

Using Formative Assessment Results to Predict Student Achievement on High Stakes
Tests

A Dissertation

Presented to

The Faculty of the School of Education

Liberty University

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Education

By

Lisa W. Smith

April, 2008

Using Formative Assessment Results to Predict Student Achievement on High Stakes
Tests
By Lisa W. Smith

APPROVED:

COMMITTEE CHAIR

Karen L. Parker, Ed.D.

COMMITTEE MEMBERS

Lisa A. Sheehy, Ph.D.

Branson L. Woodard, D.A.

CHAIR, GRADUATE STUDIES

Scott B. Watson, Ph.D.

Abstract

Lisa W. Smith. USING FORMATIVE ASSESSMENT RESULTS TO PREDICT STUDENT ACHIEVEMENT ON HIGH STAKES TESTS. (Under the direction of Dr. Karen Parker) School of Education, Liberty University, March, 2008.

The purpose of this study was to examine whether comprehensive post formative assessments can accurately predict student academic achievement on AYP (Adequate Yearly Progress) indicators as measured by standardized criterion-referenced tests.

The primary participant populations for this study were sixth, seventh, and eighth grade students enrolled in a middle school in north Georgia from 2004-2007. Over 2,900 student assessments were used to conduct the statistical research and variables such as gender, race, and socio-economic levels were not disaggregated in the data collection compilation. The data sources included the first quarter, second quarter, and third quarter post formative assessments which are administered every nine-week grading period in the school system.

The findings indicated that various grade levels exhibited a higher predictability factor with certain quarterly assessments than others. Likewise, unit gains on post assessments demonstrated a statistically significant indicator for academic achievement on high stake standardized assessments.

Dedication

I would like to dedicate this dissertation to my Lord and Savior Jesus Christ, whom without, all is meaningless and trivial. It is His mercy and grace that sustains me, refreshes my soul, and guides me towards glorifying the Father. I can do all things through Christ who strengthens me!

Acknowledgements

I want to thank my family and many friends who have lent their support and guidance through years of post-secondary education that culminates in this dissertation. Moreover, without their *belief in me*, I would surely have given up hope and squandered the gifts that the Lord has laid upon my table.

A gracious thank you is also extended to the faculty and staff of Liberty University and the Gainesville City School System in Gainesville, Georgia, members of my doctoral committee, and my most steadfast supporter—Tinky. He washed away my tears, lifted my spirits when I was down, and above all else, never left my side while professing unconditional love in a simple glance.

Table of Contents

Committee Members.....	ii
Abstract.....	iii
Dedication.....	iv
Acknowledgements.....	v
Chapter One	
Introduction	1
Background of the Study.....	1
Problem Statement.....	4
Professional Significance of the Study.....	6
Definition of Key Terms.....	8
Chapter Two	
Review of Literature.....	9
Characteristics of School Reform.....	9
The Need to Develop Best Practices.....	9
Examples of Reform Models	12
Key Factors to Influence Positive Change	16
Standards-Based Curricula - A Brief History	19
National Standards	20
State Standards	30
The Link to Performance and Formative Assessments	33
Formative Assessments and Current Research and Elements of Implementation.....	36

Formative Assessment in the Secondary School Context	39
Feedback to Students	39
Assessment Informing Teaching	40
Student Involvement in Assessment	42
Learning Goals and Assessment Criteria	45
Varied Approach to Elicit Learning	46
Chapter Three	
Basic Research Design	49
Preliminary Procedures	51
Selection of the Sample.....	51
Instruments Used in Data Collection - CRCT.....	53
Formative Assessment	55
Procedures.....	58
Analysis of Data	59
Summary	60
Chapter Four	
Research Questions	61
Null Hypotheses	62
Data Analysis	64
Results.....	66
Chapter Five	
Conclusions	75

Summary of Results	75
Implications	77
Delimitations/Limitations of the Study	77
Recommendations for Future Practice	77
Recommendations for Future Research.....	78
References.....	80

Appendices

APPENDIX A

Sample--Quarter 1 Formative Assessment—GMS 6 th Grade.....	90
---	----

Chapter 1

Introduction

Public education in the United States has become an assessment-driven system that focuses on accountability towards both school and teacher effectiveness along with student academic achievement (Marzano, 2005). This working proposal describes a proposed research project that will examine whether comprehensive post formative assessments can accurately predict student academic achievement on AYP (Adequate Yearly Progress) indicators as measured by standardized criterion-referenced tests. This first chapter of the proposal discusses the study.

