators to determine which students require increased differentiation or degree of support in instructional interventions. Those students can then receive the specific strategic, data-based instruction which is most likely to increase their acquisition of skills or educational progress (Howell, 1986, 1993).

Students who receive free and reduced lunches in schools have been designated as a specific subgroup, denoted as economically disadvantaged, in the *No Child Left Behind* legislation (U. S. Department of Education, 2002). For students from environments of economic poverty or cultural diversity, assessment using CBM measures is thought to be protective against false assumptions based on intelligence scores (Baker, O'Neil, & Linn, 1993; Jitendra & Kameenui, 1993). Information garnered from these intelligence and

other assessments far removed from the curriculum was also described as of limited generalizability (Jitendra & Kameenui, 1993).

### Reading Disability and Sex of the Student

Many samples of student reading performance have indicated a higher percentage of males identified as having specific learning disability in reading. Snow, Burns, and Griffin (1998) review the ratios typically reported as varying from 2:1 to 5:1 or greater. Other recent studies find less distinction between male and female scores, sometimes nearly erasing the gender reading performance gap (Flynn & Rahbar, 1994).

### Familial Factors in Reading Difficulties

Family indices of reading disability measure higher for parents of reading disabled children than for the general population. In Scarborough's (1998) review of 516 families across eight studies, the parental reading disability measure varies from 25% to 60%, with 37% as a median measure (Snow, Burns, & Griffin, 1998). This relationship is more powerful for fathers (46%) than it is for the mothers (33%) of the children with reading difficulties. Family and community factors such as economic disadvantage, low achieving schools, limited English skills, or limited interactive language have sometimes been related to poor reading ability. Yet in each case, some schools and some students demonstrate well-developed skills despite the combination of these factors.

Twin studies with reading disabled participants have indicated significant heritability for word recognition and spelling, although that is not true for reading comprehension (Aaron, 1995; DeFries, Fulker, & LaBuda, 1987). As Lyon has noted that word recognition may be the single best indicator of reading disability (1995), this

indicates that word recognition is a skill predictor of reading comprehension with a potential for prevention and amelioration of instructional deficit and for planning remediation.

Although a specific subgroup of students may be identified as being at a greater risk for reading difficulty, that risk does not determine whether a specific intervention might delimit the gap in achievement (Snow, et al., 1998). Indeed, even the identification of a specific skill deficit does not always indicate that addressing that specific skill will narrow the discrepant performance (Adams, 1990). Additional student instructional or ecological variables may later be identified to better explain the underachievement in reading (Felton, 1992). Although the etiology of reading difficulty remains complex, educators remain focused on preventing reading difficulty and improving reading skills.

### Web-Based Assessment Systems

Two well-known web-based reading assessment systems are the Dynamic Indicators of Basic of Early Literacy Skills (DIBELS) and the AIMSWeb sites. Each provides standardized assessment directions, prompts, and benchmark scores to evaluate student progress.

The DIBELS are a set of standardized, individually administered measures of literacy development. A backward extension of CBM, they are designed to be brief fluency measures of one-minute each, used to regularly monitor the development of pre-reading early through elementary reading skills. A creation of the University of Oregon team of Good and Kaminski, DIBELS, “includes descriptions and tutorials on each of the measures, technical reports, logistical information on implementing DIBELS in a school,

and contact information for trainers. The measures themselves are available free to download from the website and use (<http://dibels.uoregon.edu/measures/materials.php>).”

DIBELS CBM measures are specifically designed to assess three of the five Big Ideas of early literacy: phonological awareness, alphabetic principle, and fluency in text. In grades four through six, oral reading fluency and retell fluency measures are provided, as well as the accompanying benchmark standards for beginning of the year, mid-year, and end-of-year assessments. These benchmark standards are available on the DIBELS web-site for either three or four annual assessments in grades kindergarten through six. A recently developed proprietary data system allows DIBELS users to record scores on the measures and to access automatically-formatted reports based upon this data.

Similarly, the AIMSWeb is a web-based proprietary site which offers CBM reading prompts for grades kindergarten through eight, math prompts for grades kindergarten through six, and accompanying progress monitoring tools for elementary through high school students with intensive need. CBM prompts for written expression and spelling are available through this site as well. The AIMSWeb includes tools for benchmarking and Response to Intervention (RTI) as well as progress monitoring tools.

Developed from research and application by Shinn and colleagues, the AIMSWeb is a “scientifically based, formative assessment system that ‘informs’ the teaching and learning process by providing continuous student performance data and reporting improvement to parents, teachers, and administrators to enable evidence-based evaluation and data-driven instruction (<http://www.aimsweb.com/>).” For reading, with a one-year data base of reading scores, students scoring over a provided AIMSWeb target score have a high probability to score as proficient on high-stakes tests of reading.

## Curriculum-Based Assessment of Early Literacy Skills

Johnson and Allington (1991), when reviewing the range of alternative accommodations for young students with reading problems, conclude that remedial reading interventions have not been very effective in increasing early literacy skills. Rather, a variety of researchers had conclude that prevention, through the provision of explicit early literacy instruction, holds the best hope and statistical likelihood, of creating and nurturing proficient readers (Good & Kaminsky, 1996; Lundberg, et al, 1988; Masland & Masland, 1988). Others extend this thinking to posit that first grade success has depended on the pre-reading, early literacy knowledge base, which the student carries as she/he entered the first grade classroom (Adams, 1990; Anderson, et al., 1985).

The difficulty in applying this knowledge for educators is twofold. First, one must determine the specific skills needed for first grade success, and secondly, there must be data-based research on effective practices for teaching, managing records, and adapting programs (Kaminsky & Good, 1998). Masland and Masland (1988) determine that educators are uncertain which skills were pertinent; although others have defined those prerequisite skills (Byre & Fielding-Barnsley, 1991; Iverson & Tunmer, 1993; Koehler, 1996; Simmons, et al, 1995).

Adams (1990) summates the results of United States Department of Education research on needed early literacy skills. She concludes that three crucial skills were prerequisite to reading success: phonological awareness; language skills; and, print awareness. Phonological awareness appeared to be critical for children's ability to develop the alphabetic principle, and to become literate. However, neither district

professional development programs nor early reading curricula typically prioritize the implementation of this important cognitive research (Moats, 1995; Simmons, et al., 1995, 1998; Torgeson & Bryant, 1994).

An attempt to conceptualize this promising practice was Curriculum-Based Developmental Assessment (Bagnato & Neisworth, 1991). This developmental CBA includes all the age-linked landmarks or standards common to developmental research, as well as early literacy markers. Early concepts of print, language skills, and activities to build phonological awareness are assessed during purposeful activity and are related to mastery skills.

This phonological awareness refers to the student's specific and distinct awareness of the sound structure of language, including letter-sound correspondence or phonemic awareness, the manipulation of sound units (e.g., blends), rhyming, segmentation and syllable-definition, and deletion. Explicit instruction in phonemic awareness mediates differences in socio-economic status (SES), encouraging *Success for All* students (Hecht, Burgess, Torgesen, Wagner, & Rashotte, 1995). When preschoolers receive specific instruction in phonological awareness, they experience increased chances of success in first grade reading (Blachman, 1994; Iverson & Tunmer, 1993; Lundberg, Frost, & Peterson, 1988).

Phonics, a concept which refers to letter-sound correspondences and the implementation of those understandings to letter units within written language, is a separate and later sequenced skill in learning to read. Those curricula which begin by addressing phonics rather than with explicit phonological awareness instruction, are

unlikely to successfully reach and teach the largest percentage of preschoolers (Byrne & Fielding-Barnsley, 1991, 1993, 1995; Torgeson, et al., 1994).

When Stein, et al. (1999) analyzes beginning reading programs to determine whether explicit phonics instruction occurs, only two of the reviewed curricula contain this explicit instruction. Open Court, an SRA/McGraw-Hill product, and Scholastic's Phonics Readers meet this criteria; and each, additionally, relates phonics instruction to the vocabulary used within the reading text at a high potential accuracy rate: 50% and 32%, respectively (Stein, et al., 1999, p. 281).

Even so, these generally successful, phonologically-enriched curricula have not sufficed for some individual children (Blachman, 1994; Byrne & Fielding-Barnsley, 1991, 1995; Meisels & Provence, 1989; Torgeson, Wagner, & Rashotte, 1994). Torgeson et al. (1994) describes a 30% proportion of his at-risk sample who did not successfully complete the eight-week training program with significant phonological awareness growth, despite the success of similar peers. However, O'Connor, Jenkins, Leicester, and Slocum (1993) examine the instruction of young children with learning disabilities, using phonological awareness extensions and interventions, with increased success over traditional curricula.

Meta-analyses of students who are at risk for reading difficulty have been numerous (Catts, et al., 1997; Rodgers, 1983; Rutter & Yule, 1975; Shaywitz, et al., 1990, 1992; Silva, et al., 1985; Snow, Burns, & Griffin, 1998). Further, 1994-1995 statistics from the United States Department of Education study *The 18<sup>th</sup> Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act* indicates that of the 4.43% of students categorized as having Specific Learning

Disabilities, an estimated 80% of those students qualify as having specific reading disability. This 3.54% of the student population is augmented by an additional proportion of students with reading difficulty, who do not qualify as having specific learning disability. These students read less fluently than do their peers, but demonstrate less than a 2.0 standard deviation discrepancy between their actual reading skills and those predicted based on their intelligence quotient. In all, Snow, et al. (1998) finds that 14.8% of students in grades three and four, and 14% of students from grades seven through eleven underachieve in reading using this definition.

While investigating the risk of reading difficulty, Stage, et al., in a 2001 study using Native American, Hispanic, and European-American students, indicate that reading growth is not differentiated for their kindergarten population based on ethnicity, for students with the following diagnoses:

- severe cognitive disabilities (including low IQ and early receptive and expressive language disorder);
- hearing impairment and chronic otitis media (Wallace & Hooper, 1997);
- specific early language impairment (although some 40-75% of these students later qualify with reading and/or achievement difficulties [Bashir & Scavuzzo, 1992]); and,
- Attention Deficit Hyperactivity Disorder (with a 31% chance of reading difficulty in first grade to a 50% chance in ninth grade [Shaywitz, et al., 1994, 1995]).

Additional predictors of reading difficulty, useful for screening at school entry, include a variety of memory and language skills. Table 1 adapted by Snow, et al. (1998) from Scarborough (1998) addresses these skills and the strength of their relationship to later

Table 1

*Measures which Predict Reading Difficulties at School Entry*

Factors Identified	Sample Numbers	Strength of Relationship
Language		
Verbal memory for stories/sentences	11	Median $r = .49$ mean $r = .45$ ( $SD = .14$ )
		Lexical skills
1. Receptive vocabulary	20	Median $r = .33$ mean $r = .36$ ( $SD = .17$ )
2. Confrontation naming	5	Median $r = .49$ mean $r = .45$ ( $SD = .07$ )
3. Rapid serial naming	14	Median $r = .40$ mean $r = .38$ ( $SD = .09$ )
Receptive language, syntax/morphology	9	Median $r = .38$ mean $r = .37$ ( $SD = na$ )
Expressive language	11	Median $r = .37$ Mean $r = .32$ ( $SD = .16$ )
Overall language	4	Median $r = .47$ Mean $r = .46$ ( $SD = .15$ )
Phonological awareness	27	Median $r = .42$ Mean $r = .46$ ( $SD = .13$ )
Early Literacy-Related Skills		
Reading 'readiness'	21	Median $r = .56$ Mean $r = .57$ ( $SD = .12$ )
Letter identification	24	Median $r = .53$ Mean $r = .52$ ( $SD = .14$ )
Concepts of print	7	Median $r = .49$ Mean $r = .46$ ( $SD = .20$ )

*Note.* (Snow, Burns, & Griffin, 1998, p. 110).

reading difficulties. The studies reviewed contain sample sizes of 30 or more and each assesses at least one risk factor within one year of the participants' entrance to school (Snow, Burns, & Griffin, 1998).

Confrontation naming, that is the rapid naming of vocabulary items, has nearly matched the variance measured by receptive vocabulary and IQ measures, when used to predict later reading skills (Snow, et al., 1998). Additionally, rapid serial naming has been used as an estimate of current and future reading skill (Ackerman, et al., 1990; Bowers & Swanson, 1991; Cornwall, 1992; Wolf & Obregon, 1992). It is noted that these expressive language assessments may be better related to the word recognition and decoding skills which support early readers than to the comprehension of complex language in later grades (Snow, et al., 1998).

When multiple criteria are used to predict students at risk for reading difficulty, combined measures of phonological awareness and rapid serial naming allow an 83% accurate hit rate for Catt's (1991, 1993) longitudinal study. Bishop and Adams (1990) use language ability measures and IQ to predict reading difficulty slightly less accurately. Many local districts incorporate Concepts of Print and story retelling in their kindergarten screenings, to good effect (Ficco, 1999).

Twin studies with reading disabled participants indicate significant heritability for word recognition and spelling, although that is not true for reading comprehension (Aaron, 1995; DeFries, Fulker, & LaBuda, 1987). Although Reid Lyon (1995) has noted that word recognition may be the single best indicator of reading disability, the Aaron and DeFries, et al. studies indicate that reading comprehension is a skill predictor of word recognition with a potential for instructional deficit and remediation.

## Reading Curricular Design

Reading curriculum, at its most powerful, interweaves subject content and learning elements or skills, in order to build interconnected understanding. This interconnected understanding would be enhanced by the presentation, refinement, and reinforcement of learning skills over a wide range of situations or applications. As process and content goals become unified, assessment becomes integral to both. Jacobs (1989) acknowledges this concept as a metacurriculum, which consists of learning strategies and skills serving two purposes. The strategies and skills both develop the students' abilities to independently think and learn, and facilitate the students acquiring mastery of the curriculum (Jacobs, 1989). These thinking and learning skills are comprised of: decision-making; problem-solving; conceptual and symbolic thinking skills; systems related thinking; as well as the more familiar skills, including categorization and the formation of causal hypotheses (Jacobs, 1989).

Caine and Caine (1997) relate perceptual organization development to increasing classroom autonomy, self-efficacy, and to providing a classroom ecology less bound by power and authority than is commonly the case. In order to facilitate maximum learning, meaningful tasks, high challenge, and low threat or power conditions are applied (Caine & Caine, 1997).

## Data-Based Reading Supports

In the effort to focus on what works, Farone (2001) published on the University of Oregon web site, the Dynamic Indicators of Basic Early Literacy Skills (5<sup>th</sup> ed.) (DIBELS-5) (Good & Kaminsky, 1996, 1998, 2002; Kaminsky & Good, 1996, 1998), to address the big ideas in curricular assessment of young students. Within DIBELS,

fluency measures range from letter-sound match to phoneme segmentation to oral reading fluency, as the student progresses from kindergarten through third grade. Additionally, the published web site addresses instructional techniques, data-based curricula, and developmental expectations for students at specific grade levels.

A further enhancement in reading success is demonstrated when data-based assessment and decision-making models are attached to the previously described phonological curricula (Kaminsky & Good, 1998). When student progress is assessed, using a CBM problem-solving model, educators can evaluate success for individual students, apply needed accommodations and interventions, and determine effective instruction within a short, several week period (Deno, 1985, 1986, 1989), thus increasing individual student progress in reading (Fuchs & Fuchs, 1986). Fuchs and Fuchs (1986) measure this formative evaluation, reinforcement and decision-rule model effect at +0.70, for reading results.

Alternative reading curricula use these brain-based learning theories to help create and support resilient students, who maintain positive attitudes; who quickly recover from emotional reaction; and who work hard, whatever their innate abilities (Caine & Caine, 1997). Further, these curricula tend to have five characteristics in common: Teachers know their field and can apply that knowledge; students' personal meanings and purposes are respected; students experience possibilities within a sense of personal mastery; process is viewed as dynamic; and, an interconnectedness of subjects and skills relates curriculum and lifework (Caine & Caine, 1997).

Hintze and Shapiro (1997) discuss the wide variety of nontraditional reading curricula available to bridge basal readers and whole-language or literature-based series.

These curricula contain main selections, projects or activities, and classroom-based measures, which attempt to measure the breadth and depth of student reading mastery, as demonstrated in classroom samples (Hintze & Shapiro, 1997). These samples of reading behavior are stored in portfolios, which become a longitudinal record of student progress. Hintze and Shapiro review a 1992 study by Fuchs and Deno (1992a, 1992b), where the utility and validity of CBM is examined across traditional and nontraditional reading series. The results of the Fuchs and Deno study suggest that CBM remains a valid criterion, across those different curricula. Progress rates across grades show similar patterns for the two types of curricula as well. To seek these progress rates, Gickling would begin with a student skill assessment, including ratio of knowns to unknowns and rates of acquisition and retention (Burns, 2001; Burns, Tucker, Frame, Foley, & Hauser, 2001). Gickling's above described CBA procedure has demonstrated high test-retest reliability ( $r = .90-.96$ ) as a measure of student instructional level (Burns, et al., 2001).

Fuchs, Fuchs, and Maxwell (1988) demonstrated that oral reading fluency was a valid measure of reading comprehension. In that study, fluency scores are related to scores on: two subtests of the Stanford Achievement Test, measures of reading comprehension, Cloze procedure comprehension measures, question answering, and oral and written recall measures. Fluency scores were most highly related to the criterion measures (.89), with written recall being the second most highly related (.66-.82).

### Curriculum-Based Measurement

#### as a Discriminant Measure

CBM was created as a standardized and specific CBA measurement procedure by Stanley L. Deno and Phyllis Mirkin at the University of Minnesota (Shinn, Nolet, &

Knutson, 1994). A decision-making model, CBM allows educators and psychologists to identify academic problems; discriminate or certify the differences in individual and peers' performance; and explore and evaluate academic accommodations and modifications (Deno, 1989).

Curriculum-based measurement measures student curricula through producing a brief, repeatable, inexpensive measure of student fluency which is sensitive to incremental student progress. Specific methods for CBM reading assessment include counting one minute samples of words read correctly from written prompts or word lists, completing mazes and cloze procedures, and various other responses to reading prompts. An alternative source probe, using the basal reading curriculum text, was found to provide accurate instructional fluency data when literature-based curricula were used for instruction (Bradley-Klug, Shapiro, Lutz, & DuPaul, 1998; Shinn, 1988, 1989a, 1989b). During a CBM reading assessment, the fluency measure is repeated, usually three times, using separate, unfamiliar prompts or lists. The reported figure is the median score of the three fluency measures. According to Shinn (1989), these observed measures have two advantages: the reading process is observable, and creative responses can be assessed accurately. Possible disadvantages are the increased time required to assess each individual, rather than simultaneously assessing an entire group; and scoring reliability problems with untrained staff, although the scoring process is also standardized.

Oral fluency as a measure of reading comprehension has roots in statistical and theoretical research, rather than being simply intuitive. Both the oral reading fluency and maze task fluency processes are found to be reliable at greater than .80 for both alternate forms and test-retest reliability (Shin, Deno, & Espin, 2000). Correlation research yields

high reciprocal relationships (greater than .85) between the oral reading fluency measure and standardized assessments of reading comprehension (Allinder, Fuchs, & Fuchs, 1998; Marston, 1989; Tindal & Marston, 1990).

With fluency used as the discriminant reading measure, CBM was applied by Deno, Marston, Shinn, and Tindal (1983) to discriminate students with learning disability (LD) from Chapter or Title I students and from regular education students, in grades one through three. Marston, Tindal, and Deno (1983) determine that CBM can discriminate LD classification as well as the aptitude-achievement discrepancy formula. Shinn and Marston (1985) use words read aloud to discriminate students in regular education or Chapter I, and students with mild LD. In 1986, Marston, Fuchs, and Deno demonstrated that CBM fluency measures short-term reading progress across 10 and 16 week periods as well as standardized reading tests, with the additional benefits of showing greater growth and better matching teacher perceptions of improvement.

## Technical Features of Curriculum-Based Measurement

### *Psychometrics of Curriculum-Based Measures*

Fuchs and Fuchs (1992b) review the validity and reliability of CBM as a reading skill measure. They differentiate all other curriculum-based assessment as being reliant on teacher-made or chosen tests or prompts, which have unknown technical features. In CBM, the reading fluency measure has proven strong criterion-based validity, stability, inter-scorer validity, sensitivity to developmental growth rates and discriminative validity (Fuchs & Deno, 1992b; Fuchs & Fuchs, 1992b; Good & Jefferson, 1998). Fuchs and Fuchs (1992b) use an alternative, computer-based CBM system for scoring and monitoring progress, both for time efficiency and to address the issue of the measurement

of reading comprehension. These alternative measures include question answering tests, recall procedures, cloze techniques, and maze procedures. The procedures involve comprehension questions with their differently valued answers; questions about the facts of the story; fill-in-the-blank type questions regarding the prompt; or the use of mazes. Criterion validity for these four alternate measures indicate that cloze and retell procedures are inadequate measures of student progress, while the maze procedure was promising, with technical features close to those of reading fluency (Fuchs & Fuchs, 1992). They also suggest that standard error of measurement reduction would clarify and better identify true student progress, as measured by slope. One successful method of reducing this error is aggregating student data over multiple occasions and prompts (Fuchs, Deno, & Marston, 1983).

An interesting rival hypothesis to the interpretation of recent CBM construct validity data is proposed by Kranzler, Brownell, and Miller (1998). The authors examine the relative significance of several potential predictors of reading comprehension: general cognitive ability (*g*), oral reading fluency, and processing speed and efficiency. In this study, oral reading fluency ( $r = .17$ ) described less of the variability in reading comprehension than did psychometric *g* ( $r = .24$ ). After controlling for general cognitive ability and processing speed, oral reading fluency explained 11% of the variability in reading comprehension. Data from this study also suggests that the same variance in reading comprehension is measured by *g* and processing speed and efficiency. The study concludes that CBM oral reading probes are quite stable across different probes and across trained examiners.

CBM reliability has also been examined repeatedly by teams of researchers. Shinn (1989) summarizes these studies. Shinn (1981) and Marston (1982) use test-retest methods or parallel forms used one week apart, with demonstrated results from .82 to .96. Germann and Tindal, (1985b) use regular education students in grades three and four to demonstrate test-retest reliability of .97, and parallel forms results of .94. Tindal, Marston, and Deno (1983) demonstrate reliability of .89 to .99, with randomly selected first through sixth graders, using test-retest, alternate forms and inter-judge agreement.

*Curriculum-Based Measurement: Expanded Practices*

Fuchs (1998) notes that CBM data can be used both to examine and contrast total scores and to analyze patterns of skill for a particular student. Teachers can use CBM data to analyze the student skills on an item-by-item basis, determining which skills are mastered or emergent, and guiding instruction. Additionally, the use of varied curricula and varied difficulty levels has not limited the validity of the CBM oral reading metric as a survey-level assessment (Hintze, Shapiro, Conte, & Basile, 1997).

Screening students on above- or below-grade level curriculum with CBM prompts has provided a guide for establishing correct instructional levels and for use in multiple-source eligibility determination (Powell-Smith & Stewart, 1998). When continuous progress monitoring data is used within a problem-solving process, the school evaluation team can more readily establish the students' Least Restrictive Environment (LRE), that placement where the child make satisfactory progress with the greatest inclusion within regular education settings and with regular education peers (IDEA, 1991).

At the University of Oregon, Shinn and colleagues devised the Responsible Reintegration of Academically Competent Students (RRACS) system, in order to identify

students who could be reintegrated into the regular education classroom, and to continue to monitor the progress of those students (Shinn, Habedank, Rodden-Nord, & Knutson, 1993). In contrast, Fuchs and colleagues support full inclusion for students with disabilities, using a full spectrum of interventions and accommodations including out-of-grade level placements (Fuchs, Dempsey, Roberts, & Kintsch, 1995; Fuchs & Fuchs, 1990, 1994; Fuchs, Fuchs, & Fernstrom, 1992, 1993). While both approaches are CBM data-based, Shinn's ReRACS approach includes an ecological, team-based problem-solving protocol, which allows comparison of students with mild disability to the academic performance of their peers in the regular education classroom, within the same curriculum (Shinn et al., 1993).

*Curriculum-Based Measurement and  
Differentiated Instruction*

Fuchs, et al. (1991) has noted that teachers using CBM tend to adjust their students' goals more frequently and for a greater percentage of their students. The Fuchs posit that this more frequent goal and program adjustment is the significant factor which differentiates student success with CBM (Fuchs, et al., 1991, 1992). Adaptive teacher style, which monitors type of instruction as well as content level, is a marker for research-based effective instructional design. Shinn (1989) also review the benefits of repeated measurement, noting that frequent, direct performance data could guide instructional decision making across a school year.

Fuchs (1998) also describes difficulties which might limit teacher use of CBM. Teachers are often pressed for time to analyze student performance in a more than rudimentary way. Neither do the teachers welcome a task which demands extensive

preliminary learning and which provides limited or unreliable results. It is posited that the time consuming task of error analysis, which provides instructional focus, could be completed through a computer program. This computer-assisted analysis provides reliable instructional data, potentially encouraging teachers to use CBM more reliably and accurately (Fuchs & Fuchs, 1987).

### *Technological Supports for Differentiating Instruction*

In 1987-1988, software was developed to accomplish the above described task analysis in reading (Fuchs, Fuchs, & Hamlett, 1989). The skills analysis provides student performance data over two week periods, with a description of skills attained in skill clusters within the general curriculum. These descriptions range from skills not attempted, to skills not yet attained, partially mastered, probably mastered and mastered (Fuchs, 1988). This scoring system provides a reliable way of scoring student performance, with minimal teacher time consumed and graphical feedback for formative decision-making.

Similarly, the Shinn CBM software site, the AIMSweb, provides progress monitoring and general outcome measures for reading, with graphical displays and textual description of student results (Shinn & Shinn, 2002). Student results are stored and displayed on the AIMSweb over multiple school years, with choices of frequency of measurement. This system meets the requirements of a comprehensive progress monitoring system: it is useful for screening, monitoring, and reintegration data.

