

2008

# Predicting Performance on the Pennsylvania System of School Assessment Using the Developmental Reading Assessment and the 4Sight Predictive Benchmark Assessment

Todd B. Stoltz

Follow this and additional works at: <https://dsc.duq.edu/etd>

---

## Recommended Citation

Stoltz, T. (2008). Predicting Performance on the Pennsylvania System of School Assessment Using the Developmental Reading Assessment and the 4Sight Predictive Benchmark Assessment (Doctoral dissertation, Duquesne University). Retrieved from <https://dsc.duq.edu/etd/1241>

This Immediate Access is brought to you for free and open access by Duquesne Scholarship Collection. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Duquesne Scholarship Collection. For more information, please contact [phillipsg@duq.edu](mailto:phillipsg@duq.edu).

PREDICTING PERFORMANCE ON THE *PENNSYLVANIA SYSTEM OF SCHOOL*  
*ASSESSMENT* USING THE *DEVELOPMENTAL READING ASSESSMENT*  
AND THE *4SIGHT* PREDICTIVE BENCHMARK ASSESSMENT

A Dissertation

Submitted to the School of Education

Duquesne University

In partial fulfillment of the requirements for  
the degree of Doctor of Education

By

Todd B. Stoltz

December 2008

© Copyright by

Todd B. Stoltz

2008

**DUQUESNE UNIVERSITY  
SCHOOL OF EDUCATION  
INTERDISCIPLINARY DOCTORAL PROGRAM FOR  
EDUCATIONAL LEADERS**

***Dissertation***

Submitted in Partial Fulfillment of the Requirements  
For the Degree of Doctor of Education (Ed.D.)

***Presented by:***

Todd B. Stoltz  
B.A., Spanish, Lebanon Valley College, 1993  
M.ED., Curriculum and Instruction, Bloomsburg University, 2002

October 10, 2008

PREDICTING PERFORMANCE ON THE *PENNSYLVANIA SYSTEM OF SCHOOL  
ASSESSMENT* USING THE *DEVELOPMENTAL READING ASSESSMENT*  
AND THE *4SIGHT* PREDICTIVE BENCHMARK ASSESSMENT

***Approval***

\_\_\_\_\_, Chair  
Phillip Diller, Ed.D.  
Director, Professor of Educational Leadership  
Shippensburg University

\_\_\_\_\_, Member  
Winston Cleland, Ph.D.  
Retired Superintendent  
West Perry School District

\_\_\_\_\_, Member  
Jane Johnston, Ed.D.  
Director of Curriculum  
Tuscarora Area School District

***Program Director***

James E. Henderson, Ed.D.  
Professor of Educational Leadership and  
Director, Interdisciplinary Doctoral Program for Educational Leaders  
Duquesne University School of Education

## ABSTRACT

PREDICTING PERFORMANCE ON THE *PENNSYLVANIA SYSTEM OF SCHOOL  
ASSESSMENT* USING THE *DEVELOPMENTAL READING ASSESSMENT*  
AND THE *4SIGHT* PREDICTIVE BENCHMARK ASSESSMENT

By

Todd B. Stoltz

December 2008

Dissertation Supervised by Phillip Diller, Ed.D.

The purpose of the study was to determine the relationship between the *Developmental Reading Assessment (DRA)*, the *4Sight Reading Benchmark Assessment (4Sight)*, and the *Pennsylvania System of School Assessment (PSSA)* reading assessment for fifth grade students in one urban school district in Pennsylvania. A goal was to determine how well each assessment, the *DRA* and the *4Sight*, predicted student performance on the *PSSA*, and specifically to investigate whether the *DRA*, the *4Sight*, or both are significant predictors of student performance on the *PSSA*. An additional goal was to determine whether there was any added benefit to the prediction by using both assessments.

To study the relationship between the three assessments, a stepwise multiple regression approach was used to analyze student scores for a random sample of 40

students. Correlations between all assessments were significant, but only *4Sight* entered the regression equation as a significant predictor of *PSSA* scores. *4Sight* accounted for 60.5% of the variance in scores. Further analysis indicated that the *DRA* only contributed an additional 1.4%, making it an insignificant predictor.

Discussion of the results allows the reader to consider the value of both assessments based on format and philosophy. The *DRA* and the *4Sight* differ in format to one another and measure a student's reading level differently. The *DRA* is considered an authentic, performance based assessment and the *4Sight* is a more objective format that mirrors the format of the *PSSA*.

## DEDICATION

To my loving family, I dedicate this work. Your endurance through this process was as great as mine. Thank you for your support, love, patience, encouragement, and understanding of the importance of lifelong learning. This has truly been our journey together!

To my loving wife, Jodi, who is my best friend. Your sacrifice, support, and encouragement through this process has meant everything to me. You are a steady support.

To my loving children, Michael, Alexis, and Nicholas, who mean the world to me! Thank you for your patience throughout this journey. I hope that each of you will find the joy of lifelong learning as I have. No matter what, be happy in your life and pursue your interests with vigor, passion, and persistence!

## ACKNOWLEDGEMENT

The author would like to thank the members of the dissertation committee: Phillip Diller, Winston Cleland, and Jane Johnston for their wisdom and patience. You have been mentors in many ways outside of the dissertation itself. Thank you for your advice and support to me both professionally and personally. Thank you for ability to listen and offer meaningful feedback. It is your direction that has led me through this part of the journey.

A very special thank you to my family. It was your support that allowed me to devote countless hours to coursework and research. It was also your support that allowed me to juggle a career and coursework simultaneously. It goes without saying that all of us made sacrifices to see the end of this journey. To all of you, I am sincerely grateful!

To my colleagues, fellow cohort members, and my professors, I am forever indebted for stimulating conversation and being driven to reach high expectations. I am a better person because of the people that I have encountered throughout the IDPEL program. I hope that I have given as much back to you as you have given to me.

## TABLE OF CONTENTS

	Page
ABSTRACT.....	iv
DEDICATION.....	vi
ACKNOWLEDGEMENT.....	vii
LIST OF TABLES.....	xi
CHAPTER ONE.....	1
THE PROBLEM.....	1
Introduction.....	1
Statement of the Problem.....	3
Statement of Purpose.....	5
Historical and Philosophical Perspective.....	6
The Predictor Variables.....	9
The Criterion Variable.....	9
Research Questions.....	10
Need and Importance of Study.....	10
Definition of Terms.....	11
CHAPTER TWO.....	16
REVIEW OF THE LITERATURE.....	16
Introduction.....	16
Accountability in Education.....	16
State Assessment Plans.....	17
<i>Pennsylvania System of School Assessment</i> .....	19

Local Assessment Plans .....	21
<i>Developmental Reading Assessment</i> .....	23
<i>4Sight</i> .....	26
Formative Assessments as Predictors of Summative Assessments .....	29
Summary .....	30
CHAPTER THREE .....	31
METHODOLOGY .....	31
Introduction.....	31
Target Population.....	31
Method of Sampling .....	32
Measurement Devices .....	34
Data Collection Methods .....	35
Statistical Methods.....	35
Research Design and Procedures .....	38
CHAPTER FOUR.....	39
RESULTS .....	39
CHAPTER FIVE .....	44
DISCUSSION.....	44
Defining the Use of <i>4Sight</i> and <i>DRA</i> as Local Assessments .....	46
How Much Assessment Is Too Much.....	47
Possible Limitations of the <i>DRA</i> Score.....	47
Limitations of Study .....	48
Conclusions.....	49

Implications for Further Study.....	49
REFERENCES .....	51
Appendix A.....	59
22 Pa Code § 4.51 State assessment system .....	59
Appendix B.....	62
22 Pa Code § 4.52 Local assessment system .....	62

## LIST OF TABLES

	Page
1	Correlation of Concurrent Validity Between 5th Grade 4Sight Reading and PSSA .....28
2	Student Sample Distribution by Curriculum Code .....34
3	Student Sample Distribution by Ethnic Code .....34
4	Student Sample Distribution by Meal Status .....35
5	Descriptive Statistics for Student Assessment Scores .....40
6	Correlations between PSSA, DRA and 4Sight .....41
7	Stepwise Regression Model Summary for Variables Predicting PSSA Scores .....42
8	Backward Regression Model Summary for Variables Predicting PSSA Scores .....42

## CHAPTER ONE

### THE PROBLEM

#### Introduction

In the modern era of accountability, information gained from assessments has never been more critical to the field of education. States, districts, schools, and individual students are being held accountable for their progress toward educational outcomes. The *No Child Left Behind (NCLB) Act of 2001* (PL 107-110, 115 Stat. 1425) has mandated a rigorous assessment and accountability system for all students. Other federal legislation and regulations, such as Individuals with Disabilities Education Act (*IDEA*) (20 U.S.C. 1400 et seq.), and *Title I* (20 U.S.C. 6301 et seq.), ensure that all groups of students are represented and participate in formal state assessment programs.

Beyond representation, one of the major goals of the assessment and accountability movement is to ensure quality educational achievement for all students. “Under *No Child Left Behind*, states are working to close the achievement gap and make sure all students, including those who are disadvantaged, achieve academic proficiency” (United States Department of Education Website, 2008, p. 1). To ensure quality, all states have developed or, in collaboration with test developers, have adopted standardized assessments as their measures for reporting educational achievement for adequate yearly progress (AYP) reporting purposes. In Pennsylvania, the *Pennsylvania System of School Assessment (PSSA)* is used to assess students in reading and math in grades three through eight and grade eleven. Students in grades five, eight, and eleven are also assessed in writing. The *PSSA* is a summative assessment, used for accountability purposes to measure student achievement as an outcome of the student’s educational experiences (Assessment, 2006).

Summative assessments generally occur infrequently, i.e. annually, and make some levels of decision-making difficult. For example, summative assessments, such as the *PSSA*, are administered in spring and the data from the assessment is usually not available until late summer. By then, students who took the assessment have moved on to new teachers and new curricular expectations. The *PSSA* can be used to make some decisions regarding program quality or program effectiveness. However, it does not provide teachers with an immediate form of feedback that focuses on instructional improvement or individualization for students. For this, teachers generally rely on formative assessments. Formative assessments may be informal, teacher-made products or may be more formal, published products. “The crucial feature is that evidence is evoked, interpreted in terms of learning needs, and used to make adjustments to better meet those learning needs” (William, 2006, p. 285). An example of a more formal, published formative assessment is the *Developmental Reading Assessment Second Edition (DRA)* created by Joetta Beaver and Mark Carter (Beaver & Carter, 2003a). The *DRA* is designed to provide teachers with immediate feedback and information to make instructional decisions for individual students (Beaver & Carter, 2003b).

Perhaps most important, teachers want assessments that will both guide their instruction and be consistent with skills and performance tasks that are being measured on more formal state assessments. According to Nitko (2004), assessments or their results should not surprise educators. Assessments should be an indicator of expected performance. Therefore, teachers want to use formative assessments that indicate performance on a summative assessment. For example, Paris and Carpenter (2003) found that teachers who used informal reading inventories (IRI) also found them worthwhile to

report adequate yearly progress (AYP) for students. These types of assessments provide educators with information to adjust their current instruction, while at the same time measure students' performance relative to state academic content standards, thus making assessments that much more efficient and effective.

### Statement of the Problem

The Pennsylvania Department of Education (PDE) regulations do not allow school districts to administer the *PSSA* to students on a repeated basis in one school year. The format of the assessment is lengthy and does not necessarily assess students' ability while performing a task in an authentic manner. The *PSSA* consists of a multiple choice format, along with some constructed response items, where students are using old and new knowledge to construct an answer to a question (Mueller, 2006). This format does allow for measurement of reading comprehension, but does not measure reading comprehension in an authentic manner. Authentic assessments must be directly educationally meaningful.