Background of the Study

The *Elementary and Secondary Education Act of 1965* has evolved in the last 42 years into the modern-day *No Child Left Behind Act* that was recently reauthorized in 2007. Since the passing of the *No Child Left Behind Act of 2001*, volumes have been written about school reform. Efforts to improve K-12 education have increasingly recognized that improvement requires reform of a host of interconnected systems that make up education as we know it. "School reform" still means many things to many people, but there is widespread agreement that measuring student learning against standards and seeking to improve the effectiveness of schools, based on such measures, are key goals of the No Child Left Behind legislation.

NCLB has brought to the forefront accountability for student achievement. In essence, the Federal government has now required school districts or Local Education Agencies to demonstrate Adequate Yearly Progress towards student achievement goals.

LEAs that do not meet the required performance levels run the risk of losing Federal funds, face a mandated reorganization of the system, or forfeit their accreditation. One of the major additions to the Act from its original version stipulates that state departments of education must utilize a criterion-referenced formalized assessment to measure student achievement. Although the Federal government allows states to create their own assessments, the Federal government has also imposed Annual Measurable Objectives that increase annually (US Dept of Ed, 2007). For states, their standardized assessments have become the “barometer” used to measure and often times declare that a particular school or system is high achieving, low achieving, or simply adequate.

At the heart of student achievement rests the relationship between teacher and student. Often times, very early in a school year, teachers attempt to evaluate their individual students’ needs. A determination is made concerning the student’s level of mastery in a given content area or in relation to a style, for example, tactile kinesthetic, and then instruction is centered on advancing a student in that area along an achievement continuum. Moreover, curriculum directors and master teachers are well aware of the benefits to knowing a child’s skill level or level of mastery before formalized instruction begins. Formative assessments are the instruments used by teachers and school districts to help ascertain the current performance levels for students. These assessments can be in written form or often, they are simply demonstrations of student abilities informally observed by an instructor when a student attempts a task that would demonstrate content or objective mastery (Cogshall, 2004). With formalized summative assessments garnering most of the focus of the private and public sectors, formative assessments

appear to be a vital link to student achievement that is often overlooked. They are often overlooked because their diagnostic characteristics are not used to assess *for* instruction vis-à-vis curriculum mapping, pacing, and scope and sequencing but rather, they have been relegated to an assessment *of* post-instruction.

Another key factor in student achievement revolves around tailoring instruction to meet the specific needs of each learner. This differentiation of instruction has occurred for many years in special education classrooms. Teachers were trained in pedagogy that addressed a variety of learning modalities and their instructional program was, and still is, driven by individualized learning objectives. Now, educators are being required to employ these best-practices towards all students often in mixed-ability classrooms (Gregory & Chapman, 2002).

This research used the State of Georgia's CRCT (Criterion Referenced Competency Test) as one assessment measure for student achievement. Georgia, along with most other states, has adopted a standards-based curriculum for school districts to follow. Simply stated, a standards approach means that centralized themes of learning or topics are subdivided into more concise learning objectives or elements and the curriculum attempts to delve deep within a content area rather than cursorily skim the surface. Advocates of the standards-based curriculum state that this centralization of learning objectives provides a coherent and assessable level of achievement for all learners. Opponents of a standards-based curriculum argue that it is dummed down and does not provide for enrichment opportunities nor does it allow higher learning levels of learning to occur. Regardless of where someone stands on the issue of standards-based

curricula, the fact remains that students in most states will attempt to master the learning objectives offered on a high-stakes, cumulative, criterion-referenced assessment. Reeves (2001) states that “balance is notably lacking in high-stakes state tests that claim to be based on standards but that are, in fact, limited by budget and time constraints to providing a brief, basic snapshot of student proficiency on a few standards.” Moreover, this brief snapshot assessment will attempt to evaluate not only the student’s own levels of academic achievement but also, these scores will then attempt to evaluate the effectiveness of a teacher, a school, a school system, and in reality—the effectiveness of a state’s educational system.