Technology further assists teachers to differentiate instruction through the use of text-reading software which includes word prediction components, auditory feedback, tools for dictionary and thesaurus support, and visual structuring (Castellani & Jeffs,

2001). The pre-reading, post-reading, and motivational activities and supports which the technologies provide allow students with disabilities to access more complex and challenging text (Behrmann, 1998; Higgins & Boone, 1997). When these supports are provided with clear goals, active participation, focused monitoring of academic progress, and data-based decision-making, academic outcomes are promising (Christenson, Ysseldyke, & Thurlow, 1989).

Further research addresses the results in student performance when CBM is employed within instruction. When special educators are used to further examine the benefits of computer-assisted skills analysis, robust findings recommend using computer programs for data analysis as well as data management. Teachers who use data-based formative planning which includes skills analysis are able to design better fitted instruction; and their students demonstrate greater achievement than students who are instructed without the use of CBM (Fuchs et al., 1988, 1989a, 1989b; Walton, 1986).

#### *Curriculum-Based Measurement and Curriculum*

As recent PDE monitoring and best-practice indicate, educating students in the LRE must be considered in any educational placement (Powell-Smith & Stewart, 1998). The special education system has been regularly criticized for a lack of data use in decision-making when discrepancy, interventions, placement and reintegration are to be considered.

A white paper issued in 1986 by Madeline Will, then U. S. Assistant Secretary of Education, addresses the need to educate all students within the general curriculum, and to titrate specially designed modifications within that curriculum to support success for diverse learners. Additionally, the intent of the Individuals with Disabilities Education

Act (IDEA, 1991) has been to include all children in regular education to the maximum extent possible.

CBM data has provided a structure for evaluating the instruction, placement and reintegration of students of diverse ability. The usefulness of CBM in the above actions and processes has been addressed earlier. However, when only 2%-8% of students with disability are reintegrated in any school year (Lyle & Penn, 1986; Shinn, 1986), the CBM protocol for addressing reintegration must be considered. That protocol has included recommendations from the students' special educators and examination of the students' CBM data by a team consisting of the student, parents, general and special educators (Powell-Smith & Stewart, 1998). Trials of reintegration are then supported through regular CBM progress monitoring and peer group CBM comparisons. In monitoring student progress after reintegration, Shinn, Powell-Smith, and Good (1997) find that students with disability who enter the reintegration process at below the reading skill level of their Low Reading Peers (LRP) show faster progress than their LRP after four weeks of integration. Additionally, after eight weeks it is generally agreed that at least half of those students have benefited from the reintegration (Smith & Good, 1997). For the remainder of the students for whom results were more mixed, the most telling feature is the students' initial reading skill level, as their progress slopes are often equivalent to their LRP (Smith & Good, 1997).

Challenges for those reintegrated students include increased large group instruction and decreased academic encouragement (Powell-Smith & Habedank-Stewart, 1998). Progress monitoring is recommended as a support for responsible reintegration, as

students with disability are by definition individual in their profiles of difficulties and success (Powell-Smith & Stewart, 1997).

### *Curriculum-Based Measurement and Students with Disability*

While students with disability are regularly monitored for academic progress and annual plans are written to set goals for the following instruction, Shinn (1986) posits that much of the information used in this process is inaccurate or inappropriate. In his 1986 article *Does anyone care what happens after the refer-test-place sequence: The systematic evaluation of special education program effectiveness*, he states that historical or personal characteristic information is viable for research only. He recommends a data-based test-intervention-test-change the instruction model, in which the student's response to data-based intervention is the criteria for educational decision-making. This criterion is met by CBM progress monitoring. The National Association of School Psychologists (1994) corroborate this need to determine response to intervention for both instructional decision-making and for placement decisions.

An additional advantage of regular CBM practice is the opportunity to more accurately assess and identify students with learning disability. Ysseldyke, Algozzine, and Epps (1983) use 17 different definitional criteria for learning disabilities found in the literature, to examine regular education students. Of these regular educational students, at least 88% were identified as learning disabled using at least one of those definitions. When using the significant discrepancy formula to define a student as learning disabled (a 10 point discrepancy between Wechsler Intelligence Scale for Children-Revised and Woodcock-Johnson Achievement or Peabody Individual Achievement Test (PIAT) scores, 65% of regular education students meet the criterion (Ysseldyke, et al., 1983).

Further, Snow, et al. (1998) reports that poor readers within their classrooms and students identified with a specific reading disability demonstrate the same deficits and errors as well as suffering “the same kinds of negative educational and occupational outcomes” (p. 95). Data-based assessment of reading difficulty is effectively applied across the differing ability spectrum.

Various strategies have been employed to specifically address the limited reading fluency of students identified with specific learning disabilities. Allinder, et al. (1998, 2001) has examined various assessment methods and reading strategies employed to increase both fluency and reading comprehension. In the 2001 middle school study, Allinder reviews the successful use of concrete presentation and explicit instruction of specific reading strategies. Those successful strategies include pre-reading activities, pre-teaching of unfamiliar vocabulary, repeated readings, establishing personal connections, modeling fluency, extended practice on whole texts, and daily review of the literature (Allinder, et al, 2001).

In a similar process, Schmoker (2001) describes the Brazosport Independent School District’s use of standards-based reading assessment data to limit the effects of demographic disparity. The district implemented explicitly assessed standards aligned to the reading standards, along with remedial programs focused on those standards which were most difficult to reach. Results indicate that students with disability and economic disadvantage then demonstrate a much reduced gap in performance on the reading assessment.

This recommendation to focus on specific skills, rather than on disability category is shared by Shinn (1987). Shinn and others propose that the identification of students

with disability should be based upon the students' achievement and mastery skills within the regular curriculum (Shinn, et al., 1987). Further, Shinn demonstrates that CBM reliably differentiates students with disability from regular education students, using students identified by traditional standardized measures (Shinn, et al., 1988). Four additional studies have examined the usefulness of CBM in differentiating regular education and special education students (Deno, et al., 1983; Shinn, et al., 1986, 1987; Shinn & Marston, 1985). Consistently in each study, students receiving Title I (or Chapter I) services score within the bottom quartile of regular education students, while students identified with a learning disability score below the fifth percentile when compared with the reading achievement scores of regular education students. Using CBM, Shinn examines effect sizes for these differences and finds reliable differentiation of regular education students, low-achieving students, and students with disability (Marston, 1989).

Students who evidence above- or below-grade-level skills could be assessed using out-of-grade-level curricular assessment (CBM), which is variously named instructional assessment (Gickling & Thompson, 1985; Gickling & Rosenfield, 1995) or survey-level assessment (Marston & Magnusson, 1985, 1988). This above- or below-grade-level assessment allows eligibility, progress monitoring, and curricular placement decisions to be made, based strictly upon student performance data. Planning and evaluation of student progress can also produce disaggregated data for students in regular and special education programs, in order to determine whether students placed in special education begin to show a reduced skill discrepancy after receiving specially designed instruction.

Two types of goals have been selected by curriculum-based research teams (Dunn & Eckert, 2002). Short-term goals involve breaking the desired skill into sequential or hierarchical sub-skills and monitoring student progress through that sequence, with the mastery of each sub-skill being prerequisite to teaching the next, more difficult skill. Long-term goals, in contrast, established an overall reading goal, such as words-correct-per-minute to be reached within a specified time period. Meta analysis by Fuchs and Fuchs (1986b) indicates that the use of long-term goals is associated with greater progress in reading performance, as measured by standardized testing.

Specifically, student reading progress on CBM measures is measured using a linear model, and tends to increase more steeply during the early years of reading acquisition than it does in later elementary grades. Shinn and Deno recommend an appropriate goal of 1-2 words per week increase during the early elementary years, with 1.25-1.50 words per week after the rate of 30 words per minute is achieved (Deno, Fuchs, Marston & Shin, 2001; Shinn, 1989).

Wesson, a professor at the University of Wisconsin-Milwaukee, designed additional CBM procedures to increase user efficiency and implementation rates (1987). Developing a routine for random selection of passages, with 10-page samples from the beginning, middle and end of each text, allows student assessments to be chosen and initialed or recorded at each evaluation. Each text passage is pre-counted for words-per-line and this number recorded at the end of each line on only the teacher or assessor's copy. A measurement station within the classroom and a color-coded graphic organizer is suggested, to facilitate measurements. With limited instructional time, the use of trained aides and volunteers are suggested, as well as having a student read into a tape recorder,

to administer reading probes. Further, these students are able score and graph their own results, thus increasing motivation and saving teacher time (Wesson, 1987). Computer software and lists of potential instructional strategies are resources also discussed by Wesson in order to facilitate the implementation of CBM.

Yet, research continues to show that educators use CBM irregularly or inconsistently (Allinder & BeckBest, 1995). Educators often list time constraints as the reason for this irregular or inaccurate usage (Cooke, Heward, Test, Spooner, & Curson, 1991; Wesson, 1990; Wesson, King & Deno, 1984; Wesson, Skiba, Sevcik, King & Deno, 1984). The provision of limited time requirements with CBM procedures has not always improved implementation rates (Fuchs & Fuchs, 1989; Fuchs, Fuchs, Hamlett, & Stecker, 1991). Allinder and BeckBest (1995) compare teacher use of a CBM system with self-monitoring, versus CBM with expert consultant support. Results of that study indicated that teachers “appeared unwilling or unable” to use CBM consistently to monitor instructional progress, even with follow-up consultation. (Allinder & BeckBest, 1995, p. 296).

### *Reading Development Across Grade Levels*

In elementary school, curriculum-based measurement is employed as an indicator or a monitor of performance in the four basic skill areas: reading, mathematics, writing and spelling (Deno, 1985; Espin & Tindal, 1998). The content of the elementary curriculum deviates little from those areas and problem-solving focuses on academic progress and instructional changes within those discrete areas (Fuchs, 1989; Fuchs & Fuchs, 1996).

Young pre-readers in kindergarten and first grade are often monitored using the DIBELS to measure phonemic segmentation, onset recognition, and letter naming skills (Good & Kaminsky, 2002; Kaminsky & Good, 1997, 1998). The DIBELS skills are a backward extension of CBM with the named measures being sensitive to early literacy skill growth, and are helpful in focusing instruction on children at risk to be poor readers.

Juel (1988) notes that poor readers in first grade have a .88 probability of remaining poor readers when measured as fourth graders. While Good and Kaminsky (1996) identify kindergarten and the primary years as the greatest opportunity to increase the progress slopes of children, others extend this research to show that relative reading skill levels remained stable over the elementary school years (Jorm, Share, MacLean, & Matthews, 1986). The low-stakes instructional adaptations elicited by DIBELS and CBM are especially appropriate to younger children, whose performance is more frequently variable (Kaminsky & Good, 1997); while intermediate elementary students are measured more reliably over the short-term. Additionally, the longer duration of the CBM materials and process allow its use in high-stakes placement decisions for those older children (Fuchs & Deno, 1981; Shinn, 1989).

Later, as students enter secondary schools, the purposes for CBM become more complex, addressing both the four basic skill areas as well as specific content knowledge. This dual focus is paralleled in the purposes of high stakes testing, which address both basic skills and the skill area requirements of secondary schools, each directly mirroring successful content mastery (Schumaker & Deschler, 1988). While CBM progress monitoring serves to identify areas of difficulty (Deno & Fuchs, 1987), the use of broader CBM measures of actual products serves to measure mastery of the content area (Espin &

Tindal, 1998; Tindal & Parker, 1989). These broader CBM measures include total production, percentages of correct items, and mean lengths of continuously correct items (Tindal & Parker, 1989; Videen, Deno, & Marston, 1982). Tindal and Parker (1989) describe percentages correct as having especially strong holistic and predictive scores. In order to successfully graduate with a diploma, these secondary students are required to demonstrate both fluency and content mastery.

### *Reading Level and Race*

Reading progress for students of varied races remains relatively constant from 1980 until 1996, according to the National Assessment of Educational Progress (NAEP, 1997). Stage, et al., in their 2001 study including Native American, Hispanic and European American student participants, indicate that reading growth is not differentiated for their kindergarten population based on race. Most recent NAEP reporting (1996) indicates that the percentages of fourth, eighth and twelfth graders performing at or above the required scale scores are significantly smaller for students of Black or Hispanic ethnicity. Additionally, the NAEP data indicates elevated levels of basic and below basic performance for fourth grade children who are Black and Hispanic.

When two of every five Black and Hispanic students have been placed or held at least one year below the expected grade for their age, (Alsalam, et al., 1992) and the gap between the performance of White students and minority students has remained unchanged over almost thirty years (NAEP, 1991), performance monitoring of all students has become necessary. Reportedly, the mean scores of Black children have been uniformly poorer than the general population measures for both intelligence and academic skills (Humphreys, 1988). Additionally, high drop-out rates for minority youth

have ranged from 33%-70% (Arias, 1986; Office of Educational Research and Improvement, 1991). A greater percentage of youth of minority backgrounds are placed in special education than would be expected; and Black children are more likely than White children to have been labeled as mentally retarded (Reschley & Wilson, 1990).

#### Curriculum-Based Measurement: Basic and Advanced Uses

CBM provides basic information regarding student progress, which can be clearly depicted in graphs. The concept of time-series analysis or the measurement of student progress using slope or rate of improvement is central to the initial use of CBM data. Teachers and parents review the effectiveness of instructional methodology over time, shifting methods or curricula when the slope flattens or declines and continuing while progress is optimum. This data serves to evaluate the individually changing conditions, needs, and response of a particular student. Marston (1989) describes the benefits of time series analysis as including timely decision making and individualization of instruction. Additionally, Shinn (1998) asserts that formative evaluation tools should be standardized, allowing change data to better reflect student progress, rather than differing test materials and conditions. CBM provides the base data for time series analysis, a visual analysis of the relationship between the learning interventions and student progress data, over time.

#### Curriculum-Based Measurement and Ethnicity

In Table 2, the NAEP provides a national assessment of student performance and educational achievement in United States elementary and secondary schools through review of the reading achievement of male and female students as well as students of three ethnic groups. The student measures were earned in grades 4, 8, and 12, and are categorized at four scaled score ranges.

Table 2

*NAEP Long-Term Reading Results: Male, Female, Black, Hispanic and White Students**Scoring at or above Scaled Scores of 150/200/250/300*

Students	SS $\geq$ 150	SS $\geq$ 200	SS $\geq$ 250	SS $\geq$ 300
At grades 4, 8 and 12				
Male	91.6	58.3	14.1	0.3
	99.3	89.2	53.4	10.0
	99.9	96.4	77.0	33.6
Female	95.3	69.9	19.2	1.1
	99.8	94.9	66.2	16.8
	100.0	98.6	86.8	45.4
Black	84.5	41.8	6.0	0.0*
	99.4	81.5	34.0	3.0
	99.8	94.9	67.9	18.5
Hispanic	85.8	48.0	7.1	0.1*
	98.6	84.9	38.1	5.4
	99.9	94.1	65.0	20.5
White	96.4	71.4	20.4	1.0*
	99.7	95.5	68.7	16.9
	100.0	98.5	87.0	46.0

*Note.* (NAEP, 1996). \*Sample sizes were too small for reliable estimation.

## Reading Progress for Students with Reading Difficulty

Correlations between reading fluency, reading task performance, scores on standardized achievement tests and Grade Point Average (GPA) are stronger for low-achieving than for high-achieving students (Espin & Deno, 1993). These low achieving students demonstrate larger differences in task performance when they increase reading skills, while higher-achieving students improved more slowly, based on a differentiated influence of reading strategies and background knowledge in the specific content area.

Initial research by Shinn (1989), examining the use of various instructionally leveled probes (one, two, or four years advanced, and one year below instructional level), proposes that curricular samples at each instructional level can reasonably be used to determine progress slopes. Further study by Hintze, et al. (1998) examines the use of similar and challenging instructional level curricular probes. While Hintze's progress slopes based on similar instructional level material are found to be greater than those progress slopes which were based on challenging material for grades one and two; no significant differences in slope are determined for those students in grades three and four (Hintze, et al., 1998).

However, when comparing student reading progress in material at their current grade-level (similar material) versus evaluating their progress in challenging material, taken from curricula one-year in advance of grade-level, Dunn and Eckert suggest that slope is not the most accurate method for student progress monitoring (2002). Rather, the visual examination of data points, checking for stability in progress and relatively high, correctly read words-per-minute is presented as more accurate (Shinn, Gleason & Tindal, 1989). Additionally, Eckert and Dunn (2002) posit that,

despite research advising one-minute probes (Fuchs, 1989), more frequent probes or probes of three minutes duration may improve accuracy.

Bender (2002) reviews research on student supports using continuous progress monitoring as including the following central axioms:

- Academic assessments must be based on academic skills,;
- Regular and frequent assessments are necessary for effective instructional planning; and,
- Educational decision-making should be data-based.

Curriculum-based assessment and measurement blur the distinction between instruction and assessment, providing instead a system of seamless data-based instruction applicable across curricula.

### Applying Reading Research in the Curriculum

#### *Using Alternate Curriculum in Curriculum-Based Measures*

Hintze and Shapiro (1997) discuss the wide variety of nontraditional reading curricula available to bridge basal readers and whole-language or literature-based series. These curricula contained main selections, projects or activities, and classroom-based measures which attempt to measure the breadth and depth of student reading mastery, as demonstrated in classroom samples (Hintze & Shapiro, 1997). These samples of reading behavior are stored in portfolios, which become a longitudinal record of student progress. Hintze and Shapiro review a 1992 study by Fuchs and Deno, where the utility and validity of CBM is examined, across traditional and nontraditional reading series. The results of the Fuchs and Deno study suggest that CBM remains a valid criterion, across

those different curricula. Progress rates across grades show similar patterns for the two types of curricula, as well (Fuchs & Deno, 1992).

A question first raised in the 1992 Fuchs and Deno study is explored by Hintze, Shapiro, and Lutz (1994). These researchers use selective probes from one traditional, basal curriculum and one nontraditional, literature-based curriculum as CBM passages to monitor the students' reading progression. The hypothesis that student progress in both curricula can be monitored, using CBM probes from only the traditional, basal curriculum, is evaluated. Results demonstrate better sensitivity to student progress using the traditional basal curricular probes, whether the students had received traditional or nontraditional curricular instruction. However, the strongest correlation is indicated when measuring student growth in the traditional series using the traditional probe (Hintze & Shapiro, 1997). Questions for further examination include the replication of this research across grade levels (third grade participants were initially used) and the impact of increased vocabulary in the nontraditional series.

#### Accountability Measures and Reading Assessment

The choice of sensitive and accurate measures has remained dependent on federal and Pennsylvania's state level decisions regarding the PSSA assessment. Initially conceived of as a simple rubric, the state scoring of PSSA prompts floated free of any normative balances until 2001. From 1986 until 2001, the earned scores were simply divided into quartiles, with a true mean determined by all scores, including unusual outliers. Nor was there uniform participation in the PSSA across school districts and disaggregated groups of students from 1986 until 2001, with widely divergent

participation of students with disabilities, students who were highly mobile, and students who spoke English as a second language (PDE, 2002).

Many systemic and ecological factors have plagued the standardized implementation and limited the initial validity of PSSA results across Pennsylvania. In the early years of PSSA testing, many students were excused or removed from the assessment, often those students with known disability. Participation rates were more irregular in some districts than in others, and no state level accountability records yet existed to require and document high standards in participation rates (PDE, 2002). Some districts used the alternative assessment, the Pennsylvania Alternative System of Assessment (PASA) at a much higher rate than the 1% of student population which was recommended by the PDE (2002). Other districts had a much higher rate of drop-outs who were not assessed by the PSSA than was common across Pennsylvania. Even in 2003, no code yet existed for students who have been at risk because of teen pregnancy, who have been supported by Student Assistance Teams, who have been or are currently being treated for drug or alcohol abuse, or for those students with consistently high mobility rates across their instructional years.

Although state level supports have been applied to districts whose performance indicates need, those supports have not yet been factored into the normative performance of students across Pennsylvania. It would seem to be evident that performance of both students and districts could be sorted both by supports provided and by student performance, in order to determine whether supports are effective change agents for particular subgroups of students and districts.

*Curricular Data as a Predictor of Pennsylvania System of School Assessment  
Performance*

In response to the timely, existing reading research and to a systemic need for monitoring and improving reading performance, one district in Western Pennsylvania, the GJSD, attempted to study, elicit ideas, and goal-set, with groups of teachers, parents, and community-school board members. Beginning in 1996, the district selected and wrote a series of developmental reading prompts, which would directly assess student learning in the classroom three times annually (Ficco, 1999). Each student would be assessed, the reading group placement of the students would be determined and reviewed based upon curriculum-based assessment results, and an extended instructional spectrum within the reading curriculum could more often meet diverse student needs in a noncategorical setting.

Curriculum-based assessments are seen as preferable to PNAT for the district assessment program for multiple reasons. Although PNATs are accurate, reliable screening tools and measures of important reading skills (Salvia & Yseldike, 1985, 1995; Shinn & Bamonto, 1998), they are not as sensitive or useful in measuring small, incremental progress in reading (Howell, Fox, & Morehead, 1993; Marston, 1988, 1989). Because curriculum-based assessments often differ on their match between teaching curriculum and testing prompts (Mehrens & Clarizio, 1993; Shapiro & Derr, 1990), the school team chose samples from the alternate, *Houghton Mifflin* curriculum which met the standards previously discussed and served as “cold read,” unfamiliar prompts. It is hoped that these purposeful prompt selections from the basal reading curriculum will measure the students’ achievement of reading proficiency toward the annual goal.

Further, the CBA datum are intended to be used as part of a problem-solving model, to direct student learning regarding the Pennsylvania benchmarks for reading at grade level. The setting of ambitious progress goals, with the intention of reaching benchmark reading proficiency, aims at narrowing performance discrepancy for many district students (Fuchs, 1993).

It is additionally important that these CBA measures be accurately related to the Pennsylvania Standards of Education, as assessed by the PSSA. It is desirable to know if such student performance at particular developmental levels will most accurately predict an individual student's performance on the PSSA reading assessment. In light of the direct-instruction time and resources expended and the great need for this diverse population to build skill, it is critically important to determine whether reading assessment at specific elementary and middle school grades would better predict those students' later scores on the PSSA reading assessment (McGlinchey & Hixson, 2004). The JSSA was intended to serve as a general outcomes measure in reading (Reschly & Tilly, 1993), and those educational skill assessments are designed to relate directly to the tasks and skills needed for student PSSA proficiency.

Additionally, determining whether results differ by demographic characteristics of the students such as disability or sex of the student is thought to be valuable. Finally, determining whether the reading curricula and professional education afford a differentiated level of success on these measures is a focus of the analysis.

### Summary

American education, as it continues to evolve, has struggled to bridge the dichotomy of increasingly diverse students and increasingly well-defined standards of

academic performance. A significant attempt to meet this dichotomy is the use of curriculum-based assessment and curriculum-based measurement. Curriculum-based assessment and measurement are concepts which fit the educational and political objective of our times, specifically that of successfully educating high-, average-, and low-achieving students, as well as those students with diverse needs.

The psychometric properties of curriculum-based measures, the standardization of their implementation, and the ability to apply CBA across a broad base of students with diverse learning needs recommend their regular use in American education. As instructional time is prioritized and strategic instruction is designed, CBA measures provide a brief, reliable measure of developing student skills which support those practices. When curriculum-based assessment and instruction are reviewed and designed in a regular cycle, the differentiation of instruction to that population of diverse learners becomes a natural concomitant of the team-based review of that data.

Curriculum-based measurement supports data-based, effective reading instruction, relates reliably to other measures of reading proficiency, and can be usefully employed in establishing local norms. Finally, the curriculum-based measurement process may provide reliable progress data to relate to the high stakes testing which has been employed as an educational accountability measure across the state of Pennsylvania.

CHAPTER III  
METHODS AND PROCEDURES

Method

*Introduction*

The Pennsylvania System of School Assessment (PSSA) reading comprehension test, a state-wide standards-driven assessment of reading skill is now required of all Pennsylvania students in the third, fourth, fifth, sixth, seventh, eighth, and eleventh grades. It was published in separate formats and with several alternate forms. The PSSA has been scored by the Data Recognition Corporation (DRC), a professional testing service in Minnesota.

Using a similar process, the Greater Johnstown School District (GJSD) has assessed student performance on curricular reading samples, followed by scoring of these samples using the Pennsylvania Department of Education (PDE) designed, standards-based rubric.

Intended purposes for the collected curriculum-based data were instructional progress monitoring and the prediction of student success on the PSSA reading assessment. It was desirable to know if there exist precise points in the developmental assessment of reading skills, from which to most accurately predict an individual student's performance on the PSSA reading assessment. Further, in light of the direct instruction time and resources used, it was considered useful to determine whether each summative district reading assessment at fourth through seventh grades related significantly to the students' eighth grade reading scores on the PSSA. Last, the district and researcher needed to examine whether the district reading curricula, the *Houghton*

*Mifflin* basal reading series and the Johns' Hopkins *Success for All* reading program, yielded greater reading scaled score results on the PSSA reading assessment for disaggregate groups of students within the district.

Student fourth-grade Johnstown System of Student Assessment (JSSA4) through eighth-grade JSSA assessment (JSSA8) scores were available for this research, as were Spring of 2002 eighth-grade PSSA (PSSA8) scores. The collected data were stored in district files, student comprehensive records and reading portfolios, and in district reading specialist files.

## Research Questions and Hypotheses

### *Research Question One*

Did the Johnstown System of Student Assessment for reading demonstrate concurrent validity when correlated to student reading grades in the fourth grade year?

Question one examined the concurrent validity of the district curriculum-based reading assessment when measured against the student classroom reading grades. Research indicated that curriculum-based assessment was an efficient and accurate measure of student reading skill, and it was hypothesized that the district curriculum based assessment would be a valid measure.

### *Research Question Two*

Did the Johnstown System of Student Assessment (JSSA) reading scores predict the Pennsylvania System of School Assessment (PSSA) reading scaled scores?