For example, reading several long works and using them to compare and contrast different social viewpoints is directly meaningful because it is the kind of thoughtful reading educated citizens do. Reading short paragraphs and answering questions about the "main idea" or about what the characters in the passage did, on the other hand, is indirectly meaningful because it is only one fragment or component of the ultimate learning target of realistic reading. (Nitko, 2004, p. 248)

While some of the constructed response items qualify as authentic assessment, multiple-choice format, along with the description cited above, do not qualify as authentic assessment by this definition.

To measure reading comprehension authentically, teachers often rely on informal, formative assessments such as running records and anecdotal notes of a student's reading performance. Informal reading inventories (IRI) "were designed to provide teachers with

information about young children's developing oral reading fluency, retelling, and comprehension" (Paris, 2002, p. 168). Paris (2002) goes on to state that, "IRIs are regarded as authentic for young children and appropriate for teachers" (p. 168). Answering a multiple choice question does not analyze student performance while reading the text. It assumes that students read for meaning, if in fact they have not guessed at the answer. Authentic assessment analyzes a learner's skills through observable indicators (Chapman & King, 2005).

The *DRA* is a formative assessment that is used by classroom teachers throughout the nation to assess students' reading performance authentically (Washtenaw Intermediate School District Website, 2008). The *DRA* is typically administered three times per year and is used to guide teacher instruction on a regular basis. To ensure reliability, professional development in the use and administration of the *DRA* is ongoing. Professional development is offered to all new teachers, and refresher sessions are provided to veteran teachers in the scoring, analysis, and implications for instruction. It is assumed that the *DRA* can serve as a measuring stick of student performance in grade five relative to more formal assessments, such as the *PSSA*. While the format for assessing students' reading comprehension is different on the two assessments, the results indicate proficiency relative to a grade level standard. A *DRA* level 50 in fifth grade is considered on grade level (Beaver & Carter, 2003b), which is congruent with a *PSSA* performance level proficient.

Because informal assessments present challenges such as scoring reliability, interrater reliability, and assessment format to name a few, the *4Sight* assessment is an alternative formative assessment that has been developed to reduce reliability issues and

to increase validity relative to the *PSSA*. The *4Sight* reading assessment is similar to the *PSSA* in format, administration, and scoring. “*4Sight* assessments are one-hour tests that have exactly the same formats, coverage, look, and feel as your state reading assessments. They produce overall scores that predict students’ scores on your state assessment” (*4Sight* reading and math benchmarks, 2006, p. 1). The assessments can be administered to students in grades three through eight up to five times per school year. If the online versions of the assessment are used, then feedback in the form of data and reports are available almost instantaneously. These reports allow teachers to target their instruction toward specific state standards that have not been mastered by students.

#### Statement of Purpose

The purpose of the study is to determine the relationship between the *DRA*, the *4Sight*, and the *PSSA*. Realizing that these three assessments are different in construct and format, the *DRA* being more subjective than the other two, the purpose of this study is to explore the relationship between the *PSSA* and the *DRA* and the relationship between the *PSSA* and the *4Sight* benchmark reading assessment. A goal will be to determine how well each assessment, the *DRA* and the *4Sight*, predicts student performance on the *PSSA*. Additionally, this research will focus on the continued use of the *DRA* and *4Sight* as assessment tools for both informing instruction and ultimately guiding instruction to improve student achievement on the *PSSA*. Specifically, the study intends to investigate whether the *DRA*, the *4Sight*, or both are predictors of student performance on the *PSSA*. Does one of these assessments predict performance on the *PSSA* better than the other, and if so, should any school district continue to use both assessments?

## Historical and Philosophical Perspective

Assessment in education has a long history throughout all cultures. Madaus and O'Dwyer (1999) described the history of performance assessment throughout the world and the United States. Assessments have often been used to determine placement in programs, jobs, or otherwise life paths for people. In the early part of the 20<sup>th</sup> century, intelligence testing was foremost. The Army Alpha represented one of the earliest versions of large scale assessment. It was a group-administered IQ test that was developed to meet the demand for an assessment that was “more efficient, manageable, easily scored, and easily recorded” (p. 693). Over the years, large scale assessment procedures have been established, including the use of technology to administer and score assessments quickly and reliably. These advances allowed for large scale use of assessments, but “In the late 1980s standardized multiple-choice testing came under criticism, and the movement for “new” and “authentic” assessment gained momentum” (Madaus and O'Dwyer, 1999, p. 693).

Over the last 20 years, there has been significant legislation that has impacted the use of assessments in all 50 states. Goertz and Duffy (2003) described “the landscape of high-stakes testing and accountability programs” (p. 4). The Title I of the Improving America's Schools Act (IASA) of 1994, Individuals with Disabilities Education Act (IDEA) of 2004 with subsequent revisions and reauthorizations, and the No Child Left Behind (NCLB) Act of 2001 are all examples of legislation that have impacted the use of assessments in all 50 states. These examples of legislation have determined the purpose and scope for state assessments. States were initially required to develop high quality assessments aligned with state standards in reading and math. NCLB added that all

students in grades 3 through 8 would be assessed in reading and math by 2005-2006. As a result, accountability for results in student achievement has never been more intense.

With accountability comes sanctions for underperformance. According to the United States Department of Education website (No Child Left Behind Executive Summary, 2002), NCLB requires the identification of underperforming schools and school districts, along with school choice options and school restructuring. Performance targets have been set by individual states, but 100% of students must be proficient or above by school year 2013-2014, as required by NCLB. Again, the combination of legislation aforementioned requires the inclusion of all students in state assessments.

Although Pennsylvania does not use the *PSSA* as a promotion or graduation requirement, there are certainly high-stakes consequences to the results of the *PSSA*. In an analysis of the *PSSA* cut scores (Zwerling, 2003), the Pennsylvania State Education Association (PSEA) described why cut scores and, ultimately, scores on the *PSSA* matter. Because a cut score defines success or failure, Zwerling (2003) noted the following:

For students, it has the potential to impact class placements, course selections, college decisions, and lifetime earnings. For educators, it can affect job benefits and security, as well as curriculum and instructional decisions. For schools and districts, it can result in privatization. For taxpayers, it can even result in upward- or downward- pressure on housing values. (p. 10)

It would be advantageous to schools and school districts to know where students stand at any given moment in time. The *PSSA* is administered annually in early spring and results are received over the summer from this administration period. Since many funding sources, including federal Title I funds and other school improvement funds, are targeted for improving student achievement, the effective use of assessment data is important for determining effective teaching and learning. "Assessment is closely related

to instruction and achievement. Its purpose is to inform teachers about the effectiveness of their teaching and the differences in students' learning" (Williams, 1996, p. 89).

School districts are required by the Pennsylvania Department of Education (Appendix B) to administer local assessments to monitor and inform instructional practices. These assessments are less formal and generally shorter than a state assessment. The *DRA* is one example of a less formal assessment and serves these purposes in many districts. It provides assessment data about a child's progress in reading. Data gathered from this assessment is then analyzed and used to make instructional recommendations for students. Since the *DRA* is a criterion-referenced test (Thomas, 2002), determinations about a student's performance level relative to a set of criteria are similar to the set of criteria measured by the *PSSA*.

Recently, the *4Sight* predictive benchmark assessment for reading has been made available to school districts in Pennsylvania. The *4Sight* is designed to match the format, coverage, look, and feel of the *PSSA* (*4Sight* reading math benchmarks, 2006). Since the *4Sight* is administered quarterly, it allows teachers to assess student achievement in reading and adjust instruction to meet student needs. Although the *4Sight* is a formative assessment, it requires a more formal administration than the *DRA* and other similar assessments. For the researcher, this helps limit threats to validity and reliability, because it follows similar administration guidelines used with the *PSSA*.

In summary, assessment of learning has evolved over time to increase accountability for states, school districts, schools, and ultimately individual students. Federal legislation has had significant impact on state assessments, including their development, purpose, and scope. As well, federal legislation, such as *NCLB*, *IDEA*, and

other federal programs, has determined the inclusion of all students in state testing programs, as well as defined the accountability and performance levels for states, school districts, and schools. Schools are charged with improving student achievement, as measured by their state assessments. In order to estimate with some confidence the level of performance that a student will demonstrate on the state assessment, schools must rely on more periodic and less formal assessments. These assessments also guide and inform instruction that intends to improve student achievement on the state assessment. By studying the relationship of the *DRA*, the *4Sight*, and the *PSSA*, further insight will be gained about the use of a criterion-referenced test to predict student performance on a state assessment.

#### The Predictor Variables

The predictor or independent variables in this study will be student performance on the *Developmental Reading Assessment (DRA)* and the *4Sight* benchmark reading assessment in grade five. Student reading performance on these two assessments is used to determine how well each assessment can predict student reading performance on the statewide assessment, *PSSA*.

#### The Criterion Variable

The criterion or dependent variable in this study will be student performance on the *PSSA* reading test for students in grade five during the 2007 administration. Student reading performance on the *PSSA* is used as the criterion variable because it is the basis for determining student and school performance relative to the attainment of PA state academic content standards.

## Research Questions

1. What is the relationship between student performance on the *DRA* and the *PSSA* in grade five?
2. What is the relationship between student performance on the *4Sight* and the *PSSA* in grade five?
3. Do fifth grade *DRA* scores predict performance on the *PSSA* for students in grade five?
4. Do fifth grade *4Sight* scores predict performance on the *PSSA* for students in grade five?
5. Which formative assessment, the *DRA* or the *4Sight*, is a better predictor of student performance on the *PSSA*?
6. Does using both formative assessments together add significantly more power to the prediction of *PSSA* scores?

## Need and Importance of Study

Accountability in education has never been greater. State assessments have high-stakes implications for states, districts, schools, and individual students. For states, districts, and schools, funding is dependent on student achievement. The consequences for districts and schools range from instituting systematic programs for instruction to takeover by governmental agencies, such as the Pennsylvania Department of Education. Although Pennsylvania does not require minimum passing scores for graduation or promotion purposes, students are judged on their performance on the *PSSA* and thus may not be eligible for special programs or scholarships as a result. The potential impact for students is great. High stakes testing “has the potential to impact class placements, course selections, college decisions, and lifetime earnings” (Zwerling, 2003, p. 10).

The *PSSA* has become the single criterion-referenced measure used in school districts to determine student progress in grades three through eight and grade eleven. Since the *PSSA* is the mandated assessment, other standardized assessments such as the Stanford Achievement Test and Iowa Test of Basic Skills, have been discontinued in many districts. However, these formal assessments have given way to more informal, formative assessments. Formative assessments are used more frequently and allow teachers to receive immediate feedback for making instructional decisions. Since the *PSSA* is a high-stakes test, teachers need formative assessments that not only inform their daily practice but assessments that will lead to student success on the *PSSA*. In reviewing ten years of assessment research, Broadfoot and Black (2004) note “there is far more work to be done if the optimum synergy between these two, and so between assessment for learning and assessment for certification and accountability, is to be achieved” (p. 18).

It is imperative that the relationship between formative assessments and summative assessments is studied. These two forms of assessments serve similar but different purposes and should be used in conjunction with one another to improve student achievement. “While state tests provide a snapshot of a student’s performance on a given day under test conditions, formative assessments allow teachers to monitor and guide students’ performance over time in multiple problem-solving situations” (Boston, 2002, p. 4). Formative and summative assessments must work in tandem to provide the whole picture of student learning.