Problem Statement

There are many factors that now challenge the classroom teacher, the public school administrator, and the public school policy maker to create an ideal teaching and learning environment for every student, with individualization of instruction to each learner’s needs. Furthermore, the emphasis on accountability as expressed by standardized assessment results has created an educational climate reflective of most business models (Marzano, 2001). In essence, students have not only remained as the primary customer of the educational process, they have now also taken on the role as the primary producer” of achievement data. With this emphasis now on standardized assessment results serving as the qualifier for what an effective education is, local education agencies are being forced to reexamine their curricular standards and objectives, their classroom pedagogy, and their ability to accurately formatively assess students prior to completing a comprehensive, high-stakes summative assessment.

Many school districts have opted to purchase software packages that assess student mastery of various content areas. Also, school districts have hired expert consultants to perform professional development for its teachers in an attempt to realign instruction with prescribed performance standards. These strategies along with many other sound practices are intended to ensure that teachers and students are working towards a common goal, that is, successful academic achievement as measured by standardized criterion referenced exams. Considering this expenditure of resources, it is advisable to study and attempt to conclude that one of these strategies might be effective in raising student achievement. This study examined whether comprehensive post formative assessments can accurately predict student academic achievement on AYP (Adequate Yearly Progress) indicators as measured by standardized criterion-referenced tests. The content areas to be examined were mathematics and reading, the two areas that are Adequate Yearly Progress indicators.

The following research questions and null hypotheses will test whether comprehensive pre and post formative assessments can accurately predict student academic achievement on AYP (Adequate Yearly Progress) indicators as measured by standardized criterion-referenced tests.

Research question 1

Do the post test scores in each of three quarters predict CRCT (pass vs. fail) in year 1?

Research question 2

Do the post test scores in each of three quarters predict CRCT (pass vs. fail) in year 2?

Research question 3

Do the post test scores in each of three quarters predict CRCT (pass vs. fail) in year 3?

The null hypotheses to be examines are:

Null Hypothesis 1

The post test scores in each of three quarters do not predict CRCT (pass vs. fail) in year 1.

Null Hypothesis 2

The post test scores in each of three quarters do not predict CRCT (pass vs. fail) in year 2.

Null Hypothesis 3

The post test scores in each of three quarters do not predict CRCT (pass vs. fail) in year 3.

Professional Significance of the Study

In the State of Georgia, all public school students in grades K-8 must take a yearly comprehensive criterion-referenced exam. This exam, referred to as the CRCT (Criterion Referenced Competency Test), is aligned with the Georgia Department of Education's GPS curriculum (Georgia Performance Standards). This curriculum has been phased-in over the last five years in all public schools across all content areas. For students in the 3rd, 5th, and 8th grades, they must pass the Mathematics and Reading portions of the exam.

If they do not, they are required to be retained. School districts in Georgia also administer the CRCT to the other grade levels but only these three grade levels use the exam as a gateway for promotion. Also, the Mathematics and Reading content domains are used by the United States Department of Education as the AYP indicators to evaluate school system effectiveness.

The Gainesville City School System located in northeast Georgia, where the researcher is employed as a middle school principal, has developed a systemic approach to formatively measure student achievement prior to initial instruction in the classroom. The school district contracts with an outside vendor that supplies a web-based computer program for assessment construction and data analysis of results. The school system has allocated a considerably large amount of resources—both monetary and personnel—to effectively conduct these formative assessments. Many school districts are currently replicating the Gainesville method of allowing formative assessments to drive instruction. Through curriculum mapping and pacing, differentiated instruction, and a standards based curriculum, Gainesville City Schools has exemplified the effective use of formative assessments in the State of Georgia. However, what has not been determined is the validity of these formative assessments in predicting future student success on the CRCT. The significance of this study is that it will attempt to validate the use of school resources towards this formative assessment initiative. In turn, Gainesville City Schools, along with the other systems that are replicating the model, will gain insights into the variables that can deter from the overall effectiveness of the process (Ballowe & Sullivan, 2006).