In question two, the curriculum-based assessment scores in reading (JSSA4-JSSA7) were considered as predictors of the later PSSA8 reading scores for a student

cohort. It was hypothesized that those district CBA scores would be significant predictors for the later PSSA reading scores. Multiple regression analysis was used to evaluate the CBA predictors against the criterion of PSSA8 reading scaled scores.

### *Research Question Three*

Were the demographic factors of sex and educational disability of the student significant predictors of the students' eighth grade PSSA reading scaled scores?

In question three, the demographic characteristics of the student cohort (i.e., sex of the student, categorization as having an educational disability, as well as the specific type of disability for students who received special education) were examined to find any relationship between these characteristics and the students' scaled scores on the PSSA eighth-grade reading exam. It was hypothesized that female students would perform more strongly on the PSSA8 reading assessment than did male students; while students with educational disabilities would perform less well on the PSSA8 reading assessment than students in regular education with no educational disability. Multiple regression analysis was used to evaluate these predictors of the criterion, the PSSA8 reading scaled scores.

### *Research Question Four*

How predictive were the students' JSSA4 through JSSA7 scores of their eighth grade PSSA reading scores, when examining students who participated in either the John Hopkins' *Success for All* reading curriculum or the *Houghton-Mifflin* basal reading curriculum?

Question four examined differentiated curricula as a factor which might influence student reading performance on the eighth-grade PSSA. The two curricula implemented within the district were the *Success for All* reading program published by Johns Hopkins University and the *Houghton-Mifflin* reading basal program. All students registered in each elementary building received one prescribed reading curriculum throughout their elementary program. The research hypothesis was that students participating in the *Success for All* curriculum would better match student needs at the Greater Johnstown School District, yielding better results on the eighth-grade reading PSSA than the results for those students receiving the *Houghton Mifflin* reading basal curriculum. Multiple regression analysis was used to evaluate these curricula as predictors.

### Design

This research study was an exploration of various predictive factors for student performance on the PSSA assessment of reading at the eighth grade. Demographic factors included sex and educational disability, while educational factors included the two reading curricula and the student score categories on a district curriculum-based assessment of reading.

The student and educational data were collected in 2002, although the student curricula and the curriculum-based scores represented the years 1999-2002. All participants were students of the Greater Johnstown School District, and an entire grade level cohort participated in this study (see Figures 1 and 2).

Predictors for the Eighth Grade

PSSA Reading Score



*Figure 1.* Research design diagram for PSSA predictors project.

Research Path Diagram for PSSA Predictors Project

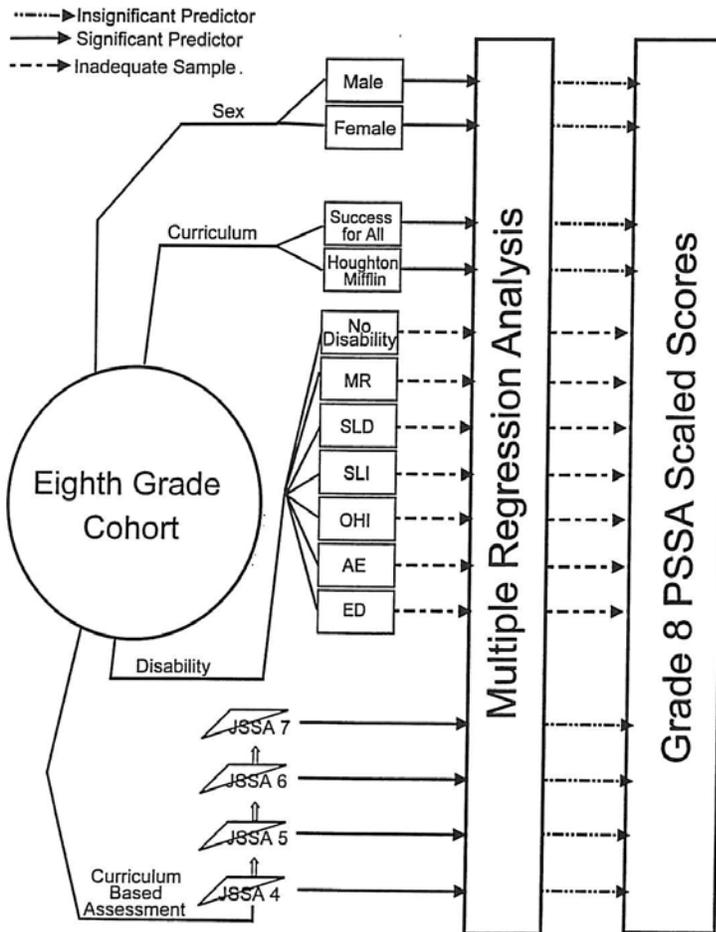


Figure 2. Research path diagram for PSSA predictors project.

## Population

For this study, the participants included two subgroups of students from the GJSD. One part of the total student cohort was composed of students who were educated at the West Side Elementary School using a *Houghton Mifflin* reading curriculum. The second subgroup consisted of students who were educated at the East Side Elementary School using the John Hopkins' *Success for All* reading curriculum. These students' subgroup memberships were determined by the majority of elementary school years spent at either the East Side or the West Side School (greater than 75% of instructional time). All participants in each subgroup attended the named elementary school from four-year Kindergarten through fifth grades, while both subgroups attended the same district middle school during sixth through eighth grades.

## Sample

The entire student cohort included 268 eighth graders who were calculated as 48% male, 22% registered as being an ethnic minority, 21% categorized as having an educational disability. This cohort included all district eighth graders who had been assessed by the district's standards-based district assessment (the JSSA) from grades four through seven. However, the data regarding any students who did not participate in that eighth grade PSSA reading assessment and those assessed while attending other districts were eliminated from the study.

## Assignment

The student participants were not assigned to their curriculum category. Rather, their home school determined which curriculum they received. No obvious assignment occurred, and all students in an entire grade level cohort were participants.

## Measurement

### *Curriculum-Based Assessment*

The JSSA curriculum-based assessment reading prompts for spring assessment were selected and scored by teams as measures of the students' instructional-level reading comprehension skills. The JSSA assessments were modeled after the Pennsylvania state standards-based assessment, the PSSA. Each passage was chosen to limit unique and difficult proper nouns and pictorial or graphic cues, to offer sufficient length for the expected range of fluencies, and to represent grade level text as assessed through readability scores (Ficco, 1999). Passage selection teams were drawn from district reading instructor volunteers; while scoring teams were either trained by the PDE at scoring sessions, or were trained by the GJSD Professional Development Coordinator using the same training protocols.

The validity of the JSSA curriculum-based measures was protected in multiple ways. All prompts were designed to be "cold read" materials, not previously encountered by students in the grade-level curriculum. Additionally, the prompts were stored centrally, with no opportunity for prior exposure of the selections to the teaching staff. The teaching proctors for the group assessments were given standardized directions and scheduling, again by the Professional Development Coordinator (Ficco, 1999).

The CBA testing protocol was evaluated for concurrent validity through correlation with district reading grades at the fourth grade level. If this correlation demonstrated a p-value of high significance (as measured by a Pearson's R or Spearman's Rho) and a check for linearity demonstrated that linearity was apparent, the collection of data would continue.

Possible threats to validity for the CBA data include the sensitivity of the measures, as well as the standardization and implementation of assessment measures. The various demographic disaggregated groups were not independent of each other. The measure may have been more sensitive to multiple skill sets, such as combined reading and writing skills, than to simple reading fluency and comprehension; and individual teachers may not have applied the testing protocol with equal precision.

The reliability of curriculum-based assessments was variously evaluated and measured; however, these CBA assessments were not standardized, nor were they consistent across content areas, grade-levels, or approaches (Gickling & Rosenfeld, 1995; Shapiro, 1990; Shinn & Good, 1993). Even with a standardized approach or procedure, CBA better described a student's performance authentically and qualitatively than it did numerically or statistically (Tindal, Marston, & Deno, 1983; Shinn, 1995). The CBA instrument, the JSSA, had not been evaluated for reliability.

The fidelity of both assessment and scoring was monitored by district administrators and the staff development coordinator (Ficco, 1999). The district scoring staff were trained and re-trained or calibrated in a cycle, allowing for substitute teachers to support the process. As this training was provided by the staff coordinator using a fixed agenda, the training process was very similar from cycle to cycle (Ficco, 1999).

The West Side students received the assessment in their small, mixed-reading-skill level reading classrooms; while the East Side students received the assessments in their small, instructionally-leveled reading skill classrooms. Middle school students were assessed in their mixed skill level language arts classrooms. In each case, the directions were read orally, repeated as needed for individual students; and the assessment was administered in small reading groups. For all students, their written response to the prompt was planned and drafted on the first day, and was prepared as a final copy on the second day. The prompt was available to the students on both days, as was the PDE reading rubric poster which had been taught to the students and was used in classrooms repeatedly prior to the assessment (Ficco, 1999).

#### *Pennsylvania System of School Assessment*

The PSSA, a state wide, standards-driven assessment of reading skill, is required of all Pennsylvania students in the third through eighth and eleventh grades in separate formats and with several alternate forms. As an accountability measure, the PSSA reading assessment is used to evaluate the effectiveness of reading instruction in public schools. The assessment is designed to measure student reading skill at the matriculated grade level, rather than at the instructional level. For disaggregated groups who were instructed above or below their matriculated grade level, this assessment could not measure actual reading skill or progress.

This PSSA assessment has been scored by the Data Recognition Corp, a professional testing service in Minnesota. The PSSA assessment scores across the years 1999 to 2002 are used in this study; and it is important to note that the PSSA did change their scoring parameters in the year 2001.

PSSA statistical equating was prepared beginning with 1996 as a base year; this was an independent process from the cut-scores. The teachers and others involved in the bookmark and borderline group methods were participants in processes which provided estimates of cut-scores. These methods were used to gather information for the policy makers at the PDE to use in their decisions about what the cut-scores for reading and mathematics should be. The State Board also was involved in policy decisions about the cut-scores. They examined the Department recommendations for the cut-scores and approved them. Right after the results from the two processes were averaged, the decision was made by the Zogby administration to raise these scores one quarter of a standard error (Masters, personal communication, December 15, 2004). Then those scores were presented to the State Board and approved. The State Board agreed with the Department's recommendation for the cut-scores, and set these cut-scores for the 2001 assessment. The 2001 to 2002 testing events provided the first opportunity to follow changes in student performance with the cut-scores in place.

Because this process occurred during the years when data were collected for this study, certainly the reliability of those PSSA measures and their relationship to the CBA measure varied significantly during the period from 1999 to 2002.

Additionally, because the PSSA designed process allowed a short time to prepare for a valid match between educational standards and PSSA content, content validity of the measures and this research study data were also compromised. As the PSSA moved toward being a measure of standards, it better measured more narrowly defined academic skills while becoming a poorer measure of broad, higher level thinking, and problem solving skills (Masters, personal communication, December 15, 2004). The 1999

assessment was the first to measure the standards. In the years following 1999, the assessment of the standards became more precise, with better coverage of each standard.

### *Curricula and Related Hypothesis*

It was hypothesized that student participation in the *Success for All* reading curriculum would predict significantly better scores on the PSSA reading assessment at the eighth grade level than would student participation in the *Houghton Mifflin* basal reading program. The following description of those curricula relates the prescribed implementation of those curricula, as well as the relevant research base.

### *Success for All*

The *Success for All* reading program prescribed a dedicated 90 minute period of reading instruction. This instruction was provided to small groups of students who required a homogenous instructional level. Teachers were trained to provide a specific sequence of instruction, with much instructional diversity within the dedicated period (Slavin, & Fashiola, 1998; Slavin, Madden, Dolan, & Wasik, 1996). While the instruction was moderately scripted, teachers were not instructed to use a word-for-word presentation. However, the vocabulary for both instructional routine and behavioral management was prescribed as a part of comprehensive school reform (Borman & Hewes, 2002; Borman, Hewes, Overman, & Brown, 2002; Madden, Slavin, Karweit, Dolan, & Wasik, 1993).

School-wide professional development, which was mandatory for administrators and teachers, was presented as central in comprehensive reform initiatives (Borman, et al., 2002; Ross, Alberg, & McNelis, 1997; Ross, McNelis, Lewis, & Loomis, 1998; Ross,

Smith, & Casey, 1999; U.S. Department of Education, n.d.). *Success for All* staff provided regular fidelity monitoring, sequenced staff development activities, and technical assistance which led to increased rates of reading achievement for students of diverse backgrounds in many urban and diverse educational sites (Smith, Ross, & Casey, 1996). It should be noted, however, that a number of the studies noted only a moderate effect size for students who were characterized as low achievers.

### *Houghton Mifflin Basal Reading*

The *Houghton Mifflin* basal reading curriculum was a commercially produced basal reading program which provided explicit, systematic instruction in phonemic awareness, phonics, fluency, vocabulary, and comprehension. In the program, direct instruction methods were used to teach intensive, systematic, and explicit phonics. Lessons had suggested topics which an expert teacher was able to independently implement. This basal programs included phonological awareness activities, however, the activities did not address dimensions of phonological awareness most highly correlated with early reading acquisition (i.e., blending and segmenting) and did not provide sufficient practice, unique materials, and teacher scaffolding. These phonological awareness instructional procedures failed to integrate critical findings from empirical research relative to addressing the needs of students with diverse learning needs (Smith, Simmons, Gleason, Kame'enui, Baker, Sprick, Gunn, Thomas, Chard, Plasencia-Peinado, & Peinado, 2001). Fluency instruction included decodable text and leveled books with assessment prompts to measure fluency, accuracy and the ability to read with expression. Comprehension instruction addressed monitoring, questioning and

summarizing for narrative and expository text with the use of graphic organizers (Pearson & Raphael, 2003; Reutzel & Cooter, 2003).

There was a very limited research base for the achievement effects of traditional basal reading program. However, *Houghton Mifflin's* program presented minimal instruction and modeling, and then asked students to perform complex summarization tasks (Murray, B., personal interview, November 2004). Further, reading programs for some underachieving students were described as effective only if they were highly structured and systematic (Tunmer & Hoover, 1993).

### Procedures

JSSA reading prompts for spring assessment, which were used in this study, were selected and scored by a team as a measure of the students' instructional-level reading comprehension skills. These selection teams were composed of district reading instructor volunteers, and separate forms were offered for the fall, winter, and spring assessments. The spring assessments were used for this study, were saved in an Excel spread-sheet with the students' sex, building and reading curriculum assignment, reading grades, PSSA reading scores, and possible disability category.

John Hopkins' *Success for All* reading students received the JSSA assessment, a grade-leveled reading passage followed by a short written response, in their smaller than regular classroom, single instructional-level reading groups. The alternative reading curriculum also provided assessment within the smaller than regular classroom reading group, which was of mixed instructional levels. Middle school students were assessed in their mixed ability level reading and language arts classrooms.

For students in each curricula, one day was provided for reading and planning; while a second day was dedicated to preparing final copies of each student's response to the prompt. Sixth and seventh grade students were assessed using the same process in a reading group or language arts classroom.

Scoring of the JSSA occurred the following week, with teacher and student feedback being provided the second week after testing. District scoring teams were trained and recalibrated at the beginning of each scoring session. A portfolio stored all of each student's JSSA scores, three for each grade level, in reading, writing, and math. JSSA scores ranged from 1-4, with 1-4 corresponding to the PDE categories of Below Basic, Basic, Proficient, and Advanced. Students were assessed with the PSSA reading assessment in their reading or language arts classroom groups with the same assessment process and supports provided for both JSSA and PSSA assessments.

#### *Data Collection*

In the district, CBA scores were stored in student portfolios, with reading specialists, and in the offices of district guidance counselors. Similarly, PSSA scores, reading grades, and demographic data were stored in building guidance offices, at the administrative offices, and in computer-based records. The described records were stored at two elementary schools, one middle school, and at the district administrative offices.

All data were encoded anonymously within Excel spread sheets; and data analysis considered that individual data using randomly generated identification. In accordance with federal regulations, this data will be maintained confidentially for three years from the completion of this project.

The PSSA reading comprehension scores were provided to the district late in the summer following assessment in each school year. Stickers coded with the student's scores were placed in each student comprehensive permanent record, while hard copies of the results were sent or mailed to each student home. Administrative data and copies of the PSSA scores were filed and stored in hard copy and on compact discs at the district central administrative offices. The concept map seen in Figure 2 describes the study, followed by details regarding the student participants, data collection, and the research hypotheses (see Table 3).

### Sample Size

An attempt was made to choose an adequate sample size by accessing an entire grade level of students at the local district. Although the entire cohort of 268 students was reviewed, only 238 of those students had recorded or available PSSA8 reading scores. When considering this entire cohort of 268 students, 111 of these students were categorized with a disability or placed within an alternative education site.

When considering the primary question of this study: Did curriculum-based assessment predict the high stakes test result for reading: the addition of variables which define disaggregate groups severely limited the sample size. When the variables included sex of the student, categorization with disability, and PSSA8, only 98 of the 111 students with disability placement could be used in the analysis. Further, when analysis considered the variables of sex of the student, categorization with disability, fourth through eighth grade JSSA categories, and PSSA8 scaled score, 62 students composed the valid cases. A further attempt to examine the questions regarding CBA and PSSA8 including specific disability categories yielded only 19 valid cases.

Table 3

*CBA and PSSA Assessment Project Task*

#	Name	Description	Begin	End	Person
1.	Project Idea	Based upon the school district's need to examine the uses of the CBA reading assessment system.	1999	11-2002	School Psychologist, Professional Development Staff, Superintendent
2.	Refine Study	Review reading assessment and benchmarking best practice. Identify questions.	2001	11-2003	School Psychologist, Administrators, IUP Professors
3.	Obtain Materials	Obtain the student data and scores. Create Excel charts.	8-2002	12-2002	School Psychologist
4.	Statistical Analysis	Meet with staff at IUP.	4-2005	8-2006	School Psychologist, Applied Research Lab Staff, Dr. Barker
5.	Initial Draft of Chapters 1 & 2	Committee review.	9-2006	12-2006	School Psychologist, IUP Professors
6.	Draft of Chapters 3 & 4	Committee review.	9-2007	12-2007	School Psychologist, IUP Professors

Table 3 (Continued)

*CBA and PSSA Assessment Project Task*

#	Name	Description	Begin	End	Person
7.	Draft of Chapter 5	Committee and Chair review.	9-2007	12-2007	School Psychologist, IUP Professors
8.	Final Report Review	Committee review.	9-2007	12-2007	School Psychologist, IUP Professors
9.	Deliver Copies To Graduate School	Deliver copies, pay fees.	11-2007	12-2007	School Psychologist

*Note.* Sample Size: An entire cohort of eighth grade students from a local school district participated in this study. This class number was 268. Statistical Analysis: The study employed correlation and multiple regression analyses.

## Statistical Analyses

### *Questions, Hypotheses, Predictors, Statistical Analyses, and*

### *Statistical Assumption for the PSSA Predictors Project*

The statistical analysis of the data was preceded by completing Pearson and Spearman's correlations between fourth grade reading grades and fourth grade curriculum-based assessment (CBA) scores, which validated the district CBA reading measure as a measure of reading skill. Then, multiple regression was completed to determine the significance of student disability and sex of the student as predictors of the PSSA reading scaled scores for the entire cohort of students. A second multiple regression analysis examined the significance of having an educational disability, sex of the student, and using the CBA scores as predictors of the PSSA reading scaled scores for those same students. The third multiple regression analysis examined the significance of having a specific educational disability, sex of the student, and CBA scores as predictors of the PSSA reading scaled scores for those same students. Finally, in the last multiple regression the predictor of reading curricula was added to all of the previous predictors to determine any significance in predicting the criterion, eighth grade PSSA reading scaled score (see Table 4).

Table 4

*Research Questions, Hypotheses, Predictors, Statistical Analyses, and Statistical Assumptions*

Research Question	Hypotheses	Predictors	Statistic	Assumptions	Assumption Appropriateness
Three Were the demographic factors of the sex and educational disability of students significant predictors of those students' eighth grade PSSA reading Scaled scores?	3a. Females will earn higher scaled scores on the PSSA reading assessment	3a. Predictor: Sex (female, male)	Multiple Regression	1. Interval or ratio data 2. Normality of error terms 3. Normality of predictors 4. Linearity	1. Examine the instruments 2. Plot or examine for linearity 3. Visual inspection of the correlation matrix using "Rules of Thumb"
	3b. Students with an educational disability will earn lower scaled scores on the PSSA reading assessment	3b. Predictor: Disability	Multiple Regression	5. Equal standard deviation terms 6. Predictors not highly correlated	

Table 4 (Continued)

*Research Questions, Hypotheses, Predictors, Statistical Analyses, and Statistical Assumptions*

Research Question	Hypotheses	Predictors	Statistic	Assumptions	Assumption Appropriateness
Four Did the students' JSSA4 through JSSA7 reading scores predict their eighth grade PSSA reading scaled scores when examining students who participated in either the John Hopkins' <i>Success For All</i> reading curriculum or the <i>Houghton-Mifflin</i> basal reading curriculum?	4. Students receiving SFA curriculum will earn higher scaled scores on the PSSA reading assessment	4. Predictor: Reading curriculum (SFA, <i>Houghton-Mifflin</i> )	Multiple Regression	<ol style="list-style-type: none"> <li>1. Interval or ratio data</li> <li>2. Normality of error terms</li> <li>3. Normality of predictors</li> <li>4. Linearity</li> <li>5. Equal standard deviation terms</li> <li>6. Predictors not highly correlated</li> </ol>	<ol style="list-style-type: none"> <li>1. Examine the instruments</li> <li>2. Plot or examine for linearity</li> <li>3. Visual inspection of the correlation matrix using "Rules of Thumb"</li> </ol>

Table 4 (Continued)

*Research Questions, Hypotheses, Predictors, Statistical Analyses, and Statistical Assumptions*

Research Question	Hypotheses	Predictors	Statistic	Assumptions	Assumption Appropriateness
One Did the Johnstown System of Student Assessment for reading demonstrate concurrent validity when correlated to student reading grades in the fourth grade year?	1. CBA reading score categories will be positively related to student grades in reading	1. Factors: Reading grades (GPA) and CBA category score	Correlation	1. Interval or ratio data 2. Normality for each group 3. Linearity 4. CBA is an accurate measure of reading skill 5. Reading grades and the JSSA reading measured similar reading skill	1. Examine the instrument 2. Descriptive statistics 3. “Rules of Thumb”

Table 4 (Continued)

*Research Questions, Hypotheses, Predictors, Statistical Analyses, and Statistical Assumptions*

Research Question	Hypotheses	Predictors	Statistic	Assumptions	Assumption Appropriateness
Two Did the Johnstown System of Student Assessment reading scores predict the Pennsylvania System of School Assessment eighth grade reading scaled scores?	2. CBA reading scores will predict PSSA reading scaled scores	2. Predictor: JSSA, the local CBA reading measure	Multiple Regression	<ol style="list-style-type: none"> <li>1. Interval or ratio data</li> <li>2. Normality of error terms</li> <li>3. Normality of predictors</li> <li>4. Linearity</li> <li>5. Equal standard deviation terms</li> <li>6. Predictors not highly correlated</li> </ol>	<ol style="list-style-type: none"> <li>1. Examine the instruments</li> <li>2. Plot or examine for linearity</li> <li>3. Visual inspection of the correlation matrix using “Rules of Thumb”</li> </ol>

*Note.* The chapter examined the concurrent validity of the district CBA measure of reading; then determined whether sex of the student, student disability, or reading curricula were significant factors which predicted reading scaled scores on the state assessment of reading.

The purpose of multiple regression was to predict a single criterion from one or more significant predictors. Multiple regression with many predictor variables is an extension of linear regression with two predictor variables. The computations are more complex, however, because the interrelationships among all the variables must be taken into account. This particular statistical process determined whether any single predictor was statistically significant and predicted the PSSA scaled scores in reading, once all variables had been considered.

Hierarchical regression added terms to the regression model in stages. At each stage, an additional term or terms was added to the model and the change in R<sup>2</sup> was calculated. Hypothesis tests were done to test whether the change in R<sup>2</sup> was significantly different from zero. The order of variable selection was designed to clarify relevant combinations of significant factors (Wang, 1999), and to determine the significance of individual predictors. This method was chosen rather than backward elimination stepwise regression, as overlapping variables may affect the PSSA-JSSA relationship differently over different steps in the regression. If those variables were eliminated early in the analysis, their later significance might have been missed (Tran, 2007; Wang, 1999). Similarly, variables found to be significant early in the analysis may have later proved statistically insignificant, and might have mistakenly been judged as significant predictors.

These predictors were then related to the criterion, the student scaled scores on the eighth-grade PSSA reading assessment (PSSA8). The researcher hoped to clarify the statistically significant predictors among all of the possible predictors: sex of the student,

educational disability, CBA results in grades four through seven, and the two alternate reading curricula.

### *Research Questions*

1. Did the Johnstown System of Student Assessment for reading demonstrate concurrent validity when correlated to student reading grades in the fourth grade year?

It was hypothesized that the district CBA reading scores would be significantly related to the reading grades for those students. Both Pearson's and Spearman's  $R$  were calculated to determine this correlation.

First, JSSA4 was correlated to the students' individual reading grades during that same year (fourth-grade grade point average--GPA in reading), in order to validate the JSSA as a measure of reading skill. As Creswell (1994) indicated that existing instruments such as the JSSA must be validated against another content measure. The students' reading grades and JSSA scores were compared using correlation measures, including Pearson's and Spearman's Rho.

2. Did the Johnstown System of Student Assessment reading scores predict the Pennsylvania System of School Assessment eighth grade reading scaled scores?

It was hypothesized that the JSSA reading scores would accurately predict the eighth grade PSSA reading scaled scores. The addition of curriculum-based reading score categories sought to determine whether these CBA performances were significant predictors of the PSSA reading scaled scores, which was the second purpose for the CBA process (Fish, 1988; Stevens, 1986). Multiple regression was used to determine the significance of the JSSA predictor.