#### Definition of Terms

*4Sight*: “*4Sight* assessments are one-hour tests that have exactly the same formats, coverage, look, and feel as your state reading assessments. They produce overall scores that predict students’ scores on your state assessment” (*4Sight* reading and math

benchmarks, 2006, p. 1). The *4Sight* assessments are available in reading and in math for grades three through eight in many states, including Pennsylvania. They are predictive benchmark assessments that are intended to inform classroom instruction for the purpose of improving student achievement. The *4Sight* has been developed by the Success for All Foundation and are available for purchase.

Adequate Yearly Progress (AYP): “AYP is an individual state's measure of progress toward the goal of 100 percent of students achieving to state academic standards in at least reading/language arts and math. It sets the minimum level of proficiency that the state, its school districts, and schools must achieve each year on annual tests and related academic indicators” (Answer, 2008, p. 1).

Authentic Assessment: “An analysis of the learner’s skills, abilities, and strengths through a variety of observable indicators. This includes skill performances, purposeful activities, portfolios, demonstrations, hands-on experiences, and projects” (Chapman & King, 2005, p. xxi). Nitko (2004) defines authentic assessment as “a type of performance assessment in which students are presented with tasks that are directly educationally meaningful instead of indirectly meaningful” (p. 513).

Criterion Variable: The criterion variable in this study will be student performance on the *PSSA* reading for students in grade five during the 2007 administration. Student reading performance on the *PSSA* is used as the criterion variable because it is the basis for determining student and school performance relative to the attainment of PA state academic content standards.

DRA: The *Developmental Reading Assessment* is a classroom-based reading assessment that is designed to 1) monitor student growth relative to skills and strategies utilized by

successful readers, 2) help teachers diagnose student needs and inform instruction, 3) prepare students to meet classroom and testing expectations, and 4) inform all stakeholders about the level of student achievement (Beaver & Carter, 2003b). The *DRA* is published by Pearson Learning, Inc. and is available for purchase.

Formative Assessment: Assessment that is used periodically by classroom teachers to provide 1) feedback to students about their learning and 2) feedback to teachers about their instructional practices to meet student learning needs (Boston, 2002).

PASA: “The *Pennsylvania Alternate System of Assessment (PASA)* uses performance tasks to measure the knowledge and skills attainment of students with significant cognitive disabilities” (*PA Alternate System of Assessment (PASA)*, 2008). The *PASA* is designed to inform educators about future instruction of students with disabilities.

Predictor Variables: The predictor variables in this study will be student performance on the *Developmental Reading Assessment (DRA)* and the *4Sight* benchmark reading assessment in grade five. Student reading performance on these two assessments is used to determine how well each assessment can predict student reading performance on the statewide assessment, *PSSA*.

PSSA: The *Pennsylvania System of School Assessment*. The state assessment that is given in reading and math annually to all students in grades three through eight and eleven in the state of Pennsylvania. The Pennsylvania Department of Education describes the *PSSA* as “a standards based criterion-referenced assessment used to measure a student’s attainment of the academic standards while also determining the degree to which school programs enable students to attain proficiency of the standards” (p. 1).

PSSA Performance Levels: The PDE has determined general performance level descriptors for advanced, proficient, basic, and below basic performance on state assessments. These performance levels are broadly defined by the PDE as:

Advanced- This level reflects superior academic performance. Advanced work indicates an in-depth understanding and exemplary display of the skills included in the Pennsylvania Academic Content Standards.

Proficient- Proficiency reflects satisfactory academic performance. Proficient work indicates a solid understanding and adequate display of the skills included in the Pennsylvania Academic Content Standards.

Basic- This level reflects marginal academic performance. Basic work indicates a partial understanding and limited display of the skills included in the Pennsylvania Academic Content Standards. This work is approaching satisfactory performance, but has not been reached. There is a need for additional instructional opportunities and/or increased student academic commitment to achieve the proficient level.

Below Basic- The lowest level reflects inadequate academic performance. Below basic work indicates little understanding and minimal display of the skills included in the Pennsylvania Academic Content Standards. There is a major need for additional instructional opportunities and/or increased student academic commitment to achieve the proficient level. (PDE Website, 2006)

Summative Assessment: Assessment that results in a judgment or evaluation, usually a score or a grade, about student learning after a period of judgment (Boston, 2002).

CHAPTER TWO  
REVIEW OF THE LITERATURE

Introduction

Accountability in Education

Of the most sweeping changes in American education over the last 40 years, the No Child Left Behind (NCLB) Act of 2001 (PL 107-110, 115 Stat. 1425) has arguably had the most significant impact on school reform and accountability for results in student achievement. It requires that all students in grades three through eight are assessed in reading and mathematics on an annual basis, relative to each individual state's standards. Test developers, policymakers, and teachers are challenged to coexist in an era of accountability (Reeves, 2002; Invernizzi, Landrum, Howell, and Warley, 2005). While educators recognize that some level of accountability is required, NCLB requires research-based, scientifically proven assessments and instructional techniques. However, "there may be significant challenges associated with selecting assessment tools and implementing a comprehensive yet efficient assessment program that (a) meets high standards of scientific rigor and (b) provides teachers with instructionally useful information" (Invernizzi et al., 2005, p. 610).

Because there are rewards and sanctions as part of this rigorous assessment and accountability movement, schools and school districts have become more concerned with alignment of their curriculum with the state standards (Wraga, 1999). As a result, districts have developed local assessment plans to meet the demands of the state assessment program and to ensure curricular alignment to state standards. In addition, local assessment plans can be used to ensure adequate progress in student achievement over shorter periods of time. It is this feature that poses a significant concern about reliability

of local assessments, but simultaneously offers useful information to inform instruction at the classroom and individual student level. In the 1999 National Center for Research on Evaluation, Standards, and Student Testing (CRESST) Conference Proceedings, Richard Elmore concluded,

Schools cannot respond to external accountability systems unless there are internal accountability systems. School organizations must be healthy, monitoring their own performance as a matter of routine, with the district role being one of helping schools to monitor the quality of instruction and to develop their own assessments. (Lewis, 2001, p. 7)

Without a high quality local assessment program aligned with the state assessment, a school district's ability to improve student achievement based solely on annual progress data is significantly compromised.

#### State Assessment Plans

As previously mentioned, states have been given the task of developing assessments to measure the educational progress of students toward state standards. In an NCLB executive summary, the outline for states was summarized as follows:

The NCLB Act will strengthen Title I accountability by requiring States to implement statewide accountability systems covering all public schools and students. These systems must be based on challenging State standards in reading and mathematics, annual testing for all students in grades 3-8, and annual statewide progress objectives ensuring that all groups of students reach proficiency within 12 years. (NCLB Executive Summary, 2002, p. 1)

As a measure of this progress, states were directed to identify performance categories. Goertz and Duffy (2001) reviewed and described the assessment and accountability systems in the 50 states during 1999-2000. Their study indicated that 48 states used "a state assessment as the principal indicator of school performance" (p. 3). The other two states left the choice of assessment instrument to local schools districts. The study noted that most state assessment systems begin in the third grade, but in many states local

districts use assessments to test reading in the early grades. Although “the expansion of state testing programs has reduced the role of local tests in state assessments and accountability systems” (p. 7), other states incorporate local assessment results in the accountability programs. The study noted Colorado, Maine, Vermont, Kentucky, California, Maryland, Texas, and Michigan as states whose legislature required or chose to supplement state tests with local assessments in order to monitor student progress throughout the year. Finally, the study described the multiple-measure requirement facing states and districts. “Because one assessment format tends to emphasize only one aspect of a complex learning target, it typically underrepresents that learning target” (Nitko, 2004, p. 6), no single assessment should be the sole basis for educational decisions that have significant impact on students.

In Turner and Williams (2002), a study was conducted to predict student proficiency on the Arkansas Benchmark Examinations; the Arkansas state assessment. The study was designed to predict student performance that would be below proficiency on the Arkansas state assessment that was given in fourth grade. The study noted the following about state assessments:

One limitation to this type of assessment system is that large amounts of data are obtained from a variety of sources, but commonly, the information is not integrated and used for diagnostic or predictive purposes in order to develop more individualized and academically appropriate learning environments. (p. 52)

The authors go on to note that “these exams are only given periodically”. To counter this concern, the study examined other assessments of student learning to predict future performance on the state assessment. The study concluded that SAT-9 Total Reading and Spelling were the best predictors of student performance on the Arkansas state reading assessment in fourth grade, accounting “for 34% to 59% of the variability in fourth grade

literacy scores using SAT-9 scores from the spring of grade 1 to the spring of grade 3” (Turner & Williams, 2002, p. 65).

The use of formative assessments and local assessments as part of the state accountability program satisfies the multiple-measure requirement. In Pennsylvania, “in order to graduate a student must demonstrate proficiency in reading, writing, and mathematics on the *PSSA* or a local assessment that is aligned with the *PSSA* and state academic standards” (Girton & Buckheit, 2006, p. 7). Classroom level assessments that are created by teachers or used by teachers regularly become extremely useful in determining student progress relative to state standards. Reeves (2002) described the elements of holistic accountability systems for schools and school systems. In discussing teacher rigor and determination of student attainment of standards, he stated, “Teacher-created classroom-level assessments are more likely to indicate a student does not meet the academic content standards than are the results of a standardized test” (p. 47). Black & William (1998) concluded that the use of formative assessments produce significant learning gains measured by test scores compared to similar groups of students not exposed to formative assessments. Finally, Paris and Hoffman (2004) described the joining of state assessment programs and commercial reading assessments to be used as formative and summative assessments of early reading.

#### *Pennsylvania System of School Assessment*

As a result of NCLB, Pennsylvania has developed its version of an annual statewide assessment. Initially, students were assessed annually in reading and mathematics in grades five, eight, and eleven. Currently, the assessment program has expanded to include all grades, three through eight, and eleven, to continue to be assessed in reading and mathematics. Grades five, eight, and eleven are also tested in writing and

assessment in science is forthcoming. The Pennsylvania Department of Education (PDE) describes the *PSSA* as follows:

The annual *Pennsylvania System of School Assessment (PSSA)* is a standards based criterion-referenced assessment used to measure a student's attainment of the academic standards while also determining the degree to which school programs enable students to attain proficiency of the standards. (Assessment, 2006, p. 1)

The PDE has determined general performance level descriptors for advanced, proficient, basic, and below basic performance on state assessments. These performance levels are broadly defined by the PDE as:

Advanced- This level reflects superior academic performance. Advanced work indicates an in-depth understanding and exemplary display of the skills included in the Pennsylvania Academic Content Standards.

Proficient- Proficiency reflects satisfactory academic performance. Proficient work indicates a solid understanding and adequate display of the skills included in the Pennsylvania Academic Content Standards.

Basic- This level reflects marginal academic performance. Basic work indicates a partial understanding and limited display of the skills included in the Pennsylvania Academic Content Standards. This work is approaching satisfactory performance, but has not been reached. There is a need for additional instructional opportunities and/or increased student academic commitment to achieve the proficient level.

Below Basic- The lowest level reflects inadequate academic performance. Below basic work indicates little understanding and minimal display of

the skills included in the Pennsylvania Academic Content Standards. There is a major need for additional instructional opportunities and/or increased student academic commitment to achieve the proficient level. (PDE Website, 2006)

According to The Pennsylvania Code, 22 Pa Code § 4.51 (Appendix A) describing the state assessment system concerning reading, the criteria for judging performance on state assessments is as follows:

1. Performance on State reading assessments shall be demonstrated by students' responses to comprehension questions about age-appropriate reading passages and by their written responses to in-depth comprehension questions about the passages.