Definition of Key Terms

To ensure a clear understanding of expression and to provide consistency throughout this study, the following terms and acronyms have been defined:

CRCT. This is Georgia's Criterion-Referenced Competency Test. It is the standardized achievement test administered throughout the state of Georgia to all first through eighth grade students. It is described in greater detail in Chapter 3.

LEA. This represents a Local Education Agency. It is synonymous with the descriptor "school system" or "school district".

AYP. This acronym stands for Adequate Yearly Progress, a term coined by the *No Child Left Behind* legislation that indicates whether a school system or school campus has sufficiently made academic achievement gains when compared to annual measurable objectives.

AMO. These are annual measurable objectives for student mastery that the federal government has dictated under the auspices of *No Child Left Behind*.

Chapter 2

Review of Literature

The Characteristics of School Reform

The Need to Develop Best Practices

The quality of public education in the United States has been under scrutiny since the launch of Sputnik in 1957. When the American politicians and the general public realized they had fallen behind the Soviet Union in the development of space exploration, the American educational experience required an infusion of ingenuity and resources. Since then, educators, legislators, and other stakeholders have continually sought best practices in public education. In an attempt to reform and reorganize the traditional factory-like setting of American public education, reform has been directed towards school administration and instructional leadership, assessment, curricula, pedagogy and instructional practices, differentiation, and behavior management just to name a few.

In 1966 and then again in 1972, educational researchers concluded that school can only positively affect about 10% of a student's gains in academic achievement (Coleman & Jenks, 1966, 1972). They concluded that most differences in test scores are due to factors that schools do not control. Researchers later reexamined these findings and discovered that the 10% gain was inaccurately expressed and that in reality, a gain in 23 percentile points. From this perspective, schools can definitely make a difference in student achievement. More importantly, since the Coleman and Jenks reports, studies have shown that an individual teacher can have a powerful effect on his or her students even if the school does not. This finding makes sense if we consider that Coleman &

Jenks examined the average effect of schools. Within a given school, there is a great deal of variance in the quality of instruction from teacher to teacher. The conclusion that individual teachers can have a profound influence on student learning even in schools that are relatively ineffective, was first noticed in the 1970s as researchers began to examine effective teaching practices. In fact, after reviewing hundreds of studies conducted in the 1970s, researchers Jere Brophy and Thomas Good (1986) commented, “the myth that teachers do not make a difference in student learning has been refuted.”

A team of researchers from Michigan State University took issue with Coleman’s conclusions. Convinced there was something wrong with the report, Professors Larry Lezotte, Ron Edmonds and Wilber Brookover launched an investigation because they wanted to know why some schools are effective and others are not. The researchers found two types of schools that Coleman’s research would have contended could not exist. The first were highly effective schools that served mainly students from low- and middle-income families. Second were schools with dismal performance records that served children from middle- and upper-income families (Edmonds, 1979 & Brookover, Lezotte, 1979). The researchers identified five characteristics, or correlates, that were common to all effective schools:

1. Strong instructional leadership by the principal that frames the school’s vision and turns it into reality
2. High expectations of student achievement by students and staff
3. A broadly understood instructional focus that centers on reading, writing, and mathematics

4. A safe and orderly school climate conducive to teaching and learning
5. Frequent measures of student achievement as a basis for program evaluation and improvement

The Effective Schools Movement, as it would come to be known, presented impressive evidence that background was not a factor in a child's ability to succeed. Truly effective schools could teach all children. This did not mean that all children can learn at the same rate, on the same day, or in the same way. Instruction must be customized to meet each child's unique needs and abilities. Eventually, this research would also spark educational reform in school leadership, formative assessments, differentiated instruction, and a host of other pedagogical strategies to address individualized students needs.