## Assumptions

Bernhardt (1998) discussed 10 levels of data interaction. This study assumed that an interaction of predictor variables: sex of the student, student disability, reading curriculum, and student JSSA performance over fourth through seventh grades could be compared over time to student performance on the eighth grade PSSA. It was assumed that a linear relationship between JSSA and PSSA performance could be established for each disaggregated group; and that the effects of the demographic factors could be established and observed over a period of four years.

The study further assumed that the JSSA and PSSA8 assessments measured similar reading standards and skills; and that the assessment scores retained a very similar meaning across the years 1999 to 2002. In fact, the Pennsylvania System of School Assessment did change their scoring parameters in the year 2001.

Additionally, the assumption was made that JSSA4 through JSSA7 assessments measured reading skills with the same degree of accuracy and against the same set of PDE grade-level standards. As these GJSD assessment teams maintained the practice of retraining and recalibrating the scorers at each assessment and scoring event, this assumption may be more valid.

3. Were the demographic factors of sex and educational disability of the student significant predictors of those students' eighth grade PSSA reading scaled scores?

It was hypothesized that the disaggregated group of females would demonstrate greater reading skill than did male students. The demographic factor of the sex of the student was examined using multiple regression analysis to determine whether that factor was significant in predicting the results of the eighth-grade PSSA reading assessment. It

was hypothesized that this predictor would be significant, as research indicated that male students were more likely to evidence reading difficulty.

It was hypothesized that the disaggregated group of students with disability would demonstrate less developed reading skill than the aggregated group of students. The demographic factor of educational disability was examined using multiple regression analysis, to determine whether that factor was significant in predicting the results of the eighth-grade PSSA reading assessment. Next, the specific disability categories were analyzed to determine their statistical significance as predictors of the PSSA reading scaled scores. Research indicated students with educational disabilities would perform less well on lengthy assessments similar to the PSSA8 reading assessment than students in regular education without disabilities. Multiple regression was used to determine which demographic factors predicted student scaled scores on the PSSA reading assessment (Cresswell, 1994). This step established the predictive significance of the categories of sex of the student and educational disability.

4. Did the students' JSSA4 through JSSA7 reading scores predict their eighth grade PSSA reading scaled scores, when examining students who participated in either the John Hopkins' *Success for All* reading curriculum or the *Houghton-Mifflin* basal reading curriculum?

It was hypothesized that students who participated in the John Hopkins' *Success for All* curriculum would perform better on the PSSA reading assessment than those students who received the alternate curriculum. This was hypothesized despite the greater economic disadvantage percentage for the *Success for All* group; and the discontinuity in

later curricular and curriculum-based assessment source material with *Success for All* materials.

Finally, the two reading curricula were evaluated as predictor variables for those same PSSA reading scaled scores. The two curricula were examined using multiple regression analysis in combination with the factors of sex of the student, the presence of an educational disability, and the curricular-based assessment score category. Within this array of variables, it was hypothesized that the *Success for All* curriculum would better match student needs at the Greater Johnstown School District, yielding better results on the eighth-grade reading PSSA than the results for those students receiving the *Houghton Mifflin* reading basal curriculum.

#### Summary

The purpose of this study was to examine the CBA program used by the Greater Johnstown School District, in order to determine the relationship of district students' CBA reading scores, reading curriculum, and student demographic factors to those same students' eighth grade PSSA reading scores. All student data were encoded within Excel spread sheets; and various statistical analyses were used to examine the student data. All individual data was identified by a randomly generated number.

Research foci of this study were to provide rationale and to describe research methods which may be implemented in the examination of district curricular-based assessments and eventual student reading performance on PSSA reading measures. Student cohort data were examined in both disaggregated and combined forms, in order to establish predictive strength of the various measures. The current study attempted to provide data to establish the reliability and validity of the school district's prompts as

they related to PDE standards, and the researcher hoped to indicate particularly valuable measurement points in predicting eventual eighth-grade PSSA reading performance.

Information collected during this study contributed to the growing body of data regarding the use of curricula-based measurement systems. Student participants' data contributed to the accurate identification of needed frequency, types, and design of these curricular-based measurement systems. Additionally, the discernment of most accurate predictive measures and most necessary years for that summative assessment related to that prediction allowing school staff to focus direct instructional and formative assessment time, remediation, and curricular adjustment for those students most in need.

## CHAPTER IV

### RESULTS

#### Introduction

In Chapter IV, the results of the study are displayed. Results for each question were presented with a brief description, followed by a summary of the significant results.

The significant results included these five items:

1. The district curriculum-based assessment (CBA) score categories were highly correlated with reading grades.
2. The sex of the student was not a significant predictor of the Pennsylvania System of School Assessment eighth grade (PSSA8) reading scaled scores.
3. Student disability was a significant predictor of PSSA8 reading scaled scores.
4. CBA data, when analyzed with sex of the student and student disability, was not a significant predictor of PSSA8 reading scaled scores.
5. The two district curricula, when analyzed with sex of the student and student disability, were not significant predictors of PSSA8 reading scaled scores.

To address the question regarding which variables predicted the PSSA eighth grade reading scaled score, several procedures were followed. After validating the Johnstown System of Student Assessment (JSSA) as a curriculum-based assessment of reading fluency through correlation with the students' reading grades during that same school year, each question examined data relating to portions of this study. The various

sequential, curriculum-based assessments were evaluated as predictors of that eighth grade reading performance, and the two district reading curricula were assessed for any additional contribution toward predicting the eighth grade PSSA reading scaled scores. The final question examined the contribution of the sex of those students and their categorization as having an educational disability or specific exceptionality toward their eighth grade PSSA reading performance.

*Research Question One*

Did the Johnstown System of Student Assessment for reading demonstrate concurrent validity when correlated to student reading grades in the fourth grade year?

*Research Question Two*

Did the Johnstown System of Student Assessment (JSSA) reading scores predict the Pennsylvania System of School Assessment (PSSA) reading scaled scores?

*Research Question Three*

Were the demographic factors of the sex and educational disability of the student significant predictors of the students' eighth grade PSSA reading scaled scores?

*Research Question Four*

Did the students' JSSA4 through JSSA7 reading scores predict their eighth grade PSSA reading scaled scores when examining students who participated in either the John Hopkins' *Success for All* reading curriculum or the *Houghton-Mifflin* basal reading curriculum?

## Complications

Few complications attended the completion of this study. However, a high number of transient students and incomplete student records limited the number of participants with complete sets of data. For this reason, the analyses which required numerous predictors used a small sample of students, and yielded results which were only suggestive.

## Computer Programs

The student data were entered into an SPSS data analysis program. SPSS (originally, Statistical Package for the Social Sciences) was released in its first version in 1968, and is widely used for statistical analysis in social science. The data were saved without individual identification markers, and was used to complete both correlation and multiple regression analyses.

## Analysis: Step One

### *Research Question One*

Did the Johnstown System of Student Assessment for reading demonstrate concurrent validity when correlated to student reading grades in the fourth grade year?

In step one, the relationship between curricular-based reading performance levels (JSSA) and reading performance measured as report card grades at that same fourth-grade level was examined. This correlation investigated the concurrent validity of the district's curricular-based reading measure (See Table 5 and Table 6).

Table 5

*Case Processing Summary*

	Cases Valid N	Percent	Missing N	Percent	Total N	Percent
JSSA and GPA Reading	185	44.3	233	55.7	418	100

*Note.* Johnstown System of Student Assessment (JSSA); Grade Point Average (GPA).

Table 6

*Correlation Matrix of JSSA Reading Performance Level and Reading Grade Point Average  
of Fourth Graders from the Greater Johnstown School District*

Reading Grades	1.0 GPA	2.0 GPA	3.0 GPA	4.0 GPA	Total	
BB JSSA 4	.31	.00	.02	.02	.00	
B JSSA 4	.01	.27	.10	.10	.08	
P JSSA 4	.02	.71	.52	.52	.08	
A JSSA 4	.00	.02	.37	.37	.85	
Total N	43	66	63	13		185

*Note.* In the Pennsylvania Department of Education rubric, reading performance levels are as follows: Below Basic (BB), Basic (B), Proficient (P), and Advanced (A). Reading GPA connotes the final or averaged grade for reading across the fourth grade marking periods. The student reading grades were estimated, rounding up or down to the nearest whole Grade Point Average (GPA). The Johnstown System of Student Assessment (JSSA) is scored at the same performance levels as the PSSA. Below Basic (BB), Basic (B), Proficient (P), and Advanced (A).

## Results of Step One

When reading grades and JSSA scores were correlated, linearity was checked and was apparent in this correlation data. The JSSA4 fourth grade curricular-based scores were significantly related to the students' classroom reading grades in fourth grade. A Pearson's  $R$  value of 0.690\* and a Spearman's Rho value of 0.685\* corroborated the highly significant relationship between reading grades and the JSSA performance level for students in fourth grade.

## Analysis: Step Two

### *Research Question Three*

Were the demographic factors of sex and educational disability of the student significant predictors of the students' eighth grade PSSA reading scaled scores?

Research question three examined the student demographics including educational disability, sex of the student, and the specific disability categories of the students, in order to determine whether the student characteristics or disability categorizations were predictors of student results on the eighth grade PSSA reading exam. It was considered desirable to determine whether particular groups of students would perform more or less well on the state exam, based on these categories and characteristics. Predictors included sex of the student and the qualification as having an educational disability or placement versus no disability and full-time regular education status. The PSSA scaled score on the students' eighth-grade reading measure was used as the criterion.

Then, specific predictors were examined, including sex of the student, and the primary educational disability and placement categories of: Emotional Disturbance (ED)

or Serious-Emotional Disturbance (SED), Learning Disability (LD) or Specific Learning Disability (SLD), Alternative Education (AE), Speech and Language Impairment (SLI), Mental Retardation (MR) with or without SLI, and Other Health Impairment (OHI). The PSSA scaled score on the students' eighth-grade reading measure was used as the criterion. The initial regression results are presented in Tables 7 and 8. This analysis predicted the eighth-grade PSSA reading results from the variables of sex of the student and the student categorization with an educational disability.

### Results of Step Two

In question three, the demographic characteristics of the student cohort (i.e. sex of the student, categorization as having an educational disability, as well as the specific type of educational disability) were examined to find any relationship between these characteristics and the students' scores on the PSSA eighth-grade reading exam. It was hypothesized that female students would perform more strongly on the PSSA8 reading assessment than did male students; while students with educational disabilities would perform less well on the PSSA8 reading assessment than students in regular education without disabilities. In this analysis, the predictor of sex of the student was demonstrated to be statistically insignificant ( $p < .05$ ).

In Table 9, the demographic factor of educational disability was examined using multiple regression analysis, to determine whether it was significant in predicting the results of the eighth-grade PSSA reading assessment. It was hypothesized that this predictor would be significant, as research indicated that students with educational disabilities would perform less well on the PSSA8 reading assessment than students in regular education without disabilities. When multiple regression analysis was used to

Table 7

*Multiple Regression Predicting the Eighth Grade PSSA Reading Performance Level from Sex of the Student and Disability*

Correlation of Predictors and Criterion				
Variable	PSSA8	Sex	Disability	
PSSA8 SS	1.00	.06	-.80**	
Sex	.06	1.00	-.21	
Disability	-.80**	-.21	1.00	

Descriptive Statistics				
Variables	n	Mean	SD	Range
PSSA8 SS	238	1252.5	227.8	758-1846
Sex	268	.51	0.5	0-1
Disability	111	.51	0.5	0-1

Regression Predicting PSSA8 SS Using Sex and Disability Categories			
Model Fit	R <sup>2</sup>	R <sup>2</sup> Adj	
93.921 [2, 98] = 134.17; p < .001	.66,	.65	

Table 7 (Continued)

*Multiple Regression Predicting the Eighth Grade PSSA Reading Performance Level from Sex of the Student and Disability*

	Variables in Equation				
	B	SE	<i>b</i>	<i>t</i>	<i>p</i>
Constant	1470.9	25.60		57.47	<.001**
Sex	-51.6	27.38	-.11	-1.89	.062
Disability	-373.4	27.31	-.83	-13.67	<.001**

*Note.* \*This is significant at the  $p < .05$  level (2-tailed). \*\*This is significant at the  $p < .01$  level (2-tailed). PSSA8 SS means the student scaled score on the Pennsylvania System of School Assessment reading exam in eighth grade.

Table 8

*Multiple Regression Predicting the Eighth Grade PSSA Reading Performance Level from Sex of the Student and Specific Disability*

Descriptive Statistics				
Variable	n	Mean	S.D.	Range
Constant	47	998.6	184.3	758-1846
SEX	47	1.4	0.5	1-2
ED	47	0.1	0.3	0-1
SLD	47	0.5	0.5	0-1
AE	47	0.1	0.1	0-1
SLI	47	0.1	0.1	0-1
MR	47	0.3	0.5	0-1

Correlation of Independent and Criterion							
Variables	PSSA8	SEX	ED	SLD	AE	SLI	MR
PSSA8 SS	1.00	.07	.19	.18	-.13	.24	-.36*
Sex		1.00	.27*	.04	-.02	.17	.19
ED			1.00	-.47**	-.11	-.06	-.26
SLD				1.00	-.24	-.14	-.55**
AE					1.00	-.03	-.13
SLI						1.00	-.07
MR							1.00

Regression Predicting PSSA8 SS Using Sex and Specific Disability Categories			
Model Fit		R <sup>2</sup>	R <sup>2</sup> Adj
	2.929 [6,40] = 164.69;	p = .018	.31, .20

Table 8 (Continued)

*Multiple Regression Predicting the Eighth Grade PSSA Reading Performance Level from Sex of the Student and Specific Disability*

	Variables in Equation				
	B	SE	<i>b</i>	<i>t</i>	<i>p</i>
Constant	869.5	172.55		5.04	.001
SEX	33.5	51.50	.09	0.65	.519
ED	233.7	178.09	.43	1.31	.197
SLD	86.9	169.47	.24	0.51	.611
AE	-16.0	232.90	-.01	-0.07	.946
SLI	507.5	238.53	.40	2.13	.040*
MR	-18.7	173.14	-.05	-0.11	.915

Regression Predicting PSSA8 Reading after Eliminating the Least Significant Sex and Disability Categories

Model Fit	R <sup>2</sup>	R <sup>2</sup> Adj
2.929 [6, 40] = 164.69; p = .018	.31	.20

*Note.* \*This is significant at the  $p < .05$  level (2-tailed). \*\*This is significant at the  $p < .01$  level (2-tailed). Specific educational disabilities and placement are considered. These include: Emotional Disturbance (ED), Specific Learning Disability (SLD), Alternative Education placement (AE), Speech and Language Impairment (SLI), and Mental Retardation (MR).

Table 9

*Multiple Regression Predicting the Eighth Grade PSSA Reading Performance Level from the Fourth- through Seventh-Grade JSSA Reading Performance Level, Sex of the Student, and Disability*

Descriptive Statistics				
Variables	n	Mean	S.D.	Range
PSSA8 SS	238	970.8	159.4	758-1846
Sex	268	1.5	0.5	1-2
Disability	111	0.5	0.5	0-1
JSSA4	188	2.2	0.9	1.4
JSSA5	192	2.4	0.8	1.4
JSSA6	198	2.3	0.7	1.4
JSSA7	202	2.2	0.5	1.4

Correlation Matrix for all Predictors and Criterion							
Variables	PSSA8	Sex	Disability	JSSA4	JSSA5	JSSA6	JSSA7
PSSA8	1.00	.06	-.80	.47	.47	.60	.52
Sex		1.00	-.21	.09	.07	.13	.08
Disability			1.00	-.64	-.60	-.63	-.50
JSSA4				1.00	.61	.55	.52
JSSA5					1.00	.56	.46
JSSA6						1.00	.63
JSSA7							1.00

Table 9 (Continued)

*Multiple Regression Predicting the Eighth Grade PSSA Reading Performance Level from the Fourth- through Seventh-Grade JSSA Reading Performance Level, Sex of the Student, and Disability*

Regression Predicting PSSA8 Performance Level from the Fourth- through Seventh-Grade JSSA Reading Performance Level, Sex, and Disability					
Model Fit			R <sup>2</sup>	R <sup>2</sup> <sub>Adj.</sub>	
23.767 [6,62] = 130.77; p < .001 .70, .67					
Variable(s) in the Equation					
Variables	B	SE	<i>b</i>	<i>t</i>	<i>p</i>
Constant	1360.9	113.25		12.02	.001
Sex	-54.9	32.5	-.12	-1.69	.096
Disability	-364.5	47.8	-.81	-7.63	.001**
JSSA4	-39.88	26.4	-.15	-1.51	.136
JSSA5	-17.11	26.3	-.06	-0.65	.517
JSSA6	41.01	32.5	.13	1.26	.212
JSSA7	62.84	39.1	.15	1.61	.113

*Note.* PSSA8 SS means the student scaled score on the Pennsylvania System of School Assessment reading exam in eighth grade; while JSSA 4-7 mean the reading categories earned on the Johnstown System of Student Assessment at grades four through seven.

examine the results of considering the predictor variables of sex of the student and educational disability, only educational disability was a significant predictor of the PSSA8 reading scaled score ( $p < .01$ ).

### Analysis: Step Three

#### *Research Question Two*

Did the Johnstown System of Student Assessment (JSSA) reading scores predict the Pennsylvania System of School Assessment (PSSA) reading scaled scores?

In examining research question two, data were collected to determine whether the addition of curriculum-based assessment score categories increased the ability to predict eighth grade PSSA reading scores. In the first model, as seen in Table 9, the question was posed using the variables of sex of the student, exceptional student status (i.e. students with an educational disability or placement), and the fourth- through seventh-grade JSSA performance levels. Then, in a second analysis shown in Table 10, the question was posed using the variables of sex of the student, specific educational disability (e.g., Mental Retardation), and the fourth- through seventh-grade JSSA performance levels. It was hypothesized that the fourth- through seventh-grade JSSA performance levels would be significant predictors of the eighth-grade PSSA reading performance on the PSSA assessment.

Multiple regression analysis was used to examine the results of adding fourth through seventh grade district CBA score datum to the predictor variables of sex of the student and specific educational disability. When considering sex of the student, generic categorization with disability, and fourth through seventh grade district CBA datum, only student disability was a significant predictor ( $p < .001$ ). When considering sex of the

student, specific disability category, and fourth through seventh grade district CBA datum, the model did not fit. The district CBA datum was regularly used as a formative assessment, guiding instructional grouping and practice. It had been hypothesized that the district CBA datum would predict the PSSA8 scaled scores. However, when analyzed in the multiple regression with other predictors, it was not a significant predictor of the PSSA8 reading scaled scores.

### Results of Step Three

Multiple regression analysis examined the results of adding CBA data to the predictor variables of sex of the student and educational disability and placement. In Table 9, the fourth through seventh grade district CBA categories were entered as predictor variables. The results of this analysis indicated that only educational disability was a significant predictor of the PSSA8 reading scaled score ( $p < .01$ ).

In Table 8, the specific educational disabilities were entered as predictor variables, rather than the generic categorization of disability which was used in Table 9. Multiple regression analysis examined the results of adding fourth through seventh grade district CBA data to the predictor variables of sex of the student and specific educational disability. Neither Table 10 nor Table 11 yielded a model which explained the data. The results of this analysis indicated that only educational disability was a significant predictor of the PSSA8 reading scaled score ( $p < .01$ ). However, this result must be viewed only as suggestive in view of the very small sample size ( $N = 30$ ). “An optimal experimental design consists of an adequate sample size from an accuracy in parameter estimation

Table 10

*Multiple Regression Predicting Eighth Grade PSSA Reading Performance Level from Sex of the Student, Specific Disability, and Fourth- through Seventh-Grade District Curriculum-Based Assessment (JSSA) Reading Performance Levels*

Variable	n	Mean	S.D.	Range
PSSA8 SS	30	1047.6	169.0	758-1846
SEX	30	1.4	0.5	1-2
ED	30	0.7	0.3	0-1
SLD	30	0.6	0.5	0-1
AE	30	0.0	0.0	0-1
SLI	30	0.7	0.3	0-1
MR	30	0.3	0.5	0-1
BB-JSSA4	30	0.6	0.5	0-1
B-JSSA4	30	0.2	0.4	0-1
P-JSSA4	30	0.2	0.4	0-1
BB-JSSA5	30	0.4	0.5	0-1
B-JSSA5	30	0.3	0.5	0-1
P-JSSA5	30	0.2	0.4	0-1
BB-JSSA6	30	0.4	0.5	0-1
B-JSSA6	30	0.6	0.5	0-1
P-JSSA6	30	0.0	0.0	0-1
BB-JSSA7	30	0.3	0.5	0-1
B-JSSA7	30	0.6	0.5	0-1
P-JSSA7	30	0.1	0.3	0-1

Table 10 (Continued)

*Multiple Regression Predicting Eighth Grade PSSA Reading Performance Level from Sex of the Student, Specific Disability, and Fourth- through Seventh-Grade District Curriculum-Based Assessment (JSSA) Reading Performance Levels*

---

Correlation of Sex, Curriculum-Based Assessments and Disabilities

---

	Sex	BBJ4	BJ4	PJ4	MR	BBJ5	BJ5	PJ5	ED	SLD	AE	SLI
Sex	1.00	-.15	.08	.07	.19	-.01	-.13	.15*	-.27*	.04	-.02	.17
BB JSSA4		1.00	-.46*	-.28*	-.03	.59**	.03	-.32**	-.24	.14	.19	-.17
B JSSA4			1.00	-.53**	-.05	-.24**	.18*	.12	.07	-.02	-.14	.23
P JSSA4				1.00	.11	-.20**	-.07	.06	.24	-.17	-.08	-.08
MR					1.00	-.03	.01	.07	-.26	-.55**	-.13	-.07
BB JSSA5						1.00	-.33**	-.25**	-.27	.08	.24	-.13
B JSSA5							1.00	-.63**	.04	.01	-.16	.20
P JSSA 5								1.00	.31	-.18	-.08	-.08
ED									1.00	-.47**	-.11	-.06
SLD										1.00	-.24	-.14

---

Correlation of Sex, Curriculum-Based Assessments and Disabilities

---

	Sex	BBJ4	BJ4	PJ4	MR	BBJ5	BJ5	PJ5	ED	SLD	AE	SLI
AE											1.00	-.03
SLI												1.00

---

Table 10 (Continued)

*Multiple Regression Predicting Eighth Grade PSSA Reading Performance Level from Sex of the Student, Specific Disability, and Fourth- through Seventh-Grade District Curriculum-Based Assessment (JSSA) Reading Performance Levels*

Regression Predicting Eighth-Grade PSSA Performance Level from Sex, Exceptionality, and Fourth- and Fifth-Grade District Curriculum-Based Assessment (JSSA) Performance Levels			
Model Fit		R <sup>2</sup>	R <sup>2</sup> <sub>Adj.</sub>
1.965	[10, 19] = 140.36; p = .099	.51,	.25

Regression Predicting Eighth-Grade PSSA Performance Level from Sex of the Student, Exceptionality, and Fourth- through Seventh-Grade District Curriculum-Based Assessment (JSSA) Performance Levels			
Model Fit		R <sup>2</sup>	R <sup>2</sup> <sub>Adj.</sub>
5.582	[11, 2] = 76.51; p = .162	.98,	.97

*Note.* In Table 8, specific educational disabilities and placement are considered. The JSSA4-JSSA7 codes represent the district CBA measures at grades four through seven, with score categories of BB,B,P, or A.

BB= Below basic performance

B= Basic performance.

P= Proficient performance

A= Advanced performance.

ED= Emotional disturbance

SLD= Specific learning disability

AE= Alternative education placement

SLI= Speech and language impairment

MR= Mental retardation

PSSA8 SS= PA System of School Assessment at eighth grade, reading scaled score.