#### Local Assessment Plans

Local assessment plans have become increasingly important because of their impact on student achievement on state assessments. While Juel (1988) supports early intervention for students as early as possible, there remain large numbers of students who are now in intermediate grades and are still not proficient readers. According to the United States Department of Education (USDoE), President George W. Bush's Striving Readers Program addresses this population. The program targets students who are reading below grade level in grades 6-12. However, there remains a significant gap in services for students in fourth and fifth grades. The USDoE has distributed more than one billion dollars to Reading First, in which "Funds are dedicated to help states and local school districts eliminate the reading deficit by establishing high-quality, comprehensive reading instruction in kindergarten through grade 3" (Reading First, 2007, p. 1). President Bush requested one hundred million dollars for the Striving Readers Program in fiscal

year 2007, “to reach more secondary students in grades 6-12 who are reading below grade level and at risk of dropping out of school” (United States Department of Education Website, 2006, p. 1). However, the program was funded at \$31,870,000 (United States Department of Education Website, 2008, p. 1).

Besides federal legislation, states have responded by increasing the accountability for results on state performance assessments, particularly in reading. For example, Florida, Ohio, and Texas have all required attainment of proficiency on the state reading assessment in order to be promoted to the next grade (Greene & Winters, 2004). Ohio referred to their guidelines as the “Fourth Grade Guarantee” (Morris, 2002). In Pennsylvania, “in order to graduate a student must demonstrate proficiency in reading, writing, and mathematics on the *PSSA* or a local assessment that is aligned with the *PSSA* and state academic standards” (Girton & Buckheit, 2006, p. 7).

Chapter 4 of the Pennsylvania Code describes the academic standards for Pennsylvania and the assessment plans at the local and state levels. Specifically, 22 Pa Code § 4.51 directs each school district to design an assessment system that uses assessment results to improve instructional practices and uses a variety of assessment strategies. While these guidelines do not suggest or recommend specific products or assessments, the assessments described in these guidelines need to be formative and at the same time summative. According to William (2006), “an assessment of a student is formative if it shapes that student’s learning. Assessments are formative ... if something is contingent on their outcome, and the information is actually used to alter what would have happened in the absence of the information” (p. 284). Therefore, formative assessments are a critical element in the local assessment plan to report summative data

as evidence of achievement of academic standards and to inform classroom practices that relate to increased student achievement on the state assessment.

According to 22 Pa Code § 4.52 (Appendix B), the local assessment system should do the following:

1. Determine the degree to which students are achieving academic standards.
2. Use assessment results to improve curriculum and instructional practices and to guide instructional strategies.

These do not represent the entire list of expectations for the local assessment system, but do indicate the specific areas relevant to the purposes of determining proficiency and informing instruction. This same code contains examples of a variety of assessments strategies that may include the following:

1. Written work by students.
2. Other demonstrations, performances, products or projects by students related to specific academic standards.
3. Examinations developed by teachers to assess specific academic standards.
4. Other measures as appropriate, which may include standardized tests.

Again, this list is not exhaustive, but it specifically pertains to examples that might be used in the assessment of student reading performance, relative to state standards.

#### *Developmental Reading Assessment*

The *Developmental Reading Assessment Second Edition 4-8* is a classroom-based reading assessment that was developed by Joetta Beaver and Mark Carter in 2003. It serves as a reading assessment for middle-childhood students, and is an extension of the *DRA K-3*. According to the *DRA Teacher Guide* (Beaver & Carter, 2003b), “the primary purposes of the assessment are to:

1. monitor student growth on a variety of crucial skills and strategies that successful readers utilize,
2. help teachers diagnose student needs and plan for timely instruction,
3. prepare students to be successful at meeting today's classroom and testing expectations, and
4. support teachers and school districts in keeping parents and other stakeholders informed about the level of student achievement. (p. 4)

The *DRA* has been used as part of the local assessment plan. As described above, it has served a dual purpose of reporting achievement toward academic standards and informed classroom practices. In Williams (2003), teachers reported finding the *DRA* helpful in among other things, "identifying students who may be reading below proficiency" (p. 11). This is an important feature because the local assessment plan should complement the state assessment system. A complete listing of the correlation of the *DRA* to Pennsylvania state standards was prepared by Pearson Publishing. This correlation allows a teacher to periodically assess a student's reading level and make informed, instructional decisions that are based on Pennsylvania state reading standards. Thus, instruction that is individualized and targeted specifically to the attainment of standards is not only possible, but expected by the regulations set forth by the PDE regarding local assessment plans in 22 Pa Code § 4.52 (Appendix B).

Morris (2002) reported choosing the *DRA* as the assessment instrument to complement the school's reading program. Describing the increasing pressure in Ohio to ensure student achievement in reading, this study described one school's attempt to address the learning needs of individual students with an appropriate formative

assessment. In addition to the results presented, a thorough review of states that either required or strongly recommended use of the *DRA* as part of their local assessment plans was included. States, including Louisiana and Washington, have done so in an attempt to identify and intervene early with struggling readers in the primary grades (Louisiana Department of Education, 1998). According to Shaywitz & Shaywitz (2006), “Seventy-five percent of children who struggle to read in third grade will continue to struggle throughout school” (p. 2).

While Morris (2002) primarily focused on the *DRA* K-3, much of the rationale used in the study was based on significant state level response to a growing epidemic of students being promoted through the elementary grades that lacked basic skills in reading. Ohio Senate Bill 55 (1997, as cited in Morris (2002)), initiated the “fourth grade guarantee,” which prohibited schools from promoting any student to fifth grade that failed the reading proficiency test in fourth grade. Since then, additional states, including Florida, Ohio, and Texas have required proficiency on state reading tests for promotion.

Finally, Williams (1999) described the reliability study for the *DRA*. Interrater reliability was good at 0.80 between the first two raters and fair at 0.74 between three raters (p. 5). When measuring construct validity, the *DRA* was correlated with the Iowa Test of Basic Skills (ITBS) Subscales: Vocabulary, Reading Comprehension, and Total Reading. Its sample (N=2470) of second grade students, whose spring *DRA* scores were correlated with their fall third grade ITBS scores, were found to be significant at the 0.01 level (2-tailed), with “the highest and most meaningful correlation for this assessment was with Total Reading ( $r=0.71$ ,  $p<.01$ )” (Williams, 1999, p. 6).

## *4Sight*

*4Sight* predictive benchmark assessments have been developed by the Success for All Foundation “to provide a formative evaluation of student progress that predicts how a group of students would perform if the *PSSA* were given on the same day” (Success for All Foundation, 2007a, p. 3). *4Sight* is “a quarterly benchmark assessment tool that helps you predict how your students will perform on state assessments, and provide useful data for focusing professional development and instructional goals” (*4Sight* reading and math benchmarks, 2006, p. 1). *4Sight* is both a predictive and formative assessment. Like other formative assessments, information gained from the *4Sight* will guide and focus instruction. According to the *4Sight* Pennsylvania Benchmark Administration and Scoring Guide (2007), “The formative data provided from these benchmarks will allow you to monitor student achievement progress over the course of a school year, making adjustments to instruction and professional development when needed” (p. 1). “The Foundation’s exams are designed to be shorter, formative assessments that will predict success on the longer, summative assessments used by the state” (Success for All Foundation, 2007a, p. 18).

The *4Sight* is a one-hour assessment that is designed to mirror the *PSSA* in format, coverage, look, and feel. According the CDDRE website, *4Sight* assessments are available in many states, including Pennsylvania. *4Sight* assessments are unique to each state since the academic content standards being measured may be different. According to the Success for All Foundation website, “*4Sight* provides a state-specific set of mini-tests so that educators can estimate how students are likely to perform throughout the year – and take immediate action in areas in which students need help. Each assessment has exactly the same format, coverage, look and feel as a state’s reading assessment”

(Success for All Website, 2007). In addition, the assessments were designed “so that schools and districts could use the assessment results to inform instruction and track progress toward proficiency during the course of a school year” (Success for All Foundation, 2007a, p. 3).

To ensure the validity and reliability of the *4Sight* assessments, a variety of processes were followed. With regard to cultural sensitivity, the assessments “were rigorously reviewed for race/ethnicity and gender bias by Success for All Foundation staff using the California Standards for Evaluation of Instructional Materials with Respect to Social Content (1986 Edition)” (Success for All Foundation, 2007a, p. 4). According to the technical report, test blueprints were developed using materials and information provided by the Pennsylvania Department of Education (PDE) so they would closely match the *PSSA*.

“In Pennsylvania the *4Sight Reading and Math Benchmarks* were piloted in a variety of districts” (Success for All Foundation, 2007a, p. 13). The assessments were administered “under conditions similar to those used in state testing” (Success for All Foundation, 2007a, p. 13). Also, inter-rater reliability was established by having Success for All Foundation staff complete the scoring on open-ended items, “using the state-specific rubrics or scoring guidelines” (Success for All Foundation, 2007a, p. 14). “*4Sight* pilot assessment scores and state test scores were matched using student numbers and correlations developed using a linear regression to provide an estimated performance of students on the state’s high-stakes reading assessment” (Success for All Foundation, 2007a, p. 13). To ensure that the most current data available was used to correlate student performance, “reading benchmarks forms 1 and 2 for all grade levels were piloted in

spring 2006 and re-correlated with spring 2007 *PSSA* data” (Success for All Foundation, 2007a, p. 13). “These forms [Reading benchmark forms 3 to 5] were piloted in spring 2007 and correlated with 2007 *PSSA* scores” (Success for All Foundation, 2007a, p. 13). Table 1 represents the concurrent validity of the 5<sup>th</sup> grade *4Sight Reading Benchmarks* and *PSSA* as reported in the technical report by the Success for All Foundation.

---

Table 1

*Correlation of Concurrent Validity Between 5<sup>th</sup> Grade 4Sight Reading and PSSA*

---

Form Number	R
Form 1	0.86
Form 2	0.87
Form 3	0.82
Form 4	0.81
Form 5	0.80

---

“*4Sight* scores were obtained from an administration window of March through May 2007. These scores were matched with spring 2007 *PSSA* scores” (Success for All Foundation, 2007a, p. 18). Generally speaking, with correlations greater than or equal to  $r=0.80$ , as is the case in all forms of the *4Sight*, the correlations are considered strong. A Pearson Correlation for inter-form reliability was completed for all *4Sight* assessments, and for grade 5, the average correlation was 0.76 and average N was 22,500 (Success for All Foundation, 2007a, p. 20).

## Formative Assessments as Predictors of Summative Assessments

Summative assessments, particularly state assessments, are useful and critical to judging academic performance of states, school districts, schools, and students. A review of research that analyzes the use of formative or local assessments to predict future performance on state assessments is used to inform the design of this study and to consider issues in making predictions. In addition to analyzing predictability, the varied purposes that schools define for local assessments and the approaches that schools use with these measures will reveal broad implications for this type of research.

In May 1998, the Louisiana State Board of Elementary and Secondary Education approved the *DRA* “as the uniform assessment instrument to be implemented statewide in the fall of 1998” (Louisiana Department of Education, 1998, p. 3). The Department indicated that the *DRA* was appropriate for assessing grade-level reading abilities and the *DRA* would provide a “more accurate and valid picture” of the reading abilities of students. From the Fall 1998 administration of the *DRA*, the Louisiana Department of Education determined that 43.64% of the second graders were reading on or above grade level, 65.52% of the third graders were reading on or above grade level, and 54.49% of the second and third graders were reading on or above grade level (Louisiana Department of Education, 1998). Finally, Thomas (2002) used a stepwise-multiple regression procedure to determine which predictor variables, including the *DRA*, were the most predictive of scores on the Louisiana Educational Assessment Program for the 21<sup>st</sup> Century (LEAP 21). The *DRA* was found to be the second most predictive variable after Iowa Test of Basic Skills (ITBS) total national percentile ranking. Thomas (2002) stated the following:

The *DRA* score added .057 to the  $R^2$  for ELA LEAP 21 scores, .022 to the  $R^2$  for Mathematics LEAP 21 scores, and .045 to the  $R^2$  for the combined ELA and Mathematics LEAP 21 scores. Although this was not very much, it was more than all of the other predictive variables combined added to the total predictive models. (p. 88)

As previously discussed, the *4Sight* assessment has been designed to mirror the *PSSA*. Its correlations make it a strong predictor of student performance on the *PSSA*. However, while the *4Sight* is a formative assessment, its construct and design differ from the *DRA* in that it is less authentic in its performance assessment of student reading than the *DRA*.

### Summary

The review of literature included a description of accountability in education for states and public schools under NCLB. With an increased accountability for student achievement, states were required to develop accountability systems to ensure that students were attaining proficiency relative to state academic content standards. Pennsylvania uses the *PSSA* as its assessment of students in grades three through eleven in reading to determine proficiency of state academic content standards. As required under state assessment plans, local school districts are required to develop local assessment plans that are designed to inform schools how well students are attaining state academic content standards and are to be used to inform curricular and instructional practices. Two formative assessments that are being used in some districts as part of the local assessment plan are the *DRA* and the *4Sight* assessments. The review of literature concluded with a brief discussion about the use of formative assessments, particularly the *DRA* and *4Sight*, as predictors of performance on criterion-referenced assessments.

## CHAPTER THREE

### METHODOLOGY

#### Introduction

This section of the research study includes the target population, method of sampling, measurement devices, data collection methods, statistical methods, research design and procedures, and the time schedule.

#### Target Population

The data source for this study was the scores obtained by all fifth grade students that were assessed in spring 2007 using the *PSSA*, the *DRA*, and the *4Sight* assessment in one urban Pennsylvania school district. There were nine elementary schools within the district that served students in fifth grade. Total student enrollment for the district was approximately 8,500 students, with approximately 575 students in fifth grade.

Of the 575 students, 5% (28) were Caucasian; 70% (401) were African-American; 22% (128) were Latino/Hispanic; and 3% (18) were Asian/Other (2007 *PSSA* Results, 2007). Approximately 94% of the students received free or reduced lunch.

Approximately 19% of the students were identified as eligible for specially designed instruction. Some students that received specially designed instruction were assessed using the Pennsylvania Alternate System of Assessment (*PASA*). These students were not included in the study because it was likely that they did not participated in the *DRA* or the *4Sight*. According to SchoolMatters.com (Standard & Poor's, 2006), 51.6% of the households in this district earned less than \$30,000 in 2005. The state average for households that earned less than \$30,000 in 2005 was 33.8%. Likewise, adult education levels were reported at levels below the state average. For adults with at least a high school diploma, the district percentage was 73.6% compared to the state average of

82.5%. For adults with at least a Bachelor's Degree, the district percentage was 16.9% compared to the state average of 23.8%.

#### Method of Sampling

The sample of the study was taken from the entire population of 575 fifth grade students in one central Pennsylvania school district who received scores on the *DRA*, the *4Sight*, and the reading portion of the *PSSA*. A random sample of 40 students was taken from the population of fifth grade students who received a score on all measures, the *DRA*, *4Sight*, and *PSSA*. Only those students who received scores aforementioned, regardless of individualized education plan (IEP), ethnicity, or socioeconomic status (SES), were considered for inclusion in the random sample. Students who participated in the PASA were excluded from this study.

The target population described 575 students that were considered for inclusion in this study. After excluding students that did not meet the criteria of having scores on all three assessments, the *PSSA*, the *DRA*, and the *4Sight*, the population was 378 students (N=378). From that population, 40 students were randomly selected for analysis using the SPSS random selection feature. Table 2 shows the sample frequency distribution by curriculum code, indicating whether students were in regular education or had an Individualized Education Plan (IEP). The sample was representative of the overall population, as described previously.

Table 2

*Student Sample Distribution by Curriculum Code*

Curriculum Code	Frequency	Percent	Valid Percent	Cumulative Percent
Regular Education	33	82.5	82.5	82.5
IEP	7	17.5	17.5	100.0
Total	40	100.0	100.0	

Table 3 shows the sample frequency distribution by ethnic code, and indicates an acceptable representation of the overall target population, as described previously.

Table 3

*Student Sample Distribution by Ethnic Code*

Ethnic Code	Frequency	Percent	Valid Percent	Cumulative Percent
Caucasian	2	5.0	5.0	5.0
Black	29	72.5	72.5	77.5
Hispanic	6	15.0	15.0	92.5
Asian	3	7.5	7.5	100.0
Total	40	100.0	100.0	

Table 4 shows the sample frequency distribution by meal status, indicated by free, reduced, or full price. Again, the sample was representative of the overall target population.

Table 4

*Student Sample Distribution by Meal Status*

Meal Status	Frequency	Percent	Valid Percent	Cumulative Percent
Free	34	85.0	85.0	85.0
Reduced	2	5.0	5.0	90.0
Full Price	4	10.0	10.0	100.0
Total	40	100.0	100.0	

Measurement Devices

DRA. The *Developmental Reading Assessment* is a classroom-based reading assessment that is designed to 1) monitor student growth relative to skills and strategies utilized by successful readers, 2) help teachers diagnose student needs and inform instruction, 3) prepare students to meet classroom and testing expectations, and 4) inform all stakeholders about the level of student achievement (Beaver & Carter, 2003b).

4Sight: “4Sight assessments are one-hour tests that have exactly the same formats, coverage, look, and feel as your state reading assessments. They produce overall scores that predict students’ scores on your state assessment” (4Sight reading and math benchmarks, 2006, p. 1). The 4Sight assessments are available in reading and in math for grades three through eight in many states, including Pennsylvania. They are predictive benchmark assessments that are intended to inform classroom instruction for the purpose of improving student achievement.

PSSA: The *Pennsylvania System of School Assessment*. The state assessment that is given in reading and math annually to all students in grades three through eight and eleven in the state of Pennsylvania. The Pennsylvania Department of Education describes

the *PSSA* as “a standards based criterion-referenced assessment used to measure a student’s attainment of the academic standards while also determining the degree to which school programs enable students to attain proficiency of the standards” (p. 1).

#### Data Collection Methods

Data were collected from school district reports using historical data that contained student *DRA* scores, *4Sight* scores, and *PSSA* reading scores from spring 2007. The data were collected using a computerized student file that contained only score information and ethnicity of subjects. The anonymity of individual students was protected at all times and students were not identified in any manner that would allow the researcher or anyone else to determine an individual student’s identity. The *DRA* score data were the independent reading level reported by teachers during the winter administration of the *DRA*. The *4Sight* score data were from the third administration. Both of these sets of data occurred prior to the administration of the 2007 *PSSA*. These two sets of data were chosen based on their proximity to one another and the *PSSA* administration. When all of the data were organized, the researcher analyzed and interpreted the data.

#### Statistical Methods

In this study, there were two independent variables, the *DRA* and the *4Sight*, that were used as predictor variables on one dependent variable, the *PSSA*. A multiple regression model was used to analyze the data set. “*Multiple regression* is a method of analyzing the collective and separate contributions of two or more independent variables,  $X_i$ , to the variation of a dependent variable,  $Y$ ” (Kerlinger & Pedhazur, 1973, p. 3). Multiple regression is described as a popular, flexible, and powerful tool for analyzing and disentangling the relative effects of two or more independent variables on a

dependent variable (Allen, 1997; Dillon & Goldstein, 1984; Foster, Barkus, & Yavorsky, 2006; Kerlinger & Pedhazur, 1973). According to Foster, Barkus, & Yavorsky (2006),

Multiple regression is used to answer three types of question:

1. What is the relative importance of the predictor variables included in the analysis?
2. A related question is to ask whether a particular variable adds to the accuracy of a prediction.
3. Given two alternative sets of predictors, which one is the more effective?

(p. 30)

This study was designed to answer all three of these types of question. Therefore, a multiple regression analysis was most appropriate.

For the purposes of analyzing the relationship between the *DRA* and the *PSSA* and the *4Sight* and the *PSSA*, a stepwise multiple regression procedure was used. A stepwise multiple regression procedure begins with all predictor variables and then continues by dropping out one variable at a time to determine the degree of prediction that is lost with each variable. “Due to the complexity of intercorrelations, the variance explained by certain variables will change when new variables enter the equation” (George & Mallery, 2005, p. 197). When the new variable enters the equation, “the Stepwise method will remove the ‘weakened’ variable” (George & Mallery, 2005, p. 197). The stepwise regression procedure described above is labeled the backward regression method in the SPSS statistical package used in this study.

Draper & Smith (1966) support the use of a stepwise regression model. They describe stepwise regression as “an improved version of the forward-selection procedure”

and note the “improvements involve the re-examination at every stage of the regression of the variables incorporated into the model in previous stages” (p. 171). Draper & Smith (1966) describe the steps of the stepwise model and are summarized to fit this study as follows:

Step 1. The stepwise procedure starts with the simple correlation matrix and enters into regression the  $X$  variable most highly correlated with the response.

Step 2. Using the partial correlation coefficients as before, it now selects, as the next variable to enter regression, that  $X$  variable whose partial correlation with the response is highest.

Step 3. Given the regression equation  $\hat{Y}=f(X_1, X_2)$ , the method now examines the contribution  $X_1$  would have made *if*  $X_2$  had been entered first and  $X_1$  entered second. (p.171)

At Step 3, a decision was made by the statistical software package to include or remove a variable based upon its partial F-value. Since there were only two independent variables, the method actually considered the contribution of both variables and in essence considered all possible regressions at the same time. Although stepwise is sometimes looked upon disapprovingly by authorities because of its inclusion of some variables rather than others (Foster, Barkus, & Yavorsky, 2006), stepwise was an appropriate method in this study because it only had two independent variables and considered both of them. SPSS, a computer statistical program, was used to analyze all data. A  $p \leq .05$  level of significance was used for all analyses in this study.

## Research Design and Procedures

The researcher used an ex post facto design to determine if there was a strong likelihood that the score from a formative assessment in reading predicted the score of a summative assessment in reading. In this study, the formative assessments to be considered were the *DRA* and the *4Sight* assessment. The summative assessment was the *PSSA*. The *DRA* and *4Sight* were considered predictor variables and the *PSSA* was considered the criterion variable or the variable to be predicted.

Since this study considered the relationship of formative assessments and a summative assessment, it was a relational study. The appropriate method is called ex post facto design “since causes are studied after they have presumably exerted their effect on another variable” (LaFountain & Bartos, 2002, p. 20). A regression model to explore the relationship of the *DRA*, *4Sight*, and the *PSSA* was considered. Gravetter & Wallnau (2004) define regression as the technique for determining the best-fitting straight line for a set of data. A straight line is helpful in describing the relationship between two variables. Since two variables, *DRA* and *4Sight*, were identified in this study, a step-wise multiple regression was used.

## CHAPTER FOUR

### RESULTS

The purpose of the study was to determine the relationship between the *DRA*, the *4Sight*, and the *PSSA*. A goal was to determine how well each assessment, the *DRA* and the *4Sight*, predicted student performance on the *PSSA*, and specifically to investigate whether the *DRA*, the *4Sight*, or both are significant predictors of student performance on the *PSSA*. The results of the data analysis are presented in this chapter.

Table 5 shows the descriptive statistics for the sample of the study. By itself, Table 5 does not answer any of the research questions, but it does provide insight into the level of performance associated with the sample.