Table 11

*Multiple Regression Predicting the Eighth-Grade PSSA Performance Level from the Fourth-Grade JSSA Performance Level, Sex of the Student, and Specific Disability*

Variable	n	Mean	S.D.	Range
PSSA8S	31	970.8	159.4	758-1846
	31	1.5	0.5	1-2
	31	0.5	0.5	0-1
	31	0.4	0.5	0-1
	31	0.2	0.4	0-1
	31	0.1	0.2	0-1
	31	0.6	0.5	0-1
	31	0.1	0.2	0-1
	31	0.1	0.2	0-1
	31	0.3	0.5	0-1

Correlation Matrix for all Predictors and Criterion

Variables	N Sex	BB-JSSA4	B-JSSA4	P-JSSA4	ED	SLD	AE	SLI	MR
N SEX	1.00	0.15*	0.08	0.07	-0.27*	0.04	-0.02	0.17	0.19
BB JSSA4		1.00	-0.46**	-0.28**	-0.24	0.14	0.19	-0.71	-0.30
B JSSA4			1.00	-0.53**	0.07	-0.02	-0.14	0.23	-0.05
P JSSA4				1.00	0.24	-0.17	-0.08	-0.08	0.11
ED					1.00	-0.47**	-0.10	-0.06	-0.26
SLD						1.00	-0.24	-0.14	-0.55**
AE							1.00	-0.03	-0.13
SLI								1.00	-0.07
MR									1.00

Regression Predicting PSSA8 Using Sex, Disability (MR, SLI, ED,AE) and CBA Performance Level (BB- and P-JSSA4)

Model Fit	$R^2$	$R^2_{Adj.}$
2.207 [7, 23] = 140.78; p = .072	.40,	.22

Table 11 (Continued)

*Multiple Regression Predicting the Eighth-Grade PSSA Performance Level from the Fourth-Grade JSSA Performance Level, Sex of the Student, and Specific Disability*

Descriptive Statistics												
Variables	n	Mean	S.D.	Range								
PSSA8 SS	31	970.8	159.4	758-1846								
SEX	31	1.5	0.5	1-2								
BB-JSSA4	31	0.5	0.5	0-1								
B-JSSA4	31	0.4	0.5	0-1								
P-JSSA4	31	0.2	0.4	0-1								
ED	31	0.1	0.2	0-1								
SLD	31	0.6	0.5	0-1								
Alt. Ed.	31	0.1	0.2	0-1								
SLI	31	0.1	0.2	0-1								
MR	31	0.3	0.5	0-1								
	Sex	BBJ4	BJ4	PJ4	MR	BBJ5	BJ5	PJ5	ED	SLD	AE	SLI
Sex	1.00	-.15*	.08	.07	.19	-.01	-.13	.15*	-.27*	.04	-.02	.17
BB JSSA4		1.00	-.46*	-.28*	-.03	.59**	.03	-.32**	-.24	.14	.19	-.17
B JSSA4			1.00	-.53**	-.05	-.24**	.18*	.12	.07	-.02	-.14	.23
P JSSA4				1.00	.11	-.20**	-.07	.06	.24	-.17	-.08	-.08
MR					1.00	-.03	.01	.07	-.26	-.55**	-.13	-.07
BB JSSA5						1.00	-.33**	-.25**	-.27	.08	.24	-.13
B JSSA5							1.00	-.63**	.04	.01	-.16	.20
P JSSA 5								1.00	.31	-.18	-.08	-.08
ED									1.00	-.47**	-.11	-.06
SLD										1.00	-.24	-.14
AE											1.00	-.03
SLI												1.00

Table 11(Continued)

*Multiple Regression Predicting the Eighth-Grade PSSA Performance Level from the Fourth-Grade JSSA Performance Level, Sex of the Student, and Specific Disability*

Regression Predicting Eighth-Grade PSSA Performance Level from Sex of the Student, Disability, and Fourth and Fifth Grade District Curriculum-Based Assessment (JSSA) Performance Levels			
Model Fit		R <sup>2</sup>	R <sup>2</sup> <sub>Adj.</sub>
1.965	[10, 19] = 140.36;	p = .099	.51, .25
Regression Predicting Eighth-Grade PSSA Performance Level from Sex, Disability, and Fourth- through Seventh-Grade District Curriculum-Based Assessment (JSSA) Performance Levels			
Model Fit		R <sup>2</sup>	R <sup>2</sup> <sub>Adj.</sub>
5.582	[11, 2] = 76.51;	p = .162	.98, .97

*Note.* \*This is significant at the  $p < .05$  level (2-tailed). \*\* This is significant at the  $p < .01$  level (2-tailed). PSSA8 SS= PA System of School Assessment at eighth grade, reading scaled score.

BB= Below basic performance  
 B= Basic performance.  
 P= Proficient performance  
 A= Advanced performance.

ED= Emotional disturbance  
 SLD= Specific learning disability  
 AE= Alternative education placement  
 SLI= Speech and language impairment  
 MR= Mental retardation

perspective, as well as an adequate sample size from the power analysis perspective. Ensuring that sample size is adequate from both perspectives leads to parameter estimates that will likely be accurate as well as statistically significant (Kelley & Maxwell, 2003).” Greene (1991) has recommended a sample size of 104 plus the number of predictors. In this case, that rule would indicate at least 108 participants.

#### Analysis: Step Four

##### *Research Question Four*

How predictive were the students’ JSSA4 through JSSA7 scores of their eighth grade PSSA reading scores, when examining students who participated in either the John Hopkins’ *Success for All* reading curriculum or the *Houghton-Mifflin* basal reading curriculum?

Step four posited that the addition of differentiated curriculum would predict the eighth grade PSSA reading scores beyond these demographic and CBA variables. Because the *Success for All* reading curriculum was both so well matched to the needs of students at the Greater Johnstown School District and was designed for explicit instruction in many of the reading skills and areas measured by the PSSA, those students receiving this curriculum were posited to perform better than the alternate subgroup of students. The performance of students who received the alternative reading curriculum was posited to be less related to their later PSSA reading score.

Table 12

*Multiple Regression Predicting PSSA Reading Performance Level from Sex of the Student, Disability, Curriculum-Based Assessments, and Two Reading Curricula*

Descriptive Statistics				
Variable	n	Mean	S.D.	Range
PSSA8	238	1252.5	226.8	758-1846
SEX	268	0.5	0.5	0-1
Disability	111	0.5	0.5	0-1
JSSA4	188	2.2	0.9	1-4
JSSA5	192	2.4	0.8	1-4
JSSA6	198	2.3	0.7	1-4
JSSA7	202	2.2	0.5	1-4
SFA	197	0.5	0.5	1-4

Correlation Matrix								
	PSSA8-SS	SEX	Disability	JSSA4	JSSA5	JSSA6	JSSA7	SFA
PSSA8	1.00	.05	-.80	.47**	.47**	.60**	.52**	.03
SEX		1.00	-.21	.09	.07	.13	.08	.04
Disability			1.00	-.64	-.60**	-.63**	-.50**	-.08
JSSA4				1.00	.61**	.55**	.52**	.12
JSSA5					1.00	.56**	.46**	.24**
JSSA6						1.00	.63**	-.04
JSSA7							1.00	.19
SFA								1.00

Table 12 (Continued)

*Multiple Regression Predicting PSSA Reading Performance Level from Sex of the Student, Disability, Curriculum-Based Assessments, and Two Reading Curricula*

Regression Predicting PSSA8-SS Using Sex, Disability, CBA Curricular Performance Level and Curriculum						
Model Fit					R <sup>2</sup>	R <sup>2</sup> Adj
20.116 [7, 61] = 131.67; p < .001					.70,	.66
Variable(s) in Equation						
	B	SE	<i>b</i>	t	<i>p</i>	
Constant	1354.6	115.14		11.77	.001	
SEX	-55.5	32.74	-.12	-1.69	.095	
Disability	-364.6	48.08	-.81	-7.58	.001**	
JSSA4	-40.5	26.65	-.16	-1.52	.133	
JSSA5	-20.2	27.56	-.07	-0.73	.467	
JSSA6	42.5	32.98	.14	1.29	.202	
JSSA7	65.6	39.90	.16	1.64	.106	
SFA	13.5	34.29	.03	0.40	.694	

*Note.* The JSSA4-JSSA7 codes represent the district CBA reading measures in grades four through seven. PSSA8 SS= PA System of School Assessment at eighth grade, reading scaled score. SFA=Participation in the *Success for All* reading curriculum.

Table 13

*Multiple Regression Predicting PSSA Reading Performance Level from Sex of the Student, Specific Disability Curriculum-Based Assessments, and Two Reading Curricula*

Variable	Descriptive Statistics			
	n	Mean	S.D.	Range
PSSA8 SS	13	1062.7	165.8	758-1846
SEX	13	1.4	0.5	1-2
ED	13	0.1	0.3	0-1
SLD	13	0.6	0.5	0-1
AE	13	0.0	0.0	0-1
SLI	13	0.1	0.3	0-1
MR	13	0.2	0.4	0-1
JSSA4-BB	13	0.6	0.5	0-1
JSSA4-B	13	0.2	0.4	0-1
JSSA4-P	13	0.2	0.4	0-1
JSSA5-BB	13	0.5	0.5	0-1
JSSA5-B	13	0.3	0.5	0-1
JSSA5-P	13	0.2	0.4	0-1
JSSA6-BB	13	0.3	0.5	0-1
JSSA6-B	13	0.7	0.5	0-1
JSSA6-P	13	0.0	0.0	0-1
JSSA7-BB	13	0.3	0.5	0-1
JSSA7-B	13	0.6	0.5	0-1
JSSA7-P	13	0.1	0.3	0-1
SFA	13	0.5	0.5	0-1

Table 13 (Continued)

*Multiple Regression Predicting PSSA Reading Performance Level from Sex of the Student, Specific Disability, Curriculum-Based Assessments, and Two Reading Curricula*

Correlation Matrix						
	PSSA8-SS	SLD	J5-BB	J5-B	J7-BB	J7-P
PSSA8-SS	1.00	.04	.36**	.12	-.27**	.42**
SLD		1.00	.08	.01	.21	.42**
J5-88			1.00	-.33**	.32**	-.15
J5-B				1.00	-.07	-.25**
J7-BB					1.00	-.11
J7-P						1.00
SFA						

Regression Predicting PSSA8 Using Sex, Specific Disability, CBA, and Curricular Performance Level

Model Fit	R <sup>2</sup>	R <sup>2</sup> Adj
5.436 [10, 2] = 76.51; p = .165	.98,	.97

*Note.* The JSSA4-JSSA7 codes represent the district CBA reading measures in grades four through seven, with score categories of BB,B,P, or A.

- BB= Below basic performance
- B= Basic performance.
- P= Proficient performance
- A= Advanced performance.
- ED= Emotional disturbance
- SLD= Specific learning disability
- AE= Alternative education placement
- SLI= Speech and language impairment
- MR= Mental retardation

PSSA8 SS= PA System of School Assessment at eighth grade, reading scaled score.  
 SFA=Participation in the *Success for All* reading curriculum.

## Results of Question Four

Multiple regression analysis was used with this interval/ratio data and yielded a statistically insignificant result for each predictor represented in this hypothesis. The addition of the *Success for All* reading curriculum did not better predict student performance on the eighth grade PSSA reading exam than did the alternative reading curriculum, when analyzed with the predictors of sex of the student, exceptionality or disability and CBA performance level scores. This analysis was limited by the very small sample size (N=13) which remained after all predictors had been considered. All of the above results should be considered in the light of that small sample size.

## Summary

A review of the results indicated the various significant predictors of PSSA8 reading performance. Although student demographics beyond having an educational disability were not a significant contribution to the prediction of this performance, data suggests that students who qualified as having speech and language impairment (SLI) have demonstrated this significant relationship. Similarly, when curriculum-based assessment and reading curriculum data were included, only disability significantly predicted the students' PSSA eighth grade reading performance level. Finally, these significant results remain:

1. The district CBA score categories were highly correlated to reading grades.
2. The sex of the student was not a significant predictor of PSSA8 reading scaled scores ( $p > .05$ ).

3. Student disability was a significant predictor of PSSA8 reading scaled scores ( $p < .01$ ).
4. CBA data, when analyzed with sex of the student and student disability, was not a significant predictor of PSSA8 reading scaled scores ( $p > .05$ ).
5. The two district curricula, when analyzed with sex of the student and student disability, were not significant predictors of PSSA8 reading scaled scores ( $p > .05$ ).

## CHAPTER V

### DISCUSSION

#### Introduction

Chapter V reviews the purpose of this study and summarizes the findings reviewed in Chapter IV. These findings are related to the literature and implications for the profession are discussed. Recommendations for further research in the practices of curriculum-based assessment and benchmarking are offered. The chapter concludes with a final review and summary of this study.

#### Purpose of the Study

Few studies have examined repeated, sequential curriculum-based assessment used as benchmarking over several years (Juel, 1988; Schulte, et al., 2001). The purposes of this data collection included the use of the data to drive and inform instruction, and to predict the scores that this cohort of students would earn on the summative statewide system of assessment in reading. This study examined whether the curriculum-based assessment data, which was collected to inform instruction, also predicted the students' scores on that statewide assessment.

#### A Review of the Procedures and Analyses

This study followed the student performance on the spring (third annual) curriculum-based benchmarking assessment in reading for an entire grade level cohort of district students, from their fourth through their seventh grade years. The curriculum-based assessment data, reading curricula, and student demographic characteristics were

then evaluated as predictors of the student cohort's performance on the Pennsylvania high-stakes test of reading, the PSSA.

The participants in this study were an entire grade-level cohort of students from the Greater Johnstown public schools in Pennsylvania. This cohort included 48% males, 21% students with an educational disability or special-educational placement, and 22% students with ethnic minority status. All students were assessed three times per year using a CBA reading measure from fourth through seventh grade.

Before considering the predictor variables, the students' classroom reading grades were correlated to their curriculum-based assessment performance on the JSSA at the fourth-grade level in order to establish the concurrent validity of the local measure. It was important to determine whether the curriculum-based assessment measured similar and clearly related skills when compared to the grade-level reading measure. As this correlation was significant and positive, as Shinn and Good (1993) indicated should be the case, the study continued.

This multiple regression analysis was completed in three separate examinations of the datum. The district CBA datum was regularly used as a formative assessment, guiding instructional grouping and practice. However, when analyzed in the multiple regression with other predictors, it was not a significant ( $p > .05$ ) predictor of the PSSA8 reading scaled scores.

The characteristics of sex and disability status were also appraised as possible predictors. Although there existed a body of research suggesting that male students with CBA skills that were similar to those of their female peers might not perform as well on lengthy, problem-solving formats like the PSSA (Griffin, 1998), the Johnstown study did

not provide data that corroborated that research. The categorization of having an educational disability versus having no such disability was, however, a significant predictor of the students' performance on the PSSA reading assessment.

In the next analysis, specific disability categories were evaluated as potential predictors of this PSSA performance. The student disability categories were broadly represented, yet the sample size within each category was small. Still, the possibility of differentiating PSSA performance based upon the specific exceptionality was of high interest. Although the data could only be considered as suggestive, the disability of SLI alone was a significant predictor of the PSSA reading performance when all demographic factors were considered.

Finally, the two Johnstown reading curricula were analyzed as variables to determine whether either better predicted to the eighth-grade PSSA reading performance of this student cohort. The curricula represented were the Johns Hopkins' University product *Success for All*, and the *Houghton-Mifflin* reading basal program. Neither curriculum proved to be a significant predictive factor.

### A Review of the Results

Four research questions were proposed in this study, and results will be summarized with the framework of those queries.

#### *Research Question One*

Did the Johnstown System of Student Assessment for reading demonstrate concurrent validity when correlated to student reading grades in the fourth grade year?

In the preliminary correlation, the relationship between curriculum-based reading performance and reading grades in the fourth-grade students' reading classrooms was examined. A Pearson's R value of 0.690 and probability of high significance ( $p < .01$ ) resulted from this initial analysis. In a second examination of this data, a Spearman's Rho value of 0.685 corroborated the results of the Pearson's correlation analysis, also yielding a probability of high significance ( $p < .01$ ). Linearity was checked and was apparent in this correlation analysis.

### *Research Question Two*

Did the Johnstown System of Student Assessment (JSSA) reading scores predict the Pennsylvania System of School Assessment (PSSA) reading scaled scores?

The second research question used the student cohort's fourth- through seventh-grade curriculum-based reading assessment performance categories to determine whether these CBA performance categories, in combination with demographic factors, would increase the ability to predict the students' eighth-grade PSSA reading scaled scores.

In the secondary analysis, the students' sex, fourth- through seventh-grade CBA performance categories, and the specific educational disabilities were used as possible predictors of the PSSA reading scaled scores. The existing disability categories included emotional disturbance, specific learning disability, alternative education placement, other health impairment, speech and language impairment, and mental retardation, while the CBA performance categories were below basic, basic, proficient, and advanced. This second analysis demonstrated that the district curriculum-based assessment was not a significant predictor of those students' eventual PSSA8 reading performance ( $p > .05$ ). The district CBA process and datum were reasonably employed as a formative measure,

guiding instructional placement and practice. The datum, however, was not technically strong enough to provide a prediction of student reading performance on the PSSA8.

The sample size deteriorated at the first step, when correlating reading grades and the fourth grade JSSA reading category. At that step, the valid cases were composed of 185 of the students. Then in the second step, when the variables included sex of the student, categorization with disability, and PSSA8, only 101 of the 111 students with disability placement could be used in the analysis. Later, when analysis considered the variables of sex of the student, categorization with disability, fourth- through seventh-grade JSSA categories, and PSSA8 scaled score, 69 students composed the valid cases.

In the current study, curricular probes were chosen using the published instructional assessment rules of that time. The district chose authentic assessments from the local basal curriculum to serve as protocols for the assessment of reading. Deno and Mirkin (1977), Shinn (1989a, 1989b), Hintze and Shapiro (1997), and Fuchs and Fuchs (1992) all concurred that assessment and instructional planning within the curriculum was linked to better progress data which potentially guided more effective reading instruction. Multiple readability formulas were applied, unique or proper words were limited, passages with pictures or graphics necessary to understanding were eliminated, and the presentation of the probes was sized and printed to resemble the typical reading materials presented to those students (Ficco, 1999). The probes were then scored using the wholistic reading rubric employed by the Pennsylvania Department of Education. Those rubric scores or categorizations of performance represented only one type of CBA used within this district.

All students were assessed within their reading classrooms at the elementary level and within an English or Language Arts placements at the middle school level, at a frequency of three times annually (Ficco, 1999). As recommended by Fuchs, Tindal, and Deno (1984), the district established consistent procedures for these repeated measures. The assessments were given on the same schedule throughout the district, and fidelity of administration was monitored by principals and reading specialists. Gickling proposed that an accurate assessment of the students' instructional level was necessary to plan and monitor the students' progress in reading (Gickling & Rosenfeld, 1995; Rosenfeld & Kuralt, 1994). These district students were assessed on their registered grade level rather than on the level of reading instruction which they received.

The study selected student assessments from the spring (third annual) assessment for this analysis, as those formative assessments were scheduled at very nearly the PSSA reading schedule for each school year (Ficco, 1999). This coordination of schedule was hoped to limit effects from seasonal changes in student motivation or performance.

The district then scored those CBA assessments using the Pennsylvania Department of Education reading rubric. This rubric was designed for the scoring of the PSSA reading assessment, yet demanded both writing and reading skills for accurate completion and to earn an at-least proficient rating. Further, the written response to this reading sample was prepared over two days, with a draft prepared on the first and a clean copy of the second day. The extended time for response may have given advantage to better organized students, while students with poor short-term memory or organizational skills may have demonstrated greater difficulty over this lengthy process (Christenson, Ysseldyke, & Thurlow, 1984; Shapiro, 1990, 1992).

The analysis of curriculum-based data was completed in two steps. The first analysis considered sex of the student, the CBA reading category scores, and whether the students were educationally disabled. The second analysis also considered sex and the CBA reading category scores, while then considering the students' specific educational disabilities. These multiple regression analyses were designed to optimize the sample size in the first step and to yield additional data about sex, CBA results, and specific disabilities and their relationship to reading performance on the PSSA in the second step. Unfortunately, the sample sizes in the second step were too small for reliable conclusions to be drawn, although the results suggested that speech and language impairment was a predictor of the PSSA reading criterion score. The results for students with educational disabilities further suggested that the district CBA instrument or process did not provide a valid vehicle to monitor their educational progress or to predict their eventual performance. General opinion in the research on students with disability clearly recommended CBA as a more reliable and useful benchmarking or monitoring tool.

### Summary

The district was an early practitioner of benchmarking student reading performance in Pennsylvania. They laudably included all students in a CBA reading assessment at a frequency of three times per year, including those students with educational disabilities. However, the choice of instrument followed a socially or politically regulated process, rather than being driven by or including expert advice. The district stakeholders were familiar with CBA and were motivated to choose passages from the traditional curriculum which met the criteria familiar to them. The district chose to emulate the categorical or wholistic rubric scoring of the PSSA (i.e., below basic,

basic, proficient, and advanced); but they did not have the expertise or tools to also create a numeric scale which was balanced across the categories or matched to the PSSA scaled scores.

When multiple regression analysis was used to examine the results of adding the district CBA datum to the predictor variables of sex of the student and educational disability, only educational disability was a significant predictor of the PSSA8 reading scaled score ( $p < .01$ ). The district CBA datum was analyzed in a multiple regression with other predictors, and was not a significant ( $p > .05$ ) predictor of the PSSA8 reading scaled scores.

Many possible explanations for this result exist. The district had created a CBA system which had the usual technical weakness of such systems, rather than implementing the technically superior CBM benchmarking. Further, the assessment was implemented at instructional levels which were not a match to some students' actual reading instruction. Then, the assessment was not regularly monitored across settings, as staff could not be dedicated to that task. For some students, testing accommodations for the assessment were not matched to those required by the students' Individual Educational Programs. Additionally, the design of the JSSA required a two-day long period, which may have been an advantage to some students, while creating an impediment to others.

### *Research Question Three*

Were the demographic factors of the sex and educational disability of the student significant predictors of the students' eighth grade PSSA reading scaled scores?

In question three, the demographic characteristics of the student cohort (i.e., sex of the student, categorization as having an educational disability, as well as the specific type of educational disability or special education placement) were examined to find any relationship between these characteristics and the students' scores on the PSSA eighth-grade reading exam. Multiple regression analyses were conducted to examine the ability of these predictor variables to predict the criterion, the students' PSSA reading score in eighth grade. This analysis was completed in two steps: the first analysis considered simply sex of the student and whether the students qualified as having an educational disability; while the second analysis considered the existing specific educational disability. These multiple regression analyses were designed to optimize the sample size in the first step, and to yield additional data about specific disabilities and their relationship to reading performance on the PSSA in the second step.

#### Educational Disability

When examined with curriculum-based as well as demographic data, the categorization with an educational disability was a significant predictor of the students' PSSA reading scaled scores, with an alpha of  $p < .001$ . The students' status of having an educational disability was a significant predictor of those students' performance on their eighth-grade PSSA assessment, and as a subgroup, the students with disability scored less well on the PSSA8 reading. As discussed earlier, possible obstacles for these students included unreliable instruction and assessment match; a lengthy procedure for testing; a need for unavailable accommodations; and the need to produce a proficient written response to the CBA reading prompt.

The Greater Johnstown School District was very hopeful that curriculum-based data could guide reading instruction more effectively for all students than the published, norm-referenced achievement tests (PNAT) of reading had done. The relationship between the district student scores on a PNAT and the reading PSSA scores was not strong (Ficco, 1999). Marston (1989) and Marston, et al. (1986) reported that standardized achievement tests had provided little instructional decision-making data. The district's staff concurred with Marston's review of reading instructional needs, including the immediate need for data on specific skill performance and progress. The use of curriculum-based assessments was uniformly applied across the district in order to serve two purposes: curriculum-based assessment was believed to be a great improvement over the previous data from PNAT assessments (Shapiro, 1990) in guiding instruction, and the district planned to follow the data to determine whether the CBA data could predict later PSSA scores. A final attempt to examine the questions regarding CBA, sex of the student, and PSSA8 including specific disability categories yielded only 14 valid cases; and those results could only be seen as suggestive. In this second portion of this analysis, the students' specific disability categories were examined as predictors of the students' eighth-grade PSSA reading scaled scores. The existing categories included emotional disturbance, specific learning disability, alternative education placement, other health impairment, speech and language impairment, and mental retardation. In this analysis, the sample size was too small for results to be considered except as suggestive. Still, the specific disability of Speech and Language Impairment ( $p < .003$ ) was indicated as a significant predictor of those PSSA8 reading results. In related discussion, Salvia and Hughes (1989) stated that the level of reading comprehension cannot exceed the

students' general language competence. The student categorization with educational disability was a significant factor ( $p < .001$ ); however, the sample sizes in the second step were too small for reliable conclusions to be drawn regarding specific disabilities.

#### Sex of the Students

At the time of this study's inception, it was generally purported that male students might more often demonstrate reading difficulty or might need an increased length of instruction to learn to read, which was later confirmed by Snow, Burns, and Griffin (1998). The ratios of males to females who experienced reading difficulties was reported as varying from 2:1 to 5:1. Additionally, those students who were determined to be at-risk for reading difficulty did not always improve reading performance when specific skills deficits were addressed (Adams, 1990; Snow, et al., 1998). However, specific phonological instruction commonly yielded greater progress (O'Connor, et al., 1993). For these reasons, the author hypothesized that males might respond differentially better to the *Success for All* curriculum, with its explicit and phonological instruction and chunked skills presentation and practice. When the students' eighth-grade PSSA scores were predicted using those students' CBA reading performance, sex, curriculum, and status as educationally disabled, multiple regression indicated no significant prediction using sex of the student ( $p > .05$ ).

When the students' eighth-grade PSSA reading scaled scores were predicted using those students' sex, their fourth- through seventh-grade CBA performance categories, and status of having an educational disability, multiple regression demonstrated a highly significant alpha ( $p < .001$ ) only for students having an educational disability. Sex of the student was not a significant predictor of eighth grade PSSA reading scaled scores.

#### *Research Question Four*

Did the students' JSSA4 through JSSA7 reading scores predict their eighth grade PSSA reading scaled scores when examining students who participated in either the John Hopkins *Success for All* reading curriculum or the *Houghton-Mifflin* basal reading curriculum?

Question four examined differentiated curricula as a factor which might influence student reading performance on the eighth-grade PSSA. The two curricula implemented within the district were the *Success for All* reading program published by Johns Hopkins University and the *Houghton-Mifflin* basal reading program. All students registered in each building received one prescribed reading curriculum throughout their elementary program.

When selecting the protocols for district benchmarking, the two different reading curricula used within the district posed a question. From what source should the prompt or reading sample be drawn? It seemed that the *Houghton-Mifflin* basal program offered many samples which met the criteria for general reading probes. Fuchs and Deno in 1992, and Hintze, Shapiro, and Lutz in 1994 each examined that very question. The results of their research demonstrated better sensitivity to student progress using traditional probes from the reading basal rather than nontraditional or literature-based probes. Although the district had not benefited by the results of these studies in their decision-making, the choice to use prompts from the basal series was further affirmed by Hintze and Shapiro in 1997.

When the students' eighth-grade PSSA reading scaled scores were predicted using those students' sex, their fourth- through seventh-grade CBA performance categories,

curriculum, and status of having an educational disability, multiple regression demonstrated a highly significant alpha ( $p < .001$ ) only for students having an educational disability. When the two curricula were examined in the array of all predictors, curriculum was not a significant predictor of the eighth grade PSSA reading scaled scores ( $p > .05$ ).

### Implications of the Study

The current study examined the practices of CBA, benchmarking, and educational problem-solving used to direct and drive instructional practice, and the ability of the CBA data to predict reading performance on a later assessment. Each practice has been refined since the collection of this data, and implications and recommendations for future practice are relevant for review.