Table 5

*Descriptive Statistics for Student Assessment Scores*

Assessment	Mean	Std. Deviation	N
PSSA Reading Score	1066.65	160.168	40
DRA Reading Score	44.10	12.169	40
4Sight Scaled Score	1130.60	125.406	40

To answer the questions relative to the relationship between the *PSSA*, *DRA*, and the *4Sight*, a correlation matrix is shown in Table 6. Correlations between assessments were significant ( $p \leq .01$ , one-tailed). This level of significance indicates a relationship between the assessments does exist.

Table 6

*Correlations between PSSA, DRA and 4Sight*

		PSSA Reading Score	DRA Reading Score	4Sight Scaled Score
Pearson Correlation	PSSA Reading Score	1.000	.543	.778
	DRA Reading Score	.543	1.000	.574
	4Sight Scaled Score	.778	.574	1.000
Sig. (1-tailed)	PSSA Reading Score	.	.000	.000
	DRA Reading Score	.000	.	.000
	4Sight Scaled Score	.000	.000	.
N	PSSA Reading Score	40	40	40
	DRA Reading Score	40	40	40
	4Sight Scaled Score	40	40	40

Since Table 6 confirms the relationship between assessments, the variables were entered into a stepwise multiple regression, as proposed in the previous chapter. Table 7 shows these results. The criteria for a variable to enter the model was  $F(2, 37) \leq .05$ , and the probability to remove was  $F(2, 37) \geq .10$ . Table 7 shows that only one variable, *4Sight*, entered the model using a stepwise approach. The resulting value was  $R=.778$ , with  $R^2=.605$ . *4Sight* accounted for 60.5% of the variance. Table 7 also shows that the confidence interval excludes zero, therefore indicating the significance of the predictor variable, *4Sight*, on the dependent variable, *PSSA*. *DRA* was not significant,  $t(1.162)$ ,  $p = .253$ , and therefore was excluded from the model.

Table 7

*Stepwise Regression Model Summary for Variables Predicting PSSA Scores*

Variable	<i>B</i>	<i>SE B</i>	$\beta$	95% Confidence Interval for <i>B</i>	
				Lower Bound	Upper Bound
4Sight	.993	.130	.778	.729	1.257
Constant	-56.239	-	-	-	-

Since *DRA* was excluded from the stepwise method, a backward method was employed to determine the additional variance, if any, was accounted for by *DRA*. By using a backward method, both variables were stepped into the equation. Table 8 shows the resulting values that include the variance accounted for using both variables. The backward method also excluded *DRA* at the second step, leaving only *4Sight*, which produced the same values as the stepwise method. As Table 8 shows, the *DRA* only accounts for an additional 1.4% of the variance in scores.

Table 8

*Backward Regression Model Summary for Variables Predicting PSSA Scores*

Variable	<i>R</i>	<i>R</i> <sup>2</sup>
DRA & 4Sight	.787	.619
4Sight	.778	.605

Both stepwise and backward methods completed the regression analysis and arrived at the same conclusion. That is, *4Sight* is the only predictor variable that reached

the threshold criterion to remain in the equation. To complete the presentation of results, a scatter plot is shown in Figure 1.

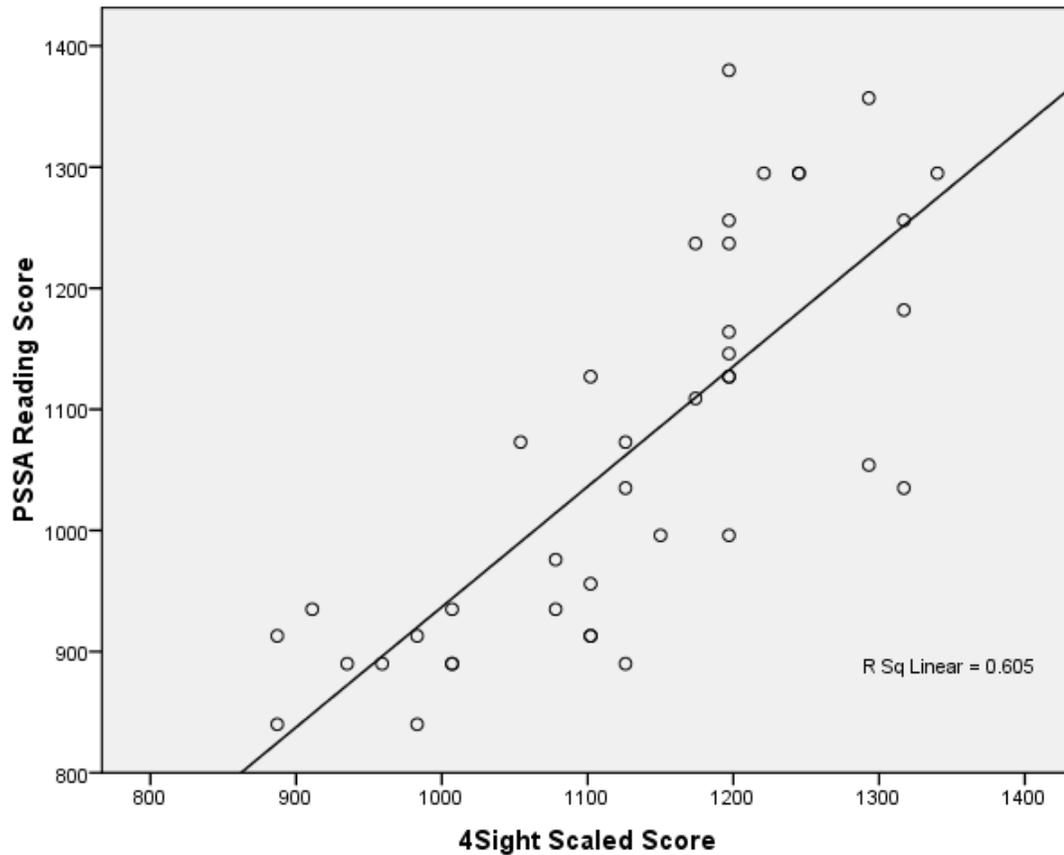


Figure 1. 4Sight Scores vs. PSSA Scores

Figure 1 confirms visually that a linear relationship between *4Sight* and *PSSA* does exist. The regression equation that describes this linear relationship is  $y = (.993)x + (-56.239)$ .

The goal of this study was to analyze the relationships between *DRA* and *4Sight* scores with *PSSA* scores. Chapter Four presented the results of the regression analysis and the resulting regression equation. From these results, the researcher can make determinations about the relationship of each assessment to *PSSA*, the strength of each assessment as a predictor variable for scores on the *PSSA*, which predictor variable is

better, if any, and finally, whether there is added benefit by using both sets of assessment scores to predict student achievement on the *PSSA*. The discussion of these results is found Chapter Five.

## CHAPTER FIVE

### DISCUSSION

The purpose of the study was to determine the relationship between the *DRA*, the *4Sight*, and the *PSSA*. A goal was to determine how well each assessment, the *DRA* and the *4Sight*, predicted student performance on the *PSSA*, and specifically to investigate whether the *DRA*, the *4Sight*, or both are significant predictors of student performance on the *PSSA*. An additional goal was to determine whether there was any added benefit to the prediction by using both assessments.

The first research question sought to determine the relationship between student performance on the *DRA* and the *PSSA* in grade five. Table 4 presented the mean score for each assessment. Mean scores for the *DRA* and *PSSA*, (44.10 and 1066.65, respectively) were below proficient levels for fifth grade students. According to the PDE website, the minimum score to be considered proficient on the *PSSA* was 1275 (Math, reading and writing performance level cut scores, 2007). Students are considered on grade level with a *DRA* score of 50 in fifth grade (Beaver & Carter, 2003b). One indication from this inspection of means is that both were predicting below proficient performance consistently. Table 5 presented the correlations between assessments and showed a Pearson  $r = .543$  to describe the correlation between *DRA* and *PSSA*. This relationship was significant at the  $p \leq .01$  level.

The second research question sought to determine the relationship between student performance on the *4Sight* and the *PSSA* in grade five. Table 4 showed the mean score for *4Sight* as  $M = 1130.60$ . The *4Sight* minimum scaled score to be considered proficient was 1293 (Success for All Foundation, 2007b). Again, *4Sight* was consistent in predicting below proficient performance. Table 5 showed a Pearson  $r = .778$  to describe

the correlation between *4Sight* and *PSSA*. This relationship was significant at the  $p \leq .01$  level.

After initial inspection of the means for each assessment and the correlation matrix, it was evident there was a stronger correlation between *4Sight* and *PSSA* than between *DRA* and *PSSA*. The strength of these correlations provides the first real insight into the relationships between each assessment and the *PSSA*. Nonetheless, a significant correlation existed between all assessments. Therefore, the proposed method of stepwise regression was appropriate.

The remaining three research questions sought to determine whether or not each assessment significantly predicted scores on the *PSSA*, and if so, which was the better predictor. Table 6 showed the resulting model summary that essentially answered all three of those questions. *4Sight* was the only variable to enter the model, and accounted for 60.5% of the variance in scores. *DRA* failed to enter the equation and was not considered to be a significant predictor of *PSSA* scores. Therefore, *4Sight* is a significant predictor of *PSSA* scores, and compared with *DRA* scores in this study, it is a better predictor.

To explore the nature of what would be added if both *DRA* and *4Sight* were considered together in the regression analysis, a backward regression method was completed and shown in Table 8. In the first step, both variables were considered, resulting in 61.9% of the variance. The difference then by considering *DRA* and *4Sight* is only 1.4%. This confirms why *DRA* failed to make the equation, considering that when all else is controlled, it only added 1.4% change in variance. Regardless of the method employed, *DRA* failed to make the final equation.

## Defining the Use of *4Sight* and *DRA* as Local Assessments

The final part of this discussion will focus attention beyond the initial research questions to consider some of the other issues described in the study and implications for future research. To begin, there is more to be considered beyond the nature of the last research question about which is the better predictor of *PSSA* scores. It has been demonstrated in the results of this study that *4Sight* is the better predictor of *PSSA* scores. A larger issue described in this study is the types of local assessment that are used to predict performance on the state assessment and the philosophy and belief system inherent with each. In particular, the *DRA* was described as a more authentic performance assessment, in that it assesses tasks, behaviors, and strategies associated with authentic reading (Beaver, 2003a) in a personalized manner. The *4Sight*, on the other hand, promotes itself as an assessment that is designed to mirror the content, look, and feel of the *PSSA* (*4Sight* Reading and Math Benchmarks, 2006), suitable for large group administration. If the intent is to administer an assessment that will reasonably predict scores on the state assessment, then *4Sight* is the better choice. If the intent is to administer an assessment that will inform teachers about the strengths of individual students reading in an authentic manner, further study is warranted. The results from this study do not allow the researcher to determine that *4Sight* would be any less effective in accomplishing that goal than the *DRA*. This issue is divided by what may be known about what good readers do versus what is measured on the state assessment, and how educators view teaching reading, that is, teaching reading or teaching to the test.

The answer to this issue may lie within the individual reader's belief system and philosophy about how children learn to read. As a practical example of this dilemma, students may be assessed using both assessments and receive the same matched scores.

What may have contributed to those results on both assessments could vary greatly for each student. For Student A the *DRA* may have identified that making inferences and summarizing are key strategies that he/she needs to work on. For Student B the *DRA* may have identified the same. The *4Sight* can only tell you that students missed the question, not why they missed the question. Assuming both Student A and Student B missed all of the same questions, they may have missed them for different reasons, including chance if they guessed at the answer or made an error in selecting their answer. This distinction makes planning for future instruction difficult because it is vague in nature. Herein lies a point for consideration in determining what the intent for the use of either assessment may be.

#### How Much Assessment Is Too Much

A related issue that was described in this study is over assessment of students. The determination of added value by using both assessments was critical in considering this issue. The contribution of *DRA* was considered using a backward model in Table 8. In other words, if there is no added benefit to using two reading assessments, why continue that practice? Part of that issue is answered in the previous paragraphs. There may be curricular and instructional value in continuing with both assessments. That decision must be made at the local level, assuming there is no mandate for specific local assessments. Local officials must make their own decisions about the costs of using two assessments, both in a financial sense and as a measure of instructional time. Further study to evaluate this issue may be considered.

#### Possible Limitations of the *DRA* Score

A final speculation about the results of this study considers the *DRA* score itself. The *DRA* is teacher scored and more subjective in nature. As described in the student

example above, students receive a score that describes the level of reader they are. They are fewer gradations in the *DRA* than in *4Sight* and *PSSA*. For example, two students may have the same score of 50, but one may in essence be higher than the other. There is no distinction between a high 50 and a low 50. Considerations for how to further distinguish individual performance on the *DRA* may impact results of further studies. Also, the ability to move from a *DRA* score to a scaled score may be helpful in further studies.

#### Limitations of Study

As with any study, there are limitations that could potentially impact the results of this study. These limitations are listed below:

1. Winter administration of the *DRA* occurred approximately eight weeks before the *PSSA* administration. This limitation does not allow for additional learning that may have occurred after the administration of the *DRA*.
2. The third administration of the *4Sight* occurred approximately seven weeks before the *PSSA* administration. This limitation does not allow for additional learning that may have occurred after the administration of the *4Sight*.
3. The *4Sight* is administered in an online format, where skills not being assessed may affect scores, e.g. scrolling on a computer screen versus having text in hard copy format. Providing students with a hard copy of the text may be an acceptable accommodation, but there still may be an issue with transfer of answers and information from one format to the other.
4. *DRA* scores may be affected by varying teacher experience and professional development in administering the assessment. Since years of teaching experience is not being measured in this study, it was assumed that some teachers were

novices and less familiar with the administration of the *DRA* than more veteran teachers that had been conducting the assessment for several years. All teachers, whether novice or veteran, had been provided with professional development in the administration of the *DRA*.

### Conclusions

If school districts were simply comparing the *DRA* and the *4Sight* to determine which assessment is the better predictor of performance on the *PSSA*, then similar results would likely follow. School districts must strive to maximize assessment that not only predicts performance on the high-stakes state assessment but that informs daily instruction and curricular initiatives. Assessments should not be redundant and use valuable instructional time for administration and scoring. In this study, both assessments vary in form and their use may be dictated based on the philosophy of reading that a school district holds or the desire to make AYP and accountability standards. In order to support the philosophical reasons for the continued use of the *DRA* as an assessment that may inform daily instruction and be a predictor, further study is necessary.

### Implications for Further Study

Often times, questions lead to more questions instead of answers. As a result of this study, several questions for future study emerged. These questions generally extend from the results of this study and they are listed below.

1. Do the *DRA* and *4Sight* predict differentially between student groups? A future study might consider how each assessment predicts scores on the *PSSA* for each performance level, i.e. Advanced, Proficient, Basic, and Below Basic. It is possible that one of the assessments would be a better predictor for a specific performance level group(s) but not others.

2. Does teacher training and professional development significantly improve the predictive power of the *DRA*? Since the *DRA* is subjective and relies heavily on teacher scoring, it would seem that teacher training would have an effect on their ability to score and administer the *DRA*. In turn, a strong correlation between teacher training and *DRA* scores may increase the predictive power of the *DRA*.
3. Do increased gradations in the scoring of the *DRA* increase its predictive power of the *PSSA*? As described in the section on the limitations of the *DRA*, a *DRA* score of 50 may represent a wide range of student abilities and scores on the *PSSA*. Combined with the subjective nature of the assessment, additional gradations may more accurately depict student performance and therefore have a stronger predictive power for scores on the *PSSA*.

## REFERENCES

- 2007 PSSA results.* (2007). The Pennsylvania Department of Education. Available from [http://www.pde.state.pa.us/a\\_and\\_t/lib/a\\_and\\_t/2007\\_District\\_Level\\_PSSA\\_Results.xls](http://www.pde.state.pa.us/a_and_t/lib/a_and_t/2007_District_Level_PSSA_Results.xls). Accessed on March 25, 2008.
- 4Sight reading and math benchmarks.* (2006). The Center for Data-Driven Reform in Education. Available from <http://www.cddre.org>.
- Allen, M. P. (1997). *Understanding Regression Analysis*. New York: Plenum Press.
- Answer.* (2008). United States Department of Education. Available from [http://answers.ed.gov/cgi-bin/education.cfg/php/enduser/std\\_adp.php?p\\_faqid=6&p\\_created=1095256734&p\\_sid=biCz1Gli&p\\_lva=&p\\_sp=cF9zcmNoPSZwX3NvcnRfYnk9JnBfZ3JpZHNvcnQ9JnBfcm93X2NudD0xMTUmcF9wcm9kc0mcF9jYXRzPSZwX3B2PSZwX2N2PSZwX3BhZ2U9MQ\\*\\*&p\\_li=&p\\_topview=1](http://answers.ed.gov/cgi-bin/education.cfg/php/enduser/std_adp.php?p_faqid=6&p_created=1095256734&p_sid=biCz1Gli&p_lva=&p_sp=cF9zcmNoPSZwX3NvcnRfYnk9JnBfZ3JpZHNvcnQ9JnBfcm93X2NudD0xMTUmcF9wcm9kc0mcF9jYXRzPSZwX3B2PSZwX2N2PSZwX3BhZ2U9MQ**&p_li=&p_topview=1). Accessed on March 25, 2008.
- Arenson, E. A. (1999). *Statistical linkages between state education assessments and the national assessment of educational progress.* (ERIC Document Reproduction Service No. ED432609)
- Assessment.* (2006). The Pennsylvania Department of Education. Available from [http://www.pde.state.pa.us/a\\_and\\_t/site/default.asp?g=0&a\\_and\\_tNav=|630|&k12Nav=|1141|](http://www.pde.state.pa.us/a_and_t/site/default.asp?g=0&a_and_tNav=|630|&k12Nav=|1141|). Accessed on October 16, 2006.
- Beaver, J. & Carter, M. A. (2003a). *Developmental Reading Assessment*. Upper Saddle River, NJ: Pearson Education, Inc.
- Beaver, J., & Carter, M. A. (2003b). *Developmental Reading Assessment Teacher Guide*. Upper Saddle River, NJ: Pearson Education, Inc.

- Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.
- Boston, C. (2002). The concept of formative assessment. *Practical Assessment, Research & Evaluation*, 8(9). Retrieved December 5, 2006 from <http://PAREonline.net/getvn.asp?v=8&n=9>
- Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of Assessment in Education. *Assessment in Education*, 11(1), 7-27.
- Chapman, C., & King, R. (2005). *Differentiated Assessment Strategies*. Thousand Oaks, CA: Corwin Press, Inc.
- Cohen, J., Cohen, P., West, S.G., Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3<sup>rd</sup> ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Dillon, W. R., & Goldstein, M. (1984). *Multivariate Analysis: Methods and applications*. New York: John Wiley & Sons, Inc.
- DRA correlation to Pennsylvania state standards*. (n.d.). Retrieved July 2, 2006, from [http://www.k12pearson.com/statepage/correlation/plg\\_corr/pa/PA\\_DRA\\_grade3.pdf](http://www.k12pearson.com/statepage/correlation/plg_corr/pa/PA_DRA_grade3.pdf)
- DRA technical manual*. (n.d.) Retrieved July 2, 2006, from [http://www.pearsonlearning.com/correlation/rsp/ResearchPaper\\_DRA\\_TechManual.pdf](http://www.pearsonlearning.com/correlation/rsp/ResearchPaper_DRA_TechManual.pdf)
- Draper, N. R., & Smith, H. (1966). *Applied Regression Analysis*. New York: John Wiley & Sons, Inc.

- Foster, J., Barkus, E., & Yavorsky, C. (2006). *Understanding and Using Advanced Statistics*. Thousand Oaks, CA: SAGE Publications, Inc.
- Four Pillars of NCLB*. (2008). United States Department of Education. Available from <http://www.ed.gov/nclb/overview/intro/4pillars.html>. Accessed on March 25, 2008.
- Funding Status- Striving Readers Program*. (2008). United States Department of Education. Available from <http://www.ed.gov/programs/strivingreaders/funding.html>. Accessed on March 18, 2008.
- George, D., & Mallery, P. (2005). *SPSS for Windows Step by Step* (5<sup>th</sup> ed.). New York: Pearson Education, Inc.
- Girton, K. R., & Buckheit, J. (2006). Pennsylvania System of School Assessment: Myths and realities. *The Pennsylvania Administrator*, 10(3), 7-9.
- Goertz, M., & Duffy, M. (2003). Mapping the landscape of high-stakes testing and accountability programs. *Theory Into Practice*, 42(1), 4-11.
- Goertz, M. E., & Duffy, M. C. (2001). *Assessment and accountability systems in the 50 states: 1999-2000* (CPRE Research Report Series RR-046). University of Pennsylvania: Consortium for Policy Research in Education.
- Gravetter, F. J., & Wallnau, L. B. (2004). *Statistics for the behavioral sciences* (6<sup>th</sup> ed.). Belmont, CA: Thomson Learning, Inc.
- Greene, J. P., & Winters, M. A. (2004). An evaluation of Florida's program to end social promotion. *Education Working Paper*, 7, 1-24.

- Guskey, T. R. (2005). *Multiple sources of evidence: An analysis of stakeholders' perceptions of different indicators of student learning*. Paper presented at the Annual Meeting of the American Educational Research Association: Montreal, Canada.
- Helping readers achieve and succeed*. (2006). United States Department of Education. Available from <http://www.ed.gov>. Accessed on October 20, 2006.
- Individuals with Disabilities Act, 20 U.S.C. 1400 et seq. (2004).
- Invernizzi, M. A., Landrum, T. J., Howell, J. L., & Warley, H. P. (2005). Toward the peaceful coexistence of test developers, policymakers, and teachers in an era of accountability. *The Reading Teacher*, 58(7), 610-618.
- 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.
- Kerlinger, F. N., & Pedhazur, E. J. (1973). *Multiple Regression in Behavioral Research*. New York: Holt, Rinehart, & Winston, Inc.
- LaFountain, R. M., & Bartos, R. B. (2002). *Research and statistics made meaningful in counseling and student affairs*. United States: Thomson Learning, Inc.
- Lewis, A. (2001). Benchmarks for accountability: Are we there yet? *Proceedings of the 1999 CRESST Conference* (CSE Technical Report 547). Los Angeles, CA.
- Local assessment system*. (2006). The Pennsylvania Code, 22 Pa Code § 4.52. Available from <http://www.pacode.com>.
- Louisiana Department of Education. (1998). *Progress report on second and third graders' reading abilities, 1998-1999, Executive Summary*. (ERIC Document Reproduction Service No. ED460353)

- Madaus, G. F., & O'Dwyer, L. M. (1999). A short history of performance assessment. *Phi Delta Kappan*, 80(9), 688-695.
- Math, reading and writing performance level cut scores.* (2007). The Pennsylvania Department of Education. Available from [http://www.pde.state.pa.us/a\\_and\\_t/lib/a\\_and\\_t/Cut\\_Scores\\_07.pdf](http://www.pde.state.pa.us/a_and_t/lib/a_and_t/Cut_Scores_07.pdf). Accessed on August 16, 2008.
- Morris, T. M. (2002). A study of reading and the reading assessment program at South Elementary School. *ProQuest Information and Learning Company*, (UMI No. 3043072)
- Mueller, J. (2006). *Authentic Assessment Toolbox*. Available from <http://jonathan.mueller.faculty.noctrl.edu/toolbox/tasks.htm>. Accessed on March 25, 2008.
- NCLB executive summary.* (2002). United States Department of Education. Available from <http://www.ed.gov>. Accessed on October 20, 2006.
- Nitko, A. J. (2004). *Educational assessment of students* (4<sup>th</sup> ed.). Upper Saddle River, NJ: Pearson Education, Inc.
- No Child Left Behind Act, PL 107-110, 115 Stat. 1425 (2002).
- PA Alternate System of Assessment (PASA).* (2008). Pennsylvania Training and Technical Assistance Network. Available from: [http://www.pattan.net/teachlead/PAAAlternate SystemofAssessment\(PASA\).aspx](http://www.pattan.net/teachlead/PAAAlternateSystemofAssessment(PASA).aspx). Accessed on August 24, 2008.
- Paris, S. G. (2002). Measuring children's reading development using leveled texts. *The Reading Teacher*, 56(2), 168-170.