The use of CBA as a predictor of critical skill development certainly affects student success in venues other than academics, as discussed by Wicks (1987). Still, the more common uses of curriculum-based assessment are for guiding instruction and for predicting high stakes or other summative assessments. Curriculum-based data is known to accurately measure student progress in academic skills areas, and is able to focus instruction for enhanced skills-based performances (Shapiro, 1990, 1992). The local district practice reviewed in this study included benchmarking curriculum-based data to guide instructional practice and to make general education reading placement decisions.

CBA is an authentic measurement system which uses student performance in the curriculum itself to determine what portions of the material and concepts students have already learned within the curriculum, to pace instruction, and to measure progress continually across short-term intervals. However, all district students were assessed at

their current assigned or matriculated grade level, although many students were instructed outside of that grade level at the instructional level assigned by the Instructional Support Team or *Success for All* periodic assessment. The district had failed to choose the instructional level for assessment, thus limiting the match between curriculum and assessment. Further, those students being instructed below grade level might well have expressed their frustration with the assessment through limited motivation or effort on a reading sample which may have seemed unduly difficult (Marston, 1989). Students receiving above-grade-level instruction may also have been less able to demonstrate accurate progress, as the challenge level of the reading sample may have been too low to be of interest (Burns, 2001), or did not access their above-grade-level level reading skill. Salvia and Hughes (1989) discussed the ability for effective assessment to limit the contribution of associated skills, while misaligned assessment surely risked increasing those contributions.

The local school district CBA assessment was not used to evaluate students for disability determination. However, the progress of students within the specific reading curriculum was certainly a factor in those students' performance on the standardized achievement tests which were used to determine whether a significant discrepancy between ability and achievement scores existed. The JSSA might be seen as a protective device leading to least restrictive placement, as it provided performance information which demonstrated instructional progress. Otherwise, the JSSA might be seen as an unfair measure which assessed content and curriculum not well matched to instruction for students who received out-of-grade-level reading instruction. Certainly, Gickling, Fuchs, and Deno each addressed the need for instructional and assessment level match. Lastly,

Cohen (1987) addressed the confused teacher focus and lessened teacher effectiveness which occurred when instruction and assessment were misaligned.

To encourage this educator focus, the use of CBA required professional education, well-designed and carefully implemented probes, and systematic supports to maintain fidelity and to share instructional decision-making. As assessment and instructional systems changed and new instructional skills were required, educational staff required assistance to implement these skills within their classroom settings. Particularly in the guided practice to independent practice step, teachers benefited from observations of the fidelity or accuracy of their implementation. Implementation at the building level and monitoring of fidelity of practice was delegated to building principals at the district, who sometimes had insufficient time and support to practice this monitoring function (Ficco, 1999). It is recommended that school psychologists serve as direct providers for this service or train teams of master-teachers or experienced co-teachers to help to assess and implement practice during the implementation of benchmarking with CBA (Eckert & Shapiro, 1999).

Instructional and curricular change also required mandatory professional education, which was limited by contractual time. One purpose of collecting curriculum-based data was to inform instruction. In this process, areas of curriculum may be identified as being poorly sequenced, as being poorly matched to the educational standards, or as needing an alternate pacing. The school psychologist, in collaboration with the curriculum director, principals, and other educational specialists, can problem solve, attempt changes, and evaluate the results of those changes. When a change in curricula is required, the school psychologist consults with other stakeholders in the

process, typically providing effectiveness data and meta-analysis data regarding the curriculum being considered. Last, the school psychologist could have assessed and presented to the district staff needs assessment results regarding their professional education, fidelity of implementation, and rates of effectiveness across different settings, sites, and student demographic groups. The local school psychologists consulted only in designing professional education at the district studied (Ficco, 1999). Given the opportunity, they would have been able to: recommend curriculum sequencing which better matched the PSSA task demands; facilitate the collaborative decision-making required for curriculum choice and design; provide research on best practice for benchmarking recommending CBM; and design implementation supports and fidelity monitoring to ensure greater uniformity during assessment. School psychologists and other expert or university advisors would have improved the process and the use of results to guide instruction.

Another needed focus for this professional education planning was the visual aspect of graphing progress, as it was well known to accrue to students' motivation and progress. School psychologists typically sought methods and technologies which included this beneficial process. When educational terminology may obfuscate, the clarity of a visual representation of progress was often welcomed. Teams of parents, students, community members, board members, or other non-educational stakeholders may have perceived a leveled playing field when data was examined visually. Thus, the presentation of data in a visual manner would have encouraged participation of all team members; may have increased student progress; and allowed the school psychologist and other professional educators to facilitate the decision-making process. Visual presentation

was not applied in the district's benchmarking process (Ficco, 1999); and the implementation and provision of visual graphing might have differentially benefited disaggregated groups of students at Johnstown, especially those students with educational disability (Shinn, et al., 1990; Shinn & Hubbard, 1992; Shinn, et al., 1998).

If visual presentation had been incorporated in the design of benchmarking at Johnstown, students, educators, parents, and the community would have benefited. Students would have clearly seen and taken greater ownership of their own educational progress, increasing their motivation to learn. The ability of educators to make instructional decisions and to strategically design instructional groupings would have been increased, while parents and community members who viewed results in this explicit format could collaborate on instructional decision-making to a greater extent and with increased confidence. In all, the visual presentation held promise to improve instructional results across curricula and student categories, to increase team effectiveness, and to communicate data in an easily understood format, so that needed instructional or curricular change could occur more quickly and easily.

The limited contractual time for mandated professional education necessitated the clear prioritization and efficient provision of this training. Other administrators and school psychologists collaborated on determining those priority areas and school psychologists provided substantial assistance in the design of those training opportunities. Two much-neglected elements in professional education were the assessment of needs and the evaluation of results, both of which can be designed or provided by school psychologists. Neither function was delegated to school psychologists at Greater Johnstown, nor were needs assessments regularly employed (Ficco, 1999), although these

needs assessments would have clarified misunderstandings and recommended goals for professional education. School psychologists could additionally have presented the assessment results to various stakeholders, encouraging strategic action through the professional education program.

Data review for administrative and management function is a critical calendar-based activity for any school entity. Fall baseline-level assessments are required for many grants. Winter staffing and budgetary planning requires an assessment of the various disaggregated groups and their needs for the next school year. Spring outcome-levels were required as post-measures for federal program planning, grants, and curricular planning. Beyond these simple presentations of data lay the more complex data exploits. Administrators assessed the quality of instruction and results of education across programs, buildings, disaggregated groups, and grade-levels, and problem-solved to improve results (Shinn, 1998). The local school district had just begun to discuss some of these data-based activities, but had not yet implemented them according to Ficco (1999). Again, the district practice was limited by both the recent development of these more advanced data practices, and by the limited collaboration with school psychologists and other educational research staff. A trend toward research-based instruction and interventions and the increasing provision of inclusive education services had shifted the role of school psychologist toward being the facilitator of problem-solving and child-find teams, manager of school-wide progress data, and the provider of consultation services in a variety of venues. The school psychologist's training and expertise in the fields of learning, behavior, individual differences, counseling, and systems theory prepared those psychologists to assess academic and behavioral needs, analyze and support systems

change, assess and evaluate instruction and learning, and coordinate services, data, and personnel efforts. It was necessary for school psychologists to collaborate and consult on district assessment activities.

In any future studies, it is essential that curriculum-based measurement, aligned with student instructional levels, be used. Additionally, an economy of time, instruction, and staff cost would be best served by the use of school psychologists and other expert or university research staff to collaborate, research best practices, and consult in the design of assessments, professional education, data-based decision-making, and fidelity monitoring.

#### Recommendations for Future Research

Curriculum-based measurement, used as benchmarking data, is a focus for future research. It is hoped that these studies would yield data indicating that the briefer and less expensive CBM measure would yield both formative instructional guidance and predictive data for the state-directed high stakes testing. Additionally, it would be valuable to know whether CBM which is used as continuous progress monitoring provides adequate formative instructional guidance and predictive data for the state-directed high stakes testing, when used without a separate proprietary or state education department propounded benchmark measure. The collaboration of university experts and state department specialists and administrators would guide both the use of benchmarking processes and the continuing design and uses of the high stakes measures of proficiency in the educational standards (Reschly & Grimes, 1991; Shephard, 1983).

As the sample size for specific disabilities is small in many individual districts, future research can address the relationship between students with specific disabilities,

across several districts, and their CBM progress data as compared to their state level high stakes testing results. It may be that students with disabilities are more likely to demonstrate growth and proficiency on the brief and concrete CBM assessment than on the lengthy and more complex problem-solving format of the high stakes test.

Future research should reach downward, to examine younger students using CBM. Several possible topics suggest themselves. An examination of student baseline skills as related to eventual high stakes testing scores in upper elementary school would be valuable. As we know through Lyon (1994) and Kamenui and others, students who have reading difficulty in fourth grade usually continue to have this difficulty throughout their school years, beginning to examine whether this is an issue of slow progression or growth versus limited or inappropriate early literacy instruction would be worthwhile.

As school districts recognize both the predictive and the formative uses of CBA, it is likely that they will need to examine more than one process. While benchmarking at grade level serves the predictive purpose; progress monitoring at instructional level will be needed to serve as a formative assessment for many regular, gifted, and special education students. Future research should focus on the relative frequency of these two assessment processes which best serve and guide instruction for both advanced and challenged students. The progress of students within each process should be compared, contrasted, and examined across a variety of disaggregated demographic subgroups of students as well.

Because instructional time is a priority and instructional focus a concern for every educator, districts should consider limiting the use of alternate assessments chosen by individual teachers. Many historical or web-based assessment instruments will not

provide valid and reliable data, and educators cannot afford to be misdirected by spurious instruments. Future research should address the number, frequency, and types of assessments which best balance dedicated instructional time and effective instructional data. It would also be useful to assay the instruments in use in a variety of subject- or content-specific classrooms.

An examination of the practices of needs assessment and fidelity monitoring, and their effectiveness at encouraging educators to practice benchmarking is imperative. Future research should address the stakeholders who participate in needs assessment and the relationship between those stakeholders who participate and successful implementation of the benchmarking process. Similarly, the relationship between the particular staff which provide fidelity monitoring and student progress, the reliability of CBA or CBM data, and accurate implementation of data rules would be useful to examine. A full understanding of the purposes of assessment, and the usefulness and implementation of that data to guide instruction depend on both practices.

An examination of the predictive validity of various CBA measures and CBM, as well as specific PNAT scores (e.g., science) should continue as the high stakes assessments required by various states continue to develop. Especially as these high stakes tests begin to measure additional content areas, we must determine the multiple curricular skills which allow proficient performance. Future research should address the validity of various curriculum-based assessments of the expanding subject areas now included in high stakes testing (e.g., science). As these high stakes tests and additional content-specific measures become the required criteria for high school graduation, every

student, educator and district will need formative assessments to guide instruction and to prepare for this culminating activity.

### Limitations of the Study

Internal threats to the validity of these results include relationships between the various predictive factors examined in this study. As interactive factors could have contaminated the clarity of the results, the choice of multiple regression analysis was made to limit these interactive factors.

In designing the study, the concurrent validity of the district curriculum-based assessment as a measure of student reading skill and guide for instruction, and the sensitivity of that CBA assessment as a predictor of the Pennsylvania reading assessment were considered. First, the CBA measure was correlated to student reading grades, the results of which indicated a very significant relationship. However, two factors may have limited the ability of CBA data to serve the previously described instructional purposes. The CBA probes were more variable in readability level than were similar CBM probes, and, unlike the unique “cold read” materials in CBM probes which were not a part of the curriculum (Fuchs & Deno, 1994), some students may have been inadvertently exposed to them although the probes were not used in grade-level classroom instruction.

Additionally, there was a mismatch in curriculum and assessment levels for some students. While all students were assessed at their matriculated and registered grade level; some students were instructed either below or above that grade level in reading. This mismatch was more likely to occur for students with disabilities and for students within the *Success for All* curriculum (Ficco, 1999), as the SFA curriculum typically assigned

reading groups based upon reading skill level. This instruction and assessment mismatch surely limited the accuracy of reading skill measurement for some members of the cohort.

In this study, the longitudinal data on reading skill for that cohort of students was examined for its ability to predict those students' PSSA scores at a later grade level. In Greater Johnstown, the students were assessed without accommodation on the district measure, and no data were available to determine which accommodations were provided on the state-wide assessment (Ficco, 1999). General research from the field indicated that CBM could assist in predicting performance on high stakes testing, but sometimes less accurately for students with disability than for the general student population, as these students required great fidelity of accommodations in order to yield accurate performance levels (Christenson, Ysseldyke, & Thurlow, 1984; Ysseldyke & Christenson, 1987a, 1987b). Because the JSSA was presented without the individual accommodations which were used on the PSSA for district students with disability, it is possible that the study results indicating a significant prediction of the criterion from student categorization with educational disability reflect the above research findings. As this may be so, the provision of strategically designed reading accommodations should be presented identically per student across the district-level and the state high stakes assessments.

Other types of CBA in district use during the study included mastery testing, performance sampling, informal inventories, and other locally designed systems, which may or may not measure the same skills and type or amount of progress across students, grade-levels and systems. Because the alternate CBA measures were used informally across the district, the JSSA assessment may not have been considered to be the primary

benchmark of progress by some educators. These educators may then have been less assiduous in implementing the JSSA or in applying the instructional recommendations which ensued from the benchmark assessments. As the only CBA measure which was validated as sensitive to growth, brief in sampling time, and reliable across students, grade-levels and systems was curriculum-based measurement (CBM); the sole choice of CBM benchmarking, rather than a variety of CBA assessments as reading measures, would have increased the usefulness and clarity of reading data collected at the district. In addition, the data collection and scoring would have been considerably simpler and less time consuming when training staff, implementing, and monitoring (Shinn, 1989a). The use of CBM would have allowed a team to share meaningful measures, to graph and analyze data visually, and to compute statistically valid changes and differences. Progress could then be examined both visually and by using data-based rules; and teams could have decided whether that progress was sufficient or whether it indicated a need for instructional change.

Additionally, in those years before 2001, the state assessment of reading was improving the link between reading standards and measurement of reading skills (Murray, 2004). Cut scores were initially applied to the PSSA scaled scores in 2001; and Masters (2004) determined the 2004 and subsequent PSSA measures as reliable and efficient measures of the standards. It is possible that the 2002 PSSA8 scaled scores also suffered from some technical weakness.

Further, the PSSA and JSSA reading assessments demanded both writing and reading skills for accurate completion and to earn an at-least proficient rating. Further, the written response to the reading sample was prepared over two days, with a draft

prepared on the first and a clean copy on the second day. The extended time for response may have given advantage to better organized students, while students with poor short-term memory or organizational skills may have demonstrated greater difficulty over this lengthy process.

Historically, teachers may have determined progress and defined problems by using classroom observations and classroom samples, but all current trends demand a research-based system. Education has followed a business model in demanding an accounting for both the effectiveness of instruction and the cost of staff. This accounting for instructional efficiency uses salaries and benefits, instructional time, and/or the cost and effectiveness of curricular materials. Two major implications of these economic trends are technology needs and professional education requirements.

The needs for technology and professional development are intimately related as a district or building designs systems to implement, review, store and evaluate the longitudinal data and short-term formative assessments of students. These assessment systems must be minimally invasive to instruction, while providing accurate information and concrete guidance to educators. Time in student instruction must be maximized, demanding high organizational skills from teachers and excellent technological supports. A well-designed problem-solving system will focus on and evaluate a few skills or behaviors across multiple measures, directing instructional practice and interventions to yield the students' best possible progress in the skill areas.

The sheer amount and variety of forms and storage locations for this educational datum, precluded the preparation and interpretation of data by any one individual. Correspondingly, this large amount of data was necessary to a wide array of educational

personnel beyond the school psychologist. Kovaleski (2006) suggested that teams of school staff provide some of this effort and function, and that this assessment team be trained and supported by school psychologists. Assessment teams can meet the need for fidelity and reliability measures across assessors, and can design the process for efficiency, and provide support for assessments across grade-levels, buildings, or settings. The teams could then provide direct assessment services and the interpretation and support which educators required to understand and employ the data. School psychologists would continue to provide professional education, consultation, and support to these teams as they applied their skills and managed their many functions. The Johnstown school district implemented the assessment team practice, using reading specialists and a scoring team to assist building staff, yet did not use school psychologists to design or support the process (Ficco, 1999). The collaboration of school psychologists would have allowed for a research base in best practices to guide the assessment team, and should be implemented.

In this study, the school district staff chose the curriculum-based assessment system most familiar to them; and one that related well to other professional education initiatives implemented at that time (Ficco, 1999). Curriculum-based measurement would have decreased instructional time spent, improved direct measurement of reading skills, and allowed more time for professional education regarding differentiated instruction.

### Summary

The population described in this study comprised an entire grade level cohort of students in the Greater Johnstown schools in Pennsylvania. This cohort included 48% males, 21% students with an educational disability, and 22% students with ethnic

minority status. All students were assessed three times per year using a CBA reading measure, from fourth through seventh grade.

This school district applied curriculum-based assessment benchmarking to guide instructional practice and to predict student scores on the state-wide assessment of reading skill. Curriculum-based assessment was a useful measure for both purposes, yet curriculum-based measurement would have provided more accurate and efficient assessment and prediction.

Multiple regression analysis was used to analyze the following predictive variables as they related to the criterion, the PSSA reading scaled score in eighth-grade. The study analyzed two reading curricula, sex of the student, curriculum-based assessment scores, and educational disability categories as predictors of the scores on that state-wide exam. When considered in an array with the curriculum-based data, only educational disability proved a significant factor ( $p < .001$ ).

## REFERENCES

- Aaron, P. G. (1995). Differential diagnosis of reading disabilities. *School Psychology Review, 24*, 345-360.
- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: The MIT Press.
- Allinder, R. M., & BeckBest, M. A. (1995). Differential effects of two approaches to supporting teachers' use of curriculum-based assessment. *School Psychology Review, 24* (2), 287-298.
- Allinder, R. M., Dunse, L., Brunken, C. D., & Obermiller-Krolikowski, H. J. (2001). Improving fluency in at-risk readers and students with learning disabilities. *Remedial and Special Education, 22*(1), 48-54.
- Allinder, R. M., Fuchs, L. S., & Fuchs, D. (1998). Curriculum-based measurement. In H. B. Vance (Ed.), *Psychological assessment of children: Best practices for school and clinical settings* (2<sup>nd</sup> ed., pp. 106-132). New York: Wiley.
- Alsalam, N., Ogle, L. T., Rogers, G. T., & Smith, T. M. (1992). *The condition of education* (NCES 92-096). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.
- American Psychological Association, American Educational Research Association, and National Council on Measurement in Education. (1985). *Standards for educational and psychological tests*. Washington, DC: American Psychological Association.
- Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22*, 261-295.
- Anderson, R., Hiebert, E., Scott, J., & Wilkinson, I. (1985). *Becoming a nation of readers: The report of the commission on reading*. Washington, DC: National Institute of Education.
- Anderson, A. B., & Stokes, S. J. (1984). Social and institutional influences on the Development and practice of literacy. In H. Goelman, A. Oberg, & F. Smith (Eds.), *Awakening to literacy*. Exeter, NH: Heinemann.
- Arias, M. B. (1986). The context of education for Hispanic students: An overview. *American Journal of Education, 26*-57.
- Atwell, N. (1991). *Side by side: Essays on teaching to learn*. Portsmouth, NH: Heinemann Educational Books, Inc.

- Bagnato, S. J., & Neisworth, J. T. (1991). *Assessment for early intervention*. New York: The Guilford Press.
- Baker, E. L., O'Neil, H. F., Jr., & Linn, R. L. (1993). Policy and validity prospects for performance-based assessment. *American Psychologist, 48*, 1210-1218.
- Bashir, A. S., & Scavuzzo, A. (1992). Children with language disorders: Natural history and academic success. *Journal of Learning Disabilities, 25*, 53-65.
- Behrmann, M. M. (1998). Assistive technology for young children in special education. In C. Dede (Ed.), *Learning with technology* (1998 ACSD Yearbook). Alexandria, VA: Association for Supervision and Curriculum Development.
- Bender, W. N. (2002). *Differentiating instruction for students with learning disabilities*. Thousand Oaks, CA: Corwin Press.
- Bishop, D. V. M., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders, and reading retardation. *Journal of Child Psychology and Psychiatry, 31*, 1027-1050.
- Blachman, B. (Ed.). *Cognitive and linguistic foundations of reading acquisition: Implications for intervention research* (pp. 67-92). Hillsdale, NJ: Erlbaum Associates.
- Bloom, B. (1964). *Stability and change in human characteristics*. New York: Wiley.
- Borman, G. D., & Hewes, G. M. (2002). The long term effects and costs effectiveness of Success for All. *Educational Evaluation and Policy Analysis, 24*(2), 243-266.
- Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2002). *Comprehensive school reform and student achievement: A meta analysis*. Baltimore: Center for Research on the Education of Students Placed At Risk, Johns Hopkins University.
- Bowers, P., & Swanson, L. B. (1991). Naming speed deficits in reading disability: Multiple measures of a singular process. *Journal of Experimental Child Psychology, 51*, 195-219.
- Bradley-Klug, K. L., Shapiro, E. S., Lutz, J. G., & DuPaul, G. J. (1998). Evaluation of oral reading rate as a curriculum-based measure within literature-based curriculum. *Journal of School Psychology, 36*(2), 183-197.
- Bredenkamp, S., & Copple, C. (Eds.) (1997). *Developmentally appropriate practice in early childhood programs*. Washington, DC: NAEYC.

- Burns, M. K. (2001). Measuring sight-word acquisition and retention rates with curriculum-based assessment. *Journal of Psychoeducational Assessment, 19*, 148-157.
- Burns, M. K., Tucker, J. A., Frame, J., Foley, S., & Hauser, A. (2000). Interscorer, alternate-form, internal consistency, and test-retest reliability of Gickling's model of curriculum-based assessment for reading. *Journal of Psychoeducational Assessment, 18*, 353-360.
- Byrne, B., & Fielding-Barnsley, R. (1989). Phonemic awareness and letter knowledge in the child's acquisition of the alphabetic principle. *Journal of Educational Psychology, 81*(3), 313-321.
- Byrne, B., & Fielding-Barnsley, R. (1991). *Sound foundations*. Artarmon, New South Wales, Australia: Leyden Educational Publishers.
- Byrne, B., & Fielding-Barnsley, R. (1993). Evaluation of a program to teach phonemic awareness to young children: A 1-year follow-up. *Journal of Educational Psychology, 85*(1), 104-111.
- Byrne B., & Fielding-Barnsley, R. (1995). Evaluation of a program to teach phonemic awareness to young children: A 2- and 3-year follow-up and a new preschool trial. *Journal of Educational Psychology, 87*, 488-503.
- Caine, R. N. B., & Caine, G. (1997). *Unleashing the power of perceptual change: The potential of brain-based teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Campbell, J. R., Donahue, P. L., Reese, C. M., & Phillips, G. W. (1996). *NAEP 1994 reading report card for the nation and the states*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.
- Castellani, J., & Jeffs, T. (2001). Reading and writing strategies using technology. *Teaching Exceptional Children, 33*(5), 60-67.
- Catts, H. W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech and Hearing Research, 36*, 948-958.
- Catts, H. W., Fey, M., & Tomblin, B. (1997). *The language base of reading disabilities*. A paper presented at the meeting of the Society for Scientific Studies of Reading, Chicago.
- Chittenden, E. (1991). In V. Perrone (Ed.), *Expanding student assessment U.S.A.* Association for Supervision and Curriculum Development.

- Christensen, S. L., Ysseldyke, J. E., & Thurlow, M. L. (1989). Critical instructional factors for students with mild handicaps: An integrated review. *Remedial and Special Education, 10*, 21-31.
- Cohen, S. A. (1987). Instructional alignment: Search for a magic bullet. *Educational Researcher, 16*, 16-20.
- Coles, G. (2004). Danger in the classroom: 'Brain glitch' research and learning to read. *Phi Delta Kappan, 85*, 344-351.
- Commission on Reading. (1985). *Becoming a nation of readers: The report of the commission on reading*. Washington: The National Institute of Education.
- Cooke, N. L., Heward, W. L., Test, D. W., Spooner, F., & Courson, F. H. (1991). Student performance in the classroom: Measurement and evaluation of student progress. *Teacher Education and Special Education, 14*(3), 155-161.
- Cooper, J. D. (1986). *Improving reading comprehension*. Boston: Houghton Mifflin Company.
- Cornwell, A. (1992). The relationship between phonological awareness, rapid naming, and verbal memory to severe reading and spelling disability. *Journal of Learning Disabilities, 25*, 532-538.
- Creswell, J. W. (1994). *Research design: Qualitative & quantitative approaches*. Thousand Oaks, CA: Sage Publications, Inc.
- DeFries, J., Fulker, D., & LaBuda, C. (1987). Evidence for a genetic aetiology in reading disability of twins. *Nature, 329*, 537-539.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232.
- Deno, S. L. (1986). Formative evaluation of individual student programs: A new role for school psychologists. *School Psychology Review, 15*, 353-374.
- Deno, S. L. (1987). Curriculum-based measurement. *Teaching Exceptional Children, 20*, 41.
- Deno, S. L. (1989). Curriculum-based measurement and special education services: A fundamental and direct relationship. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 1-17). New York: Guilford.
- Deno, S. L. (1992). The nature and development of curriculum-based measurement. *Preventing School Failure, 36*(2), 5-10.

- Deno, S. L., & Fuchs, L. S. (1987, April). Developing curriculum-based measurement systems for data-based special education problem-solving. *Focus on Exceptional Children, 19*, 1-16.
- Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30*, 507-524.
- Deno, S. L., Fuchs, L. S., Tindal, G., & Wesson, C. L. (1987). Curriculum-based measurement: Program development, graphing performance, and increasing efficiency. *Teaching Exceptional Children, 20*, 40-47.
- Deno, S., Marston, D., Mirkin, P., Lowry, L., Sindelar, P., & Jenkins, J. (1982). *The use of standard tasks to measure achievement in reading, spelling, and written expression: A normative and developmental study* (Research Report No. 87). Minneapolis, MN: University of Minnesota Institute for Research on Learning Disabilities.
- Deno, S., Marston, D., Shinn, M., & Tindal, G. (1983). Oral reading fluency: A simple datum for scaling reading disability. *Topics in Learning and Learning Disability, 2*(4), 53-59.
- Deno, S. L., Marston, D., & Tindal, G. (1985). Direct and frequent curriculum-based measurement: An alternative for educational decision-making. *Special Services in the Schools, 2*, 5-28.
- Deno, S., & Mirkin, P. (1977). *Data-based program modification: A manual*. Reston, VA: Council for Exceptional Children.
- Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49*, 36-45.
- Dettmer, P. A., Dyck, N. T., & Thurston, L. P. (1996). *Consultation, collaboration, and teamwork for students with special needs*. Boston: Allyn and Bacon.
- Dunn, E. K., & Eckert, T. L. (2002). Curriculum-based measurement in reading: A comparison of similar versus challenging material. *Psychology in the Schools, 39*, 24-46.
- Eckert, T. L., & Shapiro, E. S. (1999). Methodological issues in analog acceptability research: Are teachers' acceptability ratings of assessment methods influenced by experimental design? *School Psychology Review, 28*(1), 5-16.
- Edyburn, D., & Majsterek, D. (1993). Technology applications for individuals with LD: What can we say today? *LD Forum, 19*(1), 3-5.

- Edyburn, D. L. (1994). An equation to consider: The portfolio assessment knowledge base + technology = The Grady Profile. *LD Forum*, 19(4), 35-38.
- Enggren, P., & Kovaleski, J. (1996). *The instructional support system of Pennsylvania: Instructional assessment manual*. King of Prussia, PA: Prise EISC.
- Espin, C. A., & Deno, S. L. (1993). Content specific and general reading disabilities of secondary-level students: Identification and educational relevance. *Journal of Special Education*, 27, 321-337.
- Espin, C. A., & Tindal, G. (1998). Curriculum-based measurement for secondary students. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 214-253). New York: Guilford.
- Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3<sup>rd</sup> ed.) (pp. 235-271). Washington, DC: The American Council on Education and the National Council on Measurement in Education.
- Felton, R. H. (1993). Effects of instruction in the decoding skills of children with phonological-processing problems. *Journal of Learning Disabilities*, 20, 583-589.
- Ficco, R. A. (1999). A personal perspective and interview, regarding the JSSA. Johnstown, PA: Personal interview, prior to retirement.
- Fish, L. J. (1988). Why multivariate methods are usually vital. *Measurement and Evaluation in Counseling and Development*, 21, 130-137.
- Flynn, J. M., & Rahbar, M. J. (1994). Prevalence of reading failure in boys compared to girls. *Psychology in the Schools*, 31, 66-71.
- Fuchs, D., Fuchs, L. S., Benowitz, S., & Berringer, K. (1987). Norm-referenced tests: Are they valid for use with handicapped students? *Exceptional Children*, 54, 263-271.
- Fuchs, L. S. (1988). Effects of computer-managed instruction on teachers' implementation of systematic monitoring programs and student achievement. *Journal of Educational Research*, 81, 294-304.
- Fuchs, L. S. (1989). Evaluating solutions: Monitoring progress and revising intervention plans. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp.153-181). New York: Guilford Press.
- Fuchs, L. S. (1993). Enhancing instructional programming and student achievement with curriculum-based measurement. In J. Kramer (Ed.), *Curriculum-based measurement* (pp. 65-104). Lincoln, NE: Buros Institute of Mental Measurements.

- Fuchs, L. S., Dempsey, S., Roberts, H., & Kintsch, A. (1995). Best practices in school reintegration. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-III* (pp. 879-891). Washington, DC: National Association of School Psychologists.
- Fuchs, L. S., & Deno, S. L. (1981). *A comparison of teacher placements based on teacher judgment, standardized testing and curricular-based assessment*. (Research Rep. 56). Minneapolis: University of Minnesota, Institute for Research on Learning Disabilities.
- Fuchs, L. S., & Deno, S. L. (1991). Pragmatic distinctions between instructionally relevant measurement models. *Exceptional Children*, 57, 488-500.
- Fuchs, L. S., & Deno, S. L. (1992a). Effects of curriculum within curriculum-based measurement. *Exceptional Children*, 58(3), 232-242.
- Fuchs, L. S., & Deno, S. L. (1992b). Identifying a measure for monitoring student reading progress. *School Psychology Review*, 21(1), 45-58.
- Fuchs, L. S., & Deno, S. L. (1994). Must instructionally useful performance assessment be based in the curriculum? *Exceptional Children*, 61, 15-24.
- Fuchs, L. S., Deno, S. L., & Marston, D. (1983). Improving the reliability of curriculum-based measures of academic skills for psychological decision making. *Diagnostique*, 8, 135-149.
- Fuchs, L., Deno, S., & Mirkin, P. (1982). *Direct and frequent measurement and evaluation: Effects on instruction and assessment of student progress* (Research Report No. 97). Minneapolis, MN: University of Minnesota Institute for Research on Learning Disabilities (ERIC Document Reproduction Service No. 226 043)
- Fuchs, L., Deno, S. L., & Mirkin, P. (1984). The effects of frequent curriculum based measurement and evaluation on pedagogy, student achievement and student awareness of learning. *American Educational Research Journal*, 21, 449-460.
- Fuchs, D., & Fuchs, L. S. (1990). Framing the REI debate: Abolitionists versus conservationists, In J. W. Lloyd, N. N. Singh, & A. C. Repp (Eds.), *The regular education initiative: Alternative perspectives on concepts, issues and models* (pp. 241-255). Sycamore, IL: Sycamore.
- Fuchs, D., & Fuchs, L. S. (1994). Inclusive schools movement and the radicalization of special education reform. *Exceptional Children*, 60, 294-309.
- Fuchs, D., Fuchs, L. S., & Fernstrom, P. (1992). Case-by-case reintegration of students with learning disabilities [Special issue: Integrating learners with disabilities in regular education programs]. *Elementary School Journal*, 92, 261-281.

- Fuchs, D., Fuchs, L. S., & Fernstrom, P. (1993). A conservative approach to special education reform: Mainstreaming through transenvironmental programming and curriculum-based measurement. *American Educational Research Journal*, *30*, 149-177.
- Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, *53*, 199-208.
- Fuchs, L. S., & Fuchs, D. (1986b). Curriculum-based assessment of progress toward long- and short-term goals. *Journal of special Education*, *20*, 69-82.
- Fuchs, L. S., & Fuchs, D. (1987). The relation between methods of graphing student performance data and achievement: A meta-analysis. *Journal of Special Education Technology*, *8*(3), 5-13.
- Fuchs, L. S., & Fuchs, D. (1989). Enhancing curriculum-based measurement through computer applications: Review of research and practice. *School Psychology Review*, *18*, 317-327.
- Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review*, *21* (1), 45-58.
- Fuchs, L. S., & Fuchs, D. (1996). Combining performance assessment and curriculum-based measurement to strengthen instructional planning. *Learning Disabilities Research and Practice*, *11*, 183-192.
- Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989a). Effects of alternative goal structures within curriculum-based measurement. *Exceptional Children*, *55*, 429-438.
- Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989b). Effects of instrumental use of curriculum-based measurement to enhance instructional programs. *Remedial and Special Education*, *10*, 43-52.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Allinder, R. M. (1991). Effects of expert system advice within curriculum-based measurement on teacher planning and student achievement in spelling. *School Psychology Review*, *20*(1), 49-66.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Ferguson, C. L. (1992). Effects on expert system consultation within curriculum-based measurement using a reading task maze. *Exceptional Children*, *58*, 436-450.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Hasselbring, T. S. (1987). Using computers with curriculum-based monitoring: Effects on teacher efficiency and satisfaction. *Journal of Special Education Technology*, *8*, 14-27.

- Fuchs, S. L., Fuchs, D., Hamlett, C. L., & Phillips, N. B. (1994). Classwide curriculum-based measurement: Helping general educators meet the challenge of student diversity. *Exceptional Children, 60*, 518-537.
- Fuchs, S. L., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1990). The role of skills analysis in curriculum-based measurement in math. *School Psychology Review, 19*, 6-22.
- Fuchs, S. L., Fuchs, D., Hamlett, C. L., Waltz, L., & Germann, G. (1993). Formulative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22*(1), 27-48.
- Fuchs, L. S., Fuchs, D., & Maxwell, S. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9*(2), 20-28.
- Fuchs, L. S., Hamlett, C. L., Fuchs, D., Stecker, P. M., & Ferguson, C. (1988). Conducting curriculum-based measurement with computerized data collection: Effects on efficiency and teacher satisfaction. *Journal of Special Education Technology, 9*(2), 73-86.
- Fuchs, L. S., Tindal, G., & Deno, S. L. (1984). Methodological issues in curriculum based reading assessment. *Diagnostique, 9*, 191-207.
- Fuchs, L., Wesson, C., Tindal, G., Mirkin, P., & Deno, S. (1982). *Instructional changes, student performance, and teacher preferences: The effect of specific measurement and evaluation procedures* (Research Report No. 64). Minneapolis, MN: University of Minnesota Institute for Research on Learning Disabilities.
- Galaghan, J. E. (1985). Psychological testing: Turn out the lights, the party's over. *Exceptional Children, 52*, 288-298.
- Germann, G., & Tindal, G. (1985a). Applications of direct and repeated measurement using curriculum-based assessment. *Exceptional Children, 51*(2), 110-121.
- Germann, G., & Tindal, G. (1985b). An application of curriculum-based assessment: The use of direct and repeated measurement. *Exceptional Children, 52*(3), 244-265.
- Gesell, A., & Amatrude, C. (1954). *Developmental diagnosis*. New York: Paul B. Holber.
- Gickling, E. E., & Armstrong, D. L. (1978). Levels of instructional difficulty as related to on-task behavior, task completion, and comprehension. *Journal of Learning Disabilities, 11*, 559-566.
- Gickling, E. E., & Havertape, J. (1981). *Curriculum-based assessment (CBA)*. Minneapolis, MN: National School Psychology Inservice Training Network.

- Gickling, E., & Rosenfeld, S. (1995). Best practices in curriculum-based assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (3<sup>rd</sup> ed., pp.587-595). Washington, DC: National Association of School Psychologists.
- Gickling, E. E., & Thompson, V. P. (1985). A personal view of curriculum-based assessment. *Exceptional Children*, 52, 205-218.
- Good, R. H., III, & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp.61-88). New York: Guilford Press.
- Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic indicators of basic early literacy skills* (6<sup>th</sup> ed.). Eugene, OR: Institute for the Development of Education Achievement. Available: <http://dibels.uoregon.edu>
- Good, R. H., & Kaminsky, R. A. (1996). Assessment for instructional decisions: Toward a proactive/prevention model of decision making for early literacy skills. *School Psychology Quarterly*, 11, 1-11.
- Gordon, D. T. (2002). Curriculum access in the digital age. *Harvard Education Letter*, 18 (1), 1-5.
- Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research*, 26, 499-510.
- Greenwood, C. R., & Carta, J. J. (1987). An ecobehavioral interaction analysis of instruction within special education. *Focus on Exceptional Children*, 19(9), 1-12.
- Greenwood, C. R., Terry, B., Marquis, J., & Walker, D. (1994). Confirming a performance-based instructional model. *School Psychology Review*, 23(4), 652-668.
- Griffin, S. (1998). Re-thinking the primary school math curriculum: An approach based on cognitive science. *Issues in Education*, 4(1), 1-51.
- Harp, B. (1994). *Assessment & evaluation*. Norwood, MA: Christopher-Gordon Publishers, Inc.
- Hecht, S., Burgess, S., Torgeson, J., Wagner, R., & Rashotte, C. (1995). Explaining social class differences in growth of reading skills from beginning kindergarten through fourth-grade: The role of phonological awareness, rate of access, and print knowledge. *Reading and Writing*, 12, 99-128.
- Hennings, D. G. (1992). *Beyond the read-aloud: Learning to read through listening to and reflecting on literature*. Bloomington: Phi Delta Kappa Educational Foundation.

- Higgins, K., & Boone, R. (1997). *Technology for students with learning disabilities: Educational applications*. Austin, TX: PRO-ED.
- Hintze, J. M., Daly, E. J., & Shapiro, E. S. (1998). An investigation of the effects of passage difficulty level on outcomes of oral reading fluency progress monitoring. *School Psychology Review, 27*, 433-445.
- Hintze, J. M., & Shapiro, E. S. (1997). Curriculum-based measurement and literature-based reading: Is curriculum-based measurement meeting the needs of changing reading curricula? *Journal of School Psychology, 35*(4), 351-375.
- Hintze, J. M., Shapiro, E. S., & Lutz, J. G. (1994). The effects of curriculum on the sensitivity of curriculum-based measurement in reading. *The Journal of Special Education, 28* (2), 188-202.
- Hintze, J. M., Shapiro, E. S., Conte, K. L. & Basile, I. M. (1997). Oral reading fluency and authentic reading material: Criterion validity of the technical features of CBM survey-level assessment. *School Psychology Review, 26*(4), 535-553.
- Howell, K. W. (1986). Direct assessment of academic performance. *School Psychology Review, 15*, 324-335.
- Howell, K. W., Fox, S. L., & Morehead, M. K. (1993). *Curriculum-based evaluation: Teaching and decision making*. Pacific Gove, CA: Brooks/Cole.
- Howell, K. W., & Morehead, M. K. (1987). *Curriculum-based evaluation for special and remedial education*. Columbus, OH: Charles Merrill.
- Humphreys, L. G. (1988). Trends in levels of academic achievement of Blacks and other minorities. *Intelligence, 12*, 231-260.
- Individuals with Disabilities Education Improvement Act of 2004*, H. R. 1350, 108 Cong., 2<sup>nd</sup> Sess. (2004). Available online at [http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=108\\_cong\\_public\\_laws&docid=f:publ446.108](http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=108_cong_public_laws&docid=f:publ446.108) on June 14, 2006.
- Iverson, S., & Tunmer, W. E. (1993). Phonological processing skills and the reading recovery program. *Journal of Educational Psychology, 85*, 112-126.
- Jacobs, H. H. (1989). *Interdisciplinary curriculum: Design and implementation*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Jitendra, A. K., & Kameenui, E. J. (1993). Dynamic assessment as a compensatory assessment approach: A description and analysis. *Remedial and Special Education, 14*, 6-18.

- Johnston, P., & Allington, R. L. (1991). Remediation. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research, vol. II* (pp. 984-1012). White Plains, NY: Longman.
- Jorm, A. F., Share, D. L., Maclean, R., & Atthews, R. (1986). Cognitive factors at school entry predictive of specific reading retardation and general reading backwardness: A research note. *Journal of Child Psychology and Psychiatry, 27*, 45-54.
- Joshi, R. M. (1995). Assessing reading and spelling skills. *School Psychology Review, 24*, 361-375.
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*, 437-447.
- Kaminsky, R., & Good, R. (1996). Towards a technology for assessing basic early literacy skills. *School Psychology Review, 25*(2), 215-227.
- Kaminsky, R., & Good, R. (1998). Assessing early literacy skills in a problem-solving model: Dynamic indicators of basic early literacy skills. In M. R. Shinn (Ed.), *Advances in curriculum-based measurement and its use in a problem-solving model*. New York: Guilford Press.
- Kelly, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods, 8*, 305-321.
- King-Sears, M. E., Burgess, M., & Lawson, T. L. (1999). *Teaching Exceptional Children, 32*, 30-38.
- King-Sears, M. E. (1994). *Curriculum based assessment in special education*. San Diego: Singular Publishing Group, Inc.
- Koehler, J. (1996). Planning from second Principles. *Artificial Intelligence, 87*, 145-186.
- Kohr, R. (2004). *Assessment history in Pennsylvania*. Harrisburg, PA: Private historical document.
- Kovaleski, J. (2006). Potential pitfalls of response to intervention. To appear in: *The handbook of response to intervention: The science and practice of assessment and intervention*. S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.). New York: Springer, Inc.
- Kranzler, J. H., Brownell, M. T., & Miller, M. D. (1998). The construct validity of curriculum-based measurement of reading: An empirical test of a plausible rival hypothesis. *Journal of School Psychology, 36*(4), 399-415.

- Learning First Alliance. (1998). *Every child reading: An action plan of the learning first alliance*. Washington, DC: Learning First Alliance.
- Lesiak, J. (1988). *Approaches to reading instruction: What are they? Are they effective?* Paper presented at the 1998 NASP Annual Convention, Orlando, FL.
- Lombard, T. (1988). *Curriculum-based measurement: Megatesting or mctesting*. St. Paul: Minnesota State Department of Education.
- Lundberg, I. (1988). Lack of phonological awareness: A critical factor in developmental dyslexia. In C. von Euler, I. Lundberg, & G. Lennerstrand (Eds.), *Brain and reading* (pp. 221-231). London: MacMillan.
- Lundberg, I., Frost, J., & Petersen, O. P. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23*(3), 263-284.
- Lyle, J., & Penn, W. (1986). *Special education: Views from America's cities*. Philadelphia, PA: Research for Better Schools.
- Lyon, G. R. (1994). *Frames of reference for the assessment of learning disabilities*. Baltimore, MD: Paul Brookes.
- Lyon, G. R. (1995). Research initiatives on learning disabilities: Contributions from scientists supported by the National Institute of Child Health and Human Development. *Journal of Child Neurology, 10*, 120-126.
- Madden, N. A., Slavin, R. E., Karweit, N., Dolan, L., & Wasik, B. A. (1993). Success for all: Longitudinal effects of a restructuring program for inner-city elementary schools. *American Educational Research Journal, 30*, 123-148.
- Marston, D. (1988). The effectiveness of special education: A time-series analysis of reading performance in regular and special education settings. *The Journal of Special Education, 21*(4), 13-26.
- Marston, D. (1989). Curriculum-based measurement: What is it and why do it? In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18-78). New York: Guilford Press.
- Marston, D., Deno, S. L., & Tindal, G. (1984). Eligibility for learning disability services: A direct and repeated measurement approach. *Exceptional Children, 50*, 554-555.
- Marston, D., Fuchs, L. S., & Deno, S. L. (1986). Measuring pupil progress: A comparison of standardized achievement tests and curriculum-related measures. *Diagnostique, 11*, 77-90.

- Marston, D., & Magnusson, D. (1985). Implementing curriculum-based measurement in special and regular education settings. *Exceptional Children*, 52(3), 266-276.
- Marston, D., & Magnusson, D. (1988). Curriculum-based measurement: District level implementation. In J. Graden, J. Zins, & M. Curtis (Eds.). *Alternative education delivery systems*, (pp.137-172). Washington, DC: National Association of School Psychologists.
- Marston, D., Shinn, M. R., Tindal, G., & Deno, S. (1983). *A comparison of standardized achievement tests and direct measurement techniques in measuring student progress*. (Research Report No. 126). Minneapolis, MN: University of Minnesota Institute for Research on Learning Disabilities.
- Marzano, R. J., Brandt, R., Hughes, C. S., Jones, B. F., Presseisen, B. Z., Rankin, S. C., & Suhor, C. (1989). *Dimensions of thinking: A framework for curriculum and instruction*. Alexandria, VA: Association for Supervision and Curricular Development.
- Masland, R., & Masland, M. (1988). *Preschool prevention of reading failure*. Parkton, MD: York Press.
- Masters, J. (2004) *Telephone interview with this researcher*. Harrisburg, PA: Pennsylvania Department of Education
- McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review*, 33(2), 193-203.
- Meisels, S. J. (1989). High stakes testing in kindergarten. *Educational Leadership*, 46, 16-22.
- Meisels, S. J., & Provence, S. (1989). *Screening and assessment: Guidelines for identifying young disabled and developmentally vulnerable children and their families*. Washington, DC: National Center for Clinical Infant Programs.
- Mercer, C. D., Campbell, K. U., Miller, M. D., Mercer, K. D., & Lane, H. B. (2000). Effects of reading fluency interventions for middle schoolers with specific learning disabilities. *Learning Disabilities Research and Practice*, 15(4), 179-189.
- Mehrens, W.A., & Clarizo, H. F. (1993). Curriculum-based measurement: Conceptual and psychometric considerations. *Psychology in the Schools*, 30, 241-254.
- Minke, K. M., & Bear, G. G. (Eds.). (2000). *Preventing school problems: Promoting school success*. Bethesda, MD: NASP Publications.

- Moats, L. C. (1999). *Teaching reading is rocket science. What expert teachers of reading should know and be able to do*. Washington, DC: American Federation of Teachers.
- Murray, B. (n.d.). *A comparison of the Houghton-Mifflin (2003) and open court (2002) basal reading series*. Auburn, AL: Auburn University. [http://www.auburn.edu/academic/education/reading\\_genie/HMvOC.html](http://www.auburn.edu/academic/education/reading_genie/HMvOC.html)
- Murray, R. (2004). *Telephone interview with this research*. Harrisburg, PA: Pennsylvania Department of Education.
- National Association of School Psychologists. (1994). *Assessment and eligibility in special education: An examination of policy and practice with proposals for change*. Alexandria, VA: National Association of State Directors of Special Education.
- National Reading Panel (2001). *The report of the national reading panel: Teaching children to read*. Jessup, MD: National Institute for Literacy at EDPubs.
- National Reading Panel. (2000). *Teaching children to read: An evidence-based Assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: U.S. Department of health and Human Services.
- Neill, M., & Medina, N. (1989, May). Standardized testing: Harmful to educational health. *Phi Delta Kappan*, 70 (9), 688-697.
- Nevin, A. I., & Paolucci-Whitcomb, P. (1998). *Models of curriculum-based assessment*. Austin: Pro-Ed.
- O'Connor, R. E., Jenkins, J. R., Leicester, N., & Slocum, T. A. (1993). Teaching phonological awareness to young children with learning disabilities. *Exceptional Children*, 59(6), 532-546.
- Office of Educational Research and Improvement. (1991). *Youth indicators*. Washington, DC: U. S. Government Printing Office.
- Pany, D., & McCoy, K. (1988). Effects of corrective feedback on word accuracy and reading comprehension of readers with learning disabilities. *Journal of Learning Disabilities*, 21(9), 546-550.
- Pearson, P. D., & Raphael, T. E. (2003). Toward a more complex view of balance in the literacy curriculum. In L. M. Morrow, L. B. Gambrell, & M. Pressley (Eds.), *Best practices in literacy instruction*. New York: Guilford.

- Perrone, V. (1991). *Expanding student assessment*. USA: Association for Supervision and Curriculum Development.
- Powell-Smith, K., & Stewart, L. H. (1998). The use of curriculum-based measurement in the reintegration of students with mild disabilities. In M. R. Shinn (Ed.). *Advanced applications of curriculum-based measurement*. New York: Guilford Press.
- Reeves, D. B. (2000). *Accountability in action*. Denver, CO: Center for Performance Assessment.
- Reschly, D. J., & Grimes, J. P. (1991). State department and university cooperation: Evaluation of continuing education and consultation and curriculum-based assessment. *School Psychology Review*, 20, 522-529.
- Reschly, D. J., & Tilly, W. D. (1993). The why of system reform. *National Association of School Psychologists Communique*, 22, 1-6.
- Reschly, D. J., & Wilson, M. S. (1990). Cognitive processing versus traditional intelligence: Diagnostic utility, intervention implications, and treatment validity. *School Psychology Review*, 19(4), 443-459.
- Reutzel, D. R., & Cooter, R. B. (2003). *Teaching children to read: Putting the pieces together* (4<sup>th</sup> ed.). Upper Saddle River, NJ: Pearson
- Rogers, W. T. (1983). Use of separate answer sheets with hearing impaired and deaf school age students. *Journal of Special Education*, 7(1), 63-72.
- Rosenfield, S. A. (1987). *Instructional consultation*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rosenfield, S. A., & Gravois, T. A. (1996). *Instructional consultation teams: Collaborating for change*. New York: Guilford Press.
- Rosenfeld, S., & Kuralt, S. K. (1994). Best practices in curriculum-based assessment. In A. Thomas and J. Grimes (Eds.), *Best practices in school psychology II* (pp. 275-286). Washington, DC: The National Association of School Psychologists.
- Ross, S. M., Alberg, M., & McNelis, M. (1997). *Evaluation of elementary school school-wide programs: Clover Park School District, year 1: 1996-97*. Memphis, TN: Center for Research in Educational Policy, University of Memphis.
- Ross, S. M., McNelis, M., Lewis, T., & Loomis, S. (1998). *Evaluation of success for all programs: Little Rock School District, year 1: 1997-1998*. Memphis, TN: Center for Research in Educational Policy, University of Memphis.