- Paris, S. G., & Carpenter, R. D. (2003). FAQs about IRIs. *The Reading Teacher*, 56(6), 578-580.
- Paris, S. G., & Hoffman, J. V. (2004). Reading assessments in kindergarten through third grade: Findings from the Center for the Improvement of Early Reading Achievement. *The Elementary School Journal*, 105(2), 199-217.
- Reading First*. (2007). United States Department of Education. Available from <http://www.ed.gov/programs/readingfirst/index.html>. Accessed on November 20, 2007.
- Reeves, D. B. (2002). *Holistic accountability: Serving students, schools, and communities*. Thousand Oaks, CA: Corwin Press, Inc.
- Shaywitz, S. E., & Shaywitz, B. A. (2006). *Dyslexia at an early age and its impact on early socio-emotional development*. Retrieved November 20, 2007 from <http://www.child-encyclopedia.com/pages/PDF/ShaywitzANGxp.pdf>.
- State assessment system*. (2006). The Pennsylvania Code, 22 Pa Code § 4.51. Available from <http://www.pacode.com>.
- Standard & Poor's. (2006). *SchoolMatters*. Available from <http://schoolmatters.com>.
- Success for All Foundation. (2007a). *4Sight reading and math benchmarks 2007-2008 technical report for Pennsylvania*. Available from <https://members.successforall.net>.
- Success for All Foundation. (2007b). *Pennsylvania benchmark administration and scoring guide, 4<sup>th</sup> ed*. Available from <https://members.successforall.net>.

- Thomas, J. A. (2002). Predictor variables of performance on the Louisiana Educational Assessment Program for the 21<sup>st</sup> Century. *ProQuest Information and Learning Company*, (UMI No. 3066658)
- Turner, R. C., & Williams, T. L. (2002). Predicting student proficiency on the Arkansas Benchmark Examinations. *Arkansas Educational Research & Policy Studies Journal*. 2(1), 48-69.
- Washtenaw Intermediate School District. (2008). *Developmental Reading Assessment*. Available from <http://www.wash.k12.mi.us/files/Maisa/Developmental%20Reading.pdf>. Accessed on March 25, 2008.
- William, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment*, 11(3&4), 283-289.
- Williams, B. (Ed.). (1996). *Closing the Achievement Gap*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Williams, E. J. (1999). *Developmental Reading Assessment* reliability study. Retrieved June 22, 2006, from <http://dev.pearsonlearning.com/correlation/rsp/DRA.doc>
- Williams, E. J. (2003). Preliminary validation of the *Developmental Reading Assessment*. Retrieved July 2, 2006, from [http://www.pearsonlearning.com/dra/dra\\_validation.html](http://www.pearsonlearning.com/dra/dra_validation.html)
- Wraga, W. (1999). The educational and political implications of curriculum alignment and standards-based reform. *Journal of Curriculum and Supervision*, 15(1), 4-25.

Zwerling, H. L. (2003). *The performance levels and associated cut scores on the Pennsylvania System of School Assessment mathematics and reading tests: A critical analysis.* (ERIC Document Reproduction Service No. ED480877)

## Appendix A

### 22 Pa Code § 4.51 State assessment system

(a) The State assessment system shall be designed to serve the following purposes:

(1) Provide students, parents, educators and citizens with an understanding of student and school performance.

(2) Determine the degree to which school programs enable students to attain proficiency of academic standards under § 4.12 (relating to academic standards).

(3) Provide results to school districts, including charter schools, and AVTSs for consideration in the development of strategic plans under § 4.13 (relating to strategic plans).

(4) Provide information to State policymakers including the General Assembly and the Board on how effective schools are in promoting and demonstrating student proficiency of academic standards.

(5) Provide information to the general public on school performance.

(6) Provide results to school districts, including charter schools, and AVTSs based upon the aggregate performance of all students, for students with an Individualized Education Program (IEP) and for those without an IEP.

(b) All State assessment instruments will be standards-based and criterion referenced and include essay or open-ended response items in addition to other item formats. The proportion of type of items will vary by grade level. Neither State assessments nor academic standards under § 4.12 shall require students to hold or express particular attitudes, values or beliefs. The Department will make samples of assessment questions, instrument formats, and scoring guides available to the public after each administration of State assessments. The criteria for judging performance on State assessments are as follows:

(1) Performance on State reading assessments shall be demonstrated by students' responses to comprehension questions about age-appropriate reading passages and by their written responses to in-depth comprehension questions about the passages.

(2) Performance on State mathematics assessments shall be demonstrated by students' responses to questions about grade-appropriate content and by the quality of their responses to questions which require a written solution to a problem.

(3) Performance on State writing assessments shall be demonstrated by the quality of students' written compositions on a variety of topics and modes of writing.

(4) Levels of proficiency shall be advanced, proficient, basic and below basic. In consultation with educators, students, parents and citizens, the Department will develop and recommend to the Board for its approval specific criteria for advanced, proficient, basic and below basic levels of performance.

(c) The Department will develop or cause to be developed State assessments based on academic standards in mathematics, reading and writing under § 4.12 and contained in Appendix A. In developing assessments, the Department will consult with educators, students, parents and citizens regarding the specific methods of assessment. To ensure that information regarding student performance is available to parents and teachers, State assessments developed under this section shall include student names. Individual test results shall be used in planning instruction only by parents, teachers, administrators and guidance counselors with a need to know based upon local board policy on testing and in reporting academic progress. The Department or other Commonwealth entities are prohibited from collecting individual student test scores, and may only collect aggregate test scores by school and district.

(d) The State assessments shall be administered annually and shall include assessments of the State academic standards in mathematics and reading at grades 5, 8 and 11 and in writing at grades 6, 9 and 11. The purpose of State assessments administered in 1999 is to validate assessment instruments and to provide initial information to teachers and schools to guide the redesign of curricula and instructional strategies to enable students to achieve academic standards.

(e) Students not achieving at the proficient level in the administration of State assessments in grade 11 shall be provided one additional opportunity in grade 12 to demonstrate a proficient level on State assessments.

(f) Expansion of the State assessment system will be authorized by the Board through a revision of this chapter.

(g) The Department will implement provisions for security of the State assessment system, including the following provisions:

(1) Action by a professional employe or commissioned officer which is willfully designed to divulge test questions, falsify student scores or in some other fashion compromise the integrity of the State assessment system as determined by the school district shall be subject to disciplinary action under sections 1259—1267 of the School Code (24 P. S. § § 12-1259—12-1267).

(2) Cheating by students or employes other than those covered in paragraph (1) shall be subject to disciplinary action by the school district.

(3) Cheating or breaches of assessment security shall be reported to the Secretary as soon as detected.

(h) The Secretary has the authority to establish guidelines for the administration of the State assessment system.

(i) The Secretary will report each September to the Board and the General Assembly information and pertinent data relating to the State assessment system. The Secretary will also provide each school district (including charter schools) and AVTS information and pertinent data for the school district or AVTS and its students.

(j) Children with disabilities shall be included in the State assessment system, with appropriate accommodations, where necessary. As appropriate, the Commonwealth will develop guidelines for the participation of children with disabilities in alternate assessments for those children who cannot participate in the State assessment as determined by each child's Individualized Education Program team under the Individuals with Disabilities Education Act and this part.

### **Cross References**

This section cited in 22 Pa. Code § 4.21 (relating to elementary education: primary intermediate levels); 22 Pa. Code § 4.24 (relating to high school graduating requirements); 22 Pa. Code § 4.61 (relating to school profiles); and 22 Pa. Code § 14.142 (relating to caseload for special education).

## Appendix B

### 22 Pa Code § 4.52 Local assessment system

(a) Each school district, including charter schools, and AVTS shall design an assessment system to do the following:

(1) Determine the degree to which students are achieving academic standards under §§ 4.12 and 4.13(c)(3) (relating to academic standards; and strategic plans). The school district (including charter schools) or AVTS shall provide assistance to students not attaining academic standards at the proficient level or better and the assistance to be provided shall be indicated in the strategic plan under § 4.13.

(2) Use assessment results to improve curriculum and instructional practices, to guide instructional strategies and to develop future strategic plans under § 4.13.

(3) Provide information requested by the Department regarding the achievement of academic standards, which does not include student names, identification numbers or individually identifiable information.

(4) Provide summary information including results of assessments under this section to the general public regarding the achievement of students, which does not include student names, identification numbers or individually identifiable information.

(b) The local assessment system shall be implemented no later than 1 year after its strategic plan or revision is approved by the board of school directors under § 4.13.

(c) The local assessment system shall be described in the district's (including charter schools) or AVTS's strategic plan under § 4.13(b)(5).

(d) The local assessment system shall be designed to include a variety of assessment strategies which may include the following:

(1) Written work by students.

(2) Scientific experiments conducted by students.

(3) Works of art or musical, theatrical or dance performances by students.

(4) Other demonstrations, performances, products or projects by students related to specific academic standards.

(5) Examinations developed by teachers to assess specific academic standards.

(6) Nationally-available achievement tests.

- (7) Diagnostic assessments.
  - (8) Evaluations of portfolios of student work related to achievement of academic standards.
  - (9) Other measures as appropriate, which may include standardized tests.
- (f) Individual test information shall be maintained in a student’s educational record in a manner consistent with section 438 of the Family Educational Rights and Privacy Act of 1974 (20 U.S.C.A § 1232g) and 34 CFR Part 99 (relating to family educational rights and privacy).
- (g) Children with disabilities shall be included in the local assessment system, with appropriate accommodations, when necessary. As appropriate, the school district, including charter schools, or AVTS shall develop guidelines for the participation of children with disabilities in alternate assessments for those children who cannot participate in the local assessment as determined by each child’s Individualized Education Program team under the Individuals with Disabilities Education Act and this part.

*(Editor’s Note: The State Board of Education published a Notice of PSSA general performance level descriptors and performance level scores at 31 Pa.B. 2763 (May 26, 2001).)*

### **Cross References**

This section cited in 22 Pa. Code § 4.12 (relating to academic standards); 22 Pa. Code § 4.13 (relating to strategic plans); 22 Pa. Code § 4.21 (relating to elementary education: primary and intermediate levels); 22 Pa. Code § 4.24 (relating to high school graduation requirements); 22 Pa. Code § 4.61 (relating to school profiles); and 22 Pa. Code § 14.142 (relating to caseload for special education).