- Ross, S. M., Smith, L. J., & Casey, J. (1999). Bridging the gap: The effects of the success for all program on elementary school reading achievement as a function of student ethnicity and ability level. *School Effectiveness and School Improvement, 10*(2), 129-150.
- Rutter, M., & Yule, W. (1975). The concept of specific reading retardation. *Journal of Child Psychology and Psychiatry, 16*, 181-197.
- Salvia, J., & Hughs, C. C. (1990). *Curriculum based assessment: Testing what is taught*. New York: MacMillan Publishing Co.
- Salvia, J., & Hughs, C. C. (1989). *Curriculum based assessment: Testing what is taught*. Paramus: Prentice-Hall.
- Salvia, J., & Ysseldyke, J. E. (1985). *Assessment in special and remedial education* (3rd ed.). Boston: Houghton-Mifflin Co.
- Salvia, J. & Ysseldyke, J. E. (1995) *Assessment* (6<sup>th</sup> ed.). Boston: Houghton-Mifflin.
- Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities. In B. K. Shapiro, P. J. Accardo, & A. J. Capute (Eds.), *Specific reading disability: A view of the spectrum*. Timonium, MD: York Press.
- Schmoker, M. (2001). *The results fieldbook: Practical strategies from dramatically improved schools*. Alexandria, VA: Association for Supervision and Curricular Development.
- Schulte, A. C., Villwock, D. N., Whichard, S. M., & Stallings, C. F. (2001). High stakes testing and expected progress standards for students with learning disabilities: A five year study of one district. *School Psychology Review, 30*(4), 487-506.
- Schumaker, J. B., & Deschler, D. (1988). Implementing the regular education initiative in secondary schools: A different ball game. *Journal of Learning Disabilities, 21*(1), 36-42.
- Shapiro, E. S. (1988). Preventing academic failure. *School Psychology Review, 17* (4), 601-613.
- Shapiro, E. S. (1989). *Academic skills problems: Direct assessment and intervention*. New York: The Guilford Press.
- Shapiro, E. S. (1990). An integrated model for curriculum-based assessment. *School Psychology Review, 19* (3), 331-349.

- Shapiro, E. S. (1992). Use of Gickling's model of curriculum-based assessment to improve reading in elementary age students. *School Psychology Review, 21*, 168-176.
- Shapiro, E. S. (1996). *Academic skills problems: Direct assessment and intervention* (2<sup>nd</sup> ed.). New York: Guilford Press.
- Shapiro, E. S., & Derr, T. F. (1987). An examination of the overlap between reading curricula and standardized achievement tests. *The Journal of Special Education, 21*, 59-67.
- Shapiro, E. S., & Derr, T. F. (1990). Curriculum-based assessment. In T. B. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology*, (2<sup>nd</sup> ed.). New York: Wiley.
- Shapiro, E. S., & Lentz, F. E. (1985). Assessing academic behavior: A behavioral approach. *School Psychology Review, 14*, 325-338.
- Shapiro, E. S., & Lentz, F. E. (1986). Behavioral assessment of academic skills. In T. R. Kratochwill (Ed.), *Advances in School Psychology, 5*, 87-139.
- Shaywitz, B. A., Fletcher, J. M., & Shaywitz, S. E. (1995a). Defining and classifying learning disabilities and attention-deficit/hyperactivity disorder. *Journal of Child Neurology, 10*(Supplement 1), S50-S57.
- Shaywitz, S. (2003). *Overcoming dyslexia: A new and complete science-based program for reading problems at any level*. New York: Knopf.
- Shaywitz, S., Escobar, M., Shaywitz, B., Fletcher, J., & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine, 320*, 145-150.
- Shaywitz, S. E., Fletcher, J. M., & Shaywitz, B. A. (1994). Issues in the definition and classification of attention deficit disorder. *Topics in Language Disorders, 14*, 1-25.
- Shaywitz, S., Shaywitz, B., Fletcher, J., & Escobar, M. (1990). Relevance of reading disability in boys and girls: Results of the Connecticut longitudinal study. *Journal of the American Medical Association, 204*, 998-1002.
- Shephard, L. (1983). The role of measurement in educational policy: Lessons from the identification of learning disabilities. *Educational Measurement: Issues and Practice, 1*, 4-8.

- Shin, J., Deno, S. L., & Espin, C. (2000). Technical adequacy of the maze task for curriculum-based measurement of reading growth. *The Journal of Special Education, 34*(3), 164-172.
- Shinn, M. R. (1986). Does anyone care what happens after the refer-test-place sequence: The systematic evaluation of special education program effectiveness. *School Psychology Review, 15*, 49-58.
- Shinn, M. R. (1988). Development of curriculum-based local norms for use in special education decision-making. *School Psychology Review, 17* (1), 61-80.
- Shinn, M. R. (Ed.). (1989a). *Curriculum-based measurement: Assessing special children*. New York: Guilford Press.
- Shinn, M. R. (1989b). Identifying and defining academic problems: CBM screening and eligibility procedures. In M.R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 90-129). New York: Guilford.
- Shinn, M. R. (1995). Curriculum-based measurement and its use in a problem-solving model. In Thomas & J. Grimes (Eds.), *Best practices in school psychology III* (3rd ed., pp. 547-568). Washington, DC: National Association of School Psychologists.
- Shinn, M. R. (Ed.). (1998). *Advanced applications of CBM*. New York: Guilford Press.
- Shinn, M. R., & Bamonto, S. (1998). Advanced applications of curriculum-based measurement: "Big ideas" and avoiding confusion. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 1-31). New York: Guilford Press.
- Shinn, M. R., Collins, V. L., & Gallagher, S. (1998). Curriculum-based measurement and its use in a problem-solving model with students from minority backgrounds. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement*, (pp.143-174). New York: Guilford Press.
- Shinn, M. R., Gleason, M. M., & Tindal, G. (1989). Varying the difficulty of testing materials: Implications for curriculum-based measurement. *The Journal of Special Education, 23*, 223-233.
- Shinn, M. R., & Good, R. (1993). CBA: An assessment of its current status and a prognosis for its future. In J. Kramer (Ed.), *Curriculum-based assessment: Examining old problems, exploring new solutions*. Lincoln, NE: Buros Institute/University of Nebraska.

- Shinn, M. R., & Habedank, L. (1992). Curriculum-based measurement in special education: Problem identification and certification decisions. *Preventing School Failure, 36*(2), 11-15.
- Shinn, M. R., Habedank, L., Rodden-Nord, K., & Knutson, N. (1993). Using curriculum-based measurement to identify potential candidates for reintegration into general education. *The Journal of Special Education, 27*(2), 202-221.
- Shinn, M. R., & Hubbard, D. D. (1992). Curriculum-based measurement and problem-solving assessment: Basic procedures and outcomes. *Focus on Exceptional Children, 24* (5), 1-20.
- Shinn, M. R., & Marston, D. (1985). Differentiating mildly handicapped, low-achieving and regular education students: A curriculum-based approach. *Remedial and Special Education, 6*, 31-45.
- Shinn, M. R., Nolet, V., & Knutson, N. (1990). Best practices in curriculum-based measurement. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-II* (pp. 287-307). Washington, DC: National Association of School Psychologists.
- Shinn, M. R., Powell-Smith, K. A., & Good, R. H. (1996). Evaluating the effects of responsible reintegration into general education for students with mild disabilities on a case-by-case basis. *School Psychology Review, 25*(4), 519-539.
- Shinn, M. R., & Shinn, M. M. (2002). *Administration and scoring of reading curriculum-based measurement (R-CBM) for use in general outcome measurement*. Eden Prairie, MN: Edformation.
- Shinn, M. R., Tindal, G., Spira, D., & Marston, D. (1987). Practice of learning disabilities as social policy. *Learning Disabilities Quarterly, 10*, 17-28.
- Shinn, M. R., Ysseldyke, J. E., Deno, S. L., & Tindal, G. (1986). A comparison of differences between students labeled learning disabled and low achieving on measures of classroom performance. *Journal of Learning Disabilities, 19*, 545-552.
- Simmons, D. C., & Kameenui, E. J. (1998). *What reading research tells us about children with diverse needs*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Simmons, D. C., & Kameenui, E. J. (2003). *A consumer's guide to evaluating a core reading program grades k-3: A critical elements analysis*. Eugene, OR: Institute for the Development of Educational Achievement.

- Slavin, R. E. (2003, February). A reader's guide to scientifically based research: Learning how to assess the validity of education research is vital for creating effective, sustained reform. *Educational Leadership*, 12-16.
- Slavin, R. E. (1997). *Reading by nine: A comprehensive strategy*. Washington, DC: US Department of Education (Grant No. OERI-R-117-D40005)
- Slavin, R. F., & Fashola, O. S. (1998). *Show me the evidence! Proven and promising programs for America's schools*. Thousand Oaks, CA: Corwin.
- Slavin, R. E., Madden, N. A., Dolan, L., & Wasik, B. A. (1996). Success for all: A Summary of research. *Journal of Education for Students Placed at Risk*, 1(1), 41-76.
- Smith, L. J., Ross, S. M., & Casey, J. (1996). Multi-site comparison of the effects of success for all on reading achievement. *Journal of Literacy Research*, 28(3), 329-353.
- Smith, S. B., Simmons, D. C., Gleason, M. M., Kameenui, E. J., Bake, S. K., Sprick, M., Gunn, B., Thomas, C. L., Chard, D. J., Plasencia-Peinado, J., Peinado, R. (2001). An analysis of phonological awareness instruction in four kindergarten basal reading programs. *Reading and Writing Quarterly*, 17(1), 25-51.
- Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Stage, S. A., & Jacobsen, M. D. (2001). Predicting student success on state-mandated performance-based assessment using oral reading fluency. *School Psychology Review*, 30(3), 407-419.
- Stage, S. A., Sheppard, J., Davidson, M. M., & Browning, M. M. (2001). Prediction of first-graders growth in oral reading fluency using kindergarten letter fluency. *Journal of School Psychology*, 39(3), 225-237.
- Stein, M., Johnson, B., & Gutlohn, L. (1999). Analyzing beginning reading programs. *Remedial and Special Education*, 20(5), 275-287.
- Stevens, J. (1986). *Applied multivariate statistics for the social sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Swadener, B. B., & Kessler, S. (1991). Reconceptualizing early childhood education *Early Education and Development*, 2(Special Issue).
- Thomas, A., & Grimes, J. (1990). *Best practices in school psychology II*. Washington, DC: National Association of School Psychologists.

- Tindal, G. A., & Marston, D. B. (1990). *Classroom based assessment: Evaluating instructional outcomes*. Columbus, Ohio: Merrill.
- Tindall, G., Marston, D., & Deno, S. (1983). *The reliability of direct and repeated measurement*. (Research Report No. 183). Minneapolis, MN: University of Minnesota Institute for Research on Learning Disabilities.
- Tindal, G., & Parker, R. (1989). Assessment of written expression for students in compensatory and special education programs. *Journal of Special Education, 23*, 169-183.
- Torgesen, J. K. (1997). The prevention and remediation of reading disabilities. Evaluating what we know from research. *Journal of Academic Language Therapy, 1*, 11-47.
- Torgeson, J. K., & Bryant, B. R. (1994). *Test of phonological awareness*. Austin, TX: Pro-Ed.
- Torgeson, J. K., Wagner, R., & Rshotte, C. (1997). Prevention and remediation of severe reading disabilities: Keeping the end in mind. *Scientific Studies of Reading, 1*, 217-234.
- Tran, Q. N. (2007). A new class of term orders for elimination. *Journal of Symbolic Computation, 2*, 533-548.
- Tunmer, W. E., & Hoover, W. A. (1993). Phonological recoding skill and beginning reading. *Reading & Writing: An Interdisciplinary Journal, 5*, 161-179.
- U. S. Department of Education (1995). *Seventeenth annual report to Congress on the implementation of the Individuals with Disabilities Education Act*. Washington, DC: Author.
- U.S. Department of Education. (n.d.). *Comprehensive school reform program*. Retrieved November 8, 2005, from <http://www.ed.gov/programs/compreform/2pager.html>
- U.S. Department of Education. (2005, March 10). *Expanding the promise, continuing the progress*. Retrieved November 2, 2005, from <http://www.ed.gov/news/speeches/2005/03/03102005.html>
- U. S. Department of Education. (2002). *The No Child Left Behind Act*. Washington, DC: Author.
- U.S. Department of Education. (2002). *No Child Left Behind Act 2001*. <http://www.ed.gov/nclb/landing.jhtml>

- Videen, J., Deno, S. & Marston, D. (1982). *Correct word sequences: A valid indicator of writing proficiency in written expression*. (Research Report No. 84). Minneapolis, MN: Institute for Research on Learning Disabilities.
- Wallace, I. F., & Hooper, S. R. (1997). Otitis media and its impact on cognitive, academic, and behavioral outcomes: A review and interpretation of the findings. In J. E. Roberts, I. F. Wallace, & F. W. Henderson (Eds), *Otitis media in young children*. Baltimore, MD: Brookes Publishing Co.
- Walton, W. T. (1986). Educators' response to methods of collecting, storing and analyzing behavioral data. *Journal of Special Education Technology*, 7, 50-55.
- Wang, D. (1999). *Elimination methods*. Berlin: Springer-Verlag.
- Wesson, C. L. (1991). Curriculum-based measurement and two models of follow-up consultation. *Exceptional Children*, 57, 246-257.
- Wesson, C. L., King, R. P., & Deno, S. L. (1984). Direct and frequent measurement of student performance: If it's good for us why don't we do it? *Learning Disability Quarterly*, 7, 45-48.
- Wesson, C., Skiba, R., Sevcik, B., King, R. P., & Deno, S. (1984). The effects of technically adequate instructional data on achievement. *Remedial and Special Education*, 5, 17-22.
- White, O. R., & Haring, N. G. (1980). *Exceptional teaching* (2<sup>nd</sup> ed.). Columbus, OH: Charles Merrill.
- Wick, J. W. (1987). *School based evaluation*. Boston: Kluwer Academic Publishers Group.
- Will, M. (1986). Educating students with learning problems: A shared responsibility. *Exceptional Children*, 52, 411-415.
- Wolf, M., & Obregon, M. (1992). Early naming deficit, developmental dyslexia and a specific deficit hypothesis. *Brain and Language*, 42, 219-247.
- Ysseldyke, J. E. (1986). The use of assessment information to make decisions about students. In R. J. Morris & B. Blatt (Eds.), *Special education: Research and trends*. New York: Pergamon.
- Ysseldyke, J. E., Algozzine, B., & Epps, S. (1983). A logical and empirical analysis of current practices in classifying students as handicapped. *Exceptional Children*, 50, 160-166.

Ysseldyke, J. E., & Christenson, S. (1987a). *The instructional environment scale*. Austin, TX: Pro-Ed.

Ysseldyke, J. E., & Christenson, S. (1987b). Evaluating students' instructional environments. *Remedial and Special Education*, 8(3), 17-24.

## APPENDICES

APPENDIX A  
Standards and Rubrics

READING		WRITING	
Grade	Expectations	Grade	Expectations
PK	1 2 Proficient 3 4 Advanced	PK	4+ Advanced 1 2 3 Proficient
K	1 Basic 2 Proficient 3 4 Advanced	K	7+ Advanced 5 6 Proficient 4 and below Basic
1		1	9 10+ Advanced 7 8 Proficient 4 5 6 Basic 1 2 3 Below Basic
2		2	7 Proficient 6 Basic 5 Below Basic
3		3	
4		4	
5	58-62 → 91-99 4 Advanced - A	5	10 11 12 Advanced 8 9 Proficient 4 5 6 7 Basic 2 3 Below Basic
6	3 Proficient - P	6	
7	2 Basic - B	7	
8	1 Below Basic - B	8	
9		9	11 12 Advanced 8 9 10 Proficient 6 7 Basic 5 and below Below Basic
10		10	
11			

**PSSA  
Holistic Scoring Rubric**

**PSSA Rubric for Reading**

## PENNSYLVANIA READING ASSESSMENT RUBRIC

### LEVEL 4

- Demonstrates a thorough understanding of the complexity of the text through detailed elaboration and extensions of text with sophisticated ideas, insights and reflections. There are no errors in text-based facts.
- Exhibits a level of comprehension that extends beyond the literal, to the personal, critical and/or evaluative responses.
- Cites evidence, makes a variety of strong connections to other experiences, texts, concepts, issues and/or cultural settings.

crit. 4

### LEVEL 3 \*

- Demonstrates confident, coherent and adequate understanding and interpretation of the text through some elaboration and extension.
- There are no major errors in text-based facts.
- Exhibits a level of comprehension that reflects extensions that are more literal or personal.
- Makes connections to personal experiences, other texts and/or background knowledge.

### LEVEL 2

- Demonstrates a limited understanding and/or interpretation of the text. There may be errors in text-based facts.
- Exhibits a level of comprehension that consists primarily of literal responses to the text.
- Makes connections between other experiences and text that are disjointed, fragmented, limited and not integral to the text.

### LEVEL 1

- Demonstrates an attempt to respond with very limited evidence of understanding of the text.
- There may be errors in text-based facts.
- Exhibits a level of comprehension that consists of disjointed, incomplete or irrelevant responses.
- Might use relevant copied text
- Makes only distant connections to the text, using sketchy details.

### LEVEL 0 - consists of 3 types of responses:

- Non-Scorable (NS) papers are blank.
- Off-Task (OT) papers show no relationship to task and text, are illegible, irrelevant copied text, or written in a language other than English.
- Intentionally Off-Task (IO) papers are ones that have unrelenting profanity, are a refusal to perform, state a baseless charge of too personal or are drawings, scribbling, etc.

\*Note - Level 3 (or higher) is to be viewed as the performance standard for all students.

APPENDIX B

Letter of Approval

Greater Johnstown School District  
Department for Students with Disabilities  
1091 Broad St., Johnstown, PA 15906  
Phone: 814/533-5665, 533-5664, Fax: 814/533-5655

Dr. Gerald Zahorchak,  
Superintendent of Schools  
Greater Johnstown School District  
1091 Broad St.  
Johnstown, PA 15906

16 September 2002

Dear Dr. Zahorchak:

You have asked, at a recent administrative team meeting, about data regarding cohort performance on the JSSA/PSSA. You may recall that my dissertation reviews the reading performance of those students who attended GJSD during their 2<sup>nd</sup> through 8<sup>th</sup> grade years. With your permission, I'd like to complete my cohort data collection, with all data encoded by number, rather than student name, and share the data analysis with the administrative team.

May I complete the collection of this data?

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_ 9-16-2002  
Dr. Gerald Zahorchak

In service,

*Mary E. Rauch*  
Mary E. Rauch, NCSP

## APPENDIX C

### Curriculum-Based Assessment Paragraphs

In this story, Annie has a problem with her two grandmothers. You will be asked to write about the problem and how the problem was solved.

**Annie's Grandmas (revised)**

*by Helen Kronberg*

Annie had two grandmas. Annie's grandmas agreed on only one thing. They both loved Annie.

Grandma Sanders took Annie to art shows. She took her to museums and art galleries. She told Annie about the artists.

With Grandma Hale, Annie played ball. Grandma showed Annie how to hold the bat. She taught her how to catch the ball. And she tried to teach her how to pitch.

Grandma Sanders taught Annie to play the piano. By the time she was only three years old, Annie could play tunes on Grandma's piano. And she got better and better.

"Annie should learn how to plant a garden," said Grandma Hale. She taught Annie how to plant the seeds. She taught her to pat the soil over the seeds, just so. Annie even learned to weed the garden when the plants began to grow.

Grandma Sanders taught Annie good table manners. She taught her to take small bites and to use her napkin.

With Grandma Hale, Annie learned to hold a sandwich in one hand and a fishing pole in the other.

Annie and her grandmas always had a good time. Then, came Annie's birthday.

"This year, I want to take you to a big-league baseball game," said Grandma Hale.

"Oh, no," cried Grandma Sanders. "I have tickets for a lovely concert. The ball game will have to wait."

"The ball game is in the afternoon," said Grandma Hale.

Grandma Sanders nodded. "Perhaps there would be time for both."

"No!" said Annie.

"No?" said Grandma Sanders.

"No?" said Grandma Hale.

Annie gave her grandmothers each a hug. "I know you love me," she said. "I know you want to make me happy. Here's what I think we should do."

The task below requires that you provide a written response.  
Write your answer on the paper provided.

Remember, you may go back to the story, and you may use a dictionary to help you with your writing.

You are Annie. Write or print about the problem and what you will do on your birthday to solve the problem between your grandmothers.

As you write, be sure to:

- \*Pretend you are Annie and state the problem and how you solved it.
- \*Use details from the story and your own ideas, too.
- \*Write or print neatly and clearly.
- \*Use only the space provided.

In the story, Turtle and Spider trick each other. After reading the story, you will write about which one you would choose for a partner.

### **Hungry Spider and the Turtle**

*by Harold Courlander*

Spider was a hungry one. He always wanted to eat. Everyone in Ashanti knew about his appetite. He was greedy, too, and always wanted more than his share of things. So, people steered clear of Spider.

But, one day, a stranger came to Spider's habitation out in the back country. His name was Turtle. Turtle was a long way from his home. He had been walking all day in the hot sun, and he was tired and hungry. So, Spider had to invite Turtle into his house and offer him something to eat. He hated to do it, but, if he didn't extend hospitality to a tired traveler, it would get back around the countryside, and people would soon be talking about Spider behind his back.

So, he said to Turtle, "There is water at the spring for you to wash your feet in. Follow the trail, and you'll get there. I'll get the dinner ready."

Turtle turned and waddled down to the spring with a gourd bowl, as fast as he could. He dipped some water from the spring and carefully washed his feet in it. Then, he waddled back up the trail to the house. but the trail was dusty. By the time Turtle got back to the house, his feet were covered with dirt again.

Spider had the food all set out. It was steaming, and the smell of it made Turtle's mouth water. He hadn't eaten since sunrise.

Spider looked disapprovingly at Turtle's feet. "Your feet are awfully dirty," he said. "Don't you think you ought to wash them before you start to eat?"

Turtle looked at his feet. He was ashamed. They were so dirty. So, he turned around and waddled, as fast as he could, down to the spring again. He dipped some water out of the spring with the gourd bowl and carefully washed himself. Then, he scurried as fast as he could back to the house. But it takes a turtle awhile to get anywhere. When he came into the house, Spider was already eating.

"Excellent meal, isn't it?" Spider said. He looked at Turtle's feet with disapproval. "Hm, aren't you going to wash yourself?"

Turtle looked down at his feet. In his hurry to get back, he had stirred up a lot of dust, and his feet were covered with it again.

"I washed them," he said. "I washed them twice. It's your dusty trail that does it."

"Oh," Spider said, "so you are abusing my house now!" He took a big mouthful of food and chewed it up, looking very hurt.

"No," Turtle said, sniffing the food. "I was just explaining."

"Well, run along and wash up so we can get on with the eating," Spider said.

Turtle looked. The food was already half gone, and Spider was eating as fast as he could. Turtle spun around and hurried down to the spring. He dipped up some water in the gourd bowl and splashed it over his feet. Then, he scrambled back to the house.

This time, he didn't go on the trail, though, but on the grass and through the bushes. It took him a little longer, but he didn't get dust all over his feet.

When he got to the house, he found Spider licking his lips.

"Ah, what a fine meal we had!" Spider said.

Turtle looked in the dish. Everything was gone. Even the smell was gone. Turtle was very hungry. But he said nothing. He smiled.

"Yes, it was very good," he said. "You are certainly good to travelers in your village. If you are ever in my country, you may be sure of a welcome."

"It's nothing," Spider said. "Nothing at all."

Turtle went away. He didn't tell other people about the affair at Spider's house. He was quiet about his experience there.

But, one day many months later, Spider was a long distance from home, and he found himself in Turtle's country. He found Turtle at the shore of the lake getting a sunbath.

"Ah, friend Spider, you are far from your village," Turtle said. "Will you have something to eat with me?"

"Yes, that is the way it is when a person is far from home -- generosity merits generosity," Spider said hungrily.

"Wait here on the shore, and I'll go below and prepare the food," Turtle said. He slipped into the water and went down to the bottom of the lake. When he got there, he set out the food to eat. Then, he came to the top of the water and said to Spider, who was sitting impatiently on the shore, "All right, everything is ready. Let's go down and eat." He put his head under the water and swam down.

Spider was famished. He jumped into the water to follow Turtle. But Spider was very light. He floated. He splashed and splashed, kicked and kicked, but he stayed right there on top of the water. For a long time, he tried to get down where Turtle was eating, but nothing happened.

After a while, Turtle came up, licking his lips.

"What's the matter? Aren't you hungry?" he said. "The food is very good. Better hurry." And he went down again.

Spider made one more desperate try, but he just floated. Then, he had an idea. He went back to the shore, picked up pebbles and put them in his pockets of his jacket. He put so many pebbles in his pockets that he became very heavy. He was so heavy he could hardly walk. Then, he jumped into the water again, and this time he sank to the bottom where Turtle was eating. The food was half gone.

Spider was very hungry. He was just reaching for the food when Turtle said politely, "Excuse me, my friend. In my country, we never eat with our jackets on. Take off your jacket so that we can get down to business."

Turtle took a great mouthful of food and started chewing. In a few minutes, there wouldn't be anything left.

Spider was aching all over with hunger.

Turtle took another mouthful. So, Spider wriggled out of his coat and grabbed at the food. But, without the pebbles, he was so light again that he popped right up to the top of the water.

People always say that one good meal deserves another.

The task below requires that you provide a written response.  
Write your answer on the paper provided.

Remember, you may go back to the story, and you may use a dictionary to help you with your writing.

You will choose Turtle or Spider to work with you on an important school project. Tell which one you would choose and why.

As you write, be sure to:

- \*Tell why you would choose Turtle or Spider and explain.
- \*Use details from the story and your own ideas, too.
- \*Write or print neatly and clearly.
- \*Use only the space provided.

GREATER JOHNSTOWN SCHOOL DISTRICT  
 READING ASSESSMENT CLASS PROFILE

Assessment 1 2 3

GRADE LEVEL 3 YEAR 99  
 TEACHER TOR

SCHOOL ES WS MS HS

- above targeted benchmark
- at targeted benchmark
- below targeted benchmark

NAME	Reading Score	Performance Indicators						Continuum						SUGGESTED STRATEGIES
		A	P	B	BB	NS	NS	A	I	T	E	E		
1.														
2.	3	X												
3.	2			X										
4.	2			X										
5.	3			X										
6.	2			X										
7.	2			X										
8.	3			X										
9.	3			X										
10.	2			X										
11.	2			X										
12.	3			X										
13.	3			X										
14.	Ab													
15.	2													
16.	1													
17.	2													
18.	2													
19.	2													
20.	3													
21.	3													
22.	3													
23.														
24.														
25.														
26.														
27.														
28.														
29.														
30.		0	8	12	1	0								

SUMMARY