More recently, researchers William Sanders and his colleagues (Sanders & Horn, 1994; Wright, Horn, & Sanders, 1997) have noted that the individual classroom teacher has even more of an effect on student achievement than originally thought. As a result of analyzing the achievement scores of more than 100,000 students across hundreds of schools, their conclusion was:

The result of this study will document that the most important factor affecting student learning is the teacher. In addition, the results show wide variation in effectiveness among teachers. The immediate and clear implication of this finding is that seemingly more can be done to improve education by improving the effectiveness of teachers than by any other single factor. Effective teachers appear to be effective with students of all achievement levels, regardless of the

level of heterogeneity in their classrooms. If the teacher is ineffective, students under the teacher's tutelage will show inadequate progress academically regardless of how similar or different they are regarding their academic achievement (Wright et al., 1997).

Examples of Reform Models

A review of the literature suggests that several of the attempts at positive school reform fit into Dr. Robert F. Mager's theory of Criterion Referenced Instruction (Kearsley, 2006). Mager was influenced by Gagne, Knowles, and Rogers, and attempted to incorporate the spirit of their work into Criterion Referenced Instruction (Cooper, 2005).

Some of the critical aspects of Mager's theory include: (1) goal/task analysis – to identify what needs to be learned, (2) performance objectives – exact specification of the outcomes to be accomplished and how they are to be evaluated (the criterion), (3) criterion referenced testing – evaluation of learning in terms of the knowledge/skills specified in the objectives, (4) development of learning modules tied to specific objectives (Kearsley, 2006).

Although Mager's theory has been in use since *A Nation at Risk* (1983), today's educators with two decades of experience are looking at it from a standards-based perspective. Their task analysis identifies what standards need to be learned. Their performance objectives are standards-based. Their criterion referenced testing is based on those standards. And, of course, their learning modules are tied to those standards. A myriad of models have been developed which incorporate some or all of these facets.

The purpose of this literature review is to examine various standards-based models and their effectiveness as measured by student achievement scores.

The Los Angeles Unified School District's Standards-Based Instruction Model gives teachers the standards-based task, the assessment type, and the instructional activities. It also adds the time frame for each task and a list of materials needed (LAUSD, 1997). Objections to this type of model are summed up by Deborah Bambino (1999), "I'm convinced that lock-step, rigid, age-based assumptions are self defeating." She and many others complain that it leaves no room for individual teacher creativity.

Mid-Continent Regional Educational Laboratory in Aurora, CO (1998), developed a model which provides teachers with standards and standards-based objectives, but it uses portfolio evaluations rather than criterion referenced testing. Portfolio evaluation is time consuming for teachers. Judy Swanson (2001) feels that looking at student work has reached the status of a buzz word and that too few educators really understand how to do that, what to look for, and what value can be derived from examining student work to inform instruction.

Through a collaborative effort with educators and content providers project Virtual Information Education Web (VIEW) (Barbanell, Falco, & Newman, 2004) developed methods and models for teaching and learning using interactive video and Internet technology. Teachers worked with the content providers to develop innovative, replicable models for delivering high quality content from distant locations. These models provide students the opportunity to study and learn through interactive observation. The product allows students a wealth of opportunities that they might

otherwise never experience. They engage with scientists, historians, curators, librarians, and content experts as well as with artifacts and an array of resources, through the use of interactive videoconference and electronic technology. During the videoconferences, student skills, rated according to the six stages of Bloom's taxonomy, were observed and documented. Nearly half (42%) of the programs involved the students in the upper levels of Bloom's Taxonomy (Barbanell, et al, 2004). These models were not built on the standards of Public Law 107-110 (US Dept of Education, 2007). This process could easily be used to build standards-based models.

From this sample of articles, it is evident that educators, researchers, foundations, and corporations are interested in improving the nation's public schools. School reform models are being formed to test their ideas about the kinds of changes that will provide a better education for their students and meet the objectives of NCLB (EdSource, 2006).

To fairly assess a model's effectiveness, it is important to look at what progress the model has made over time (School Reform Models Overview, 2006). According to Keith Zvoch (2003), the most respected approach to measuring school performance is to measure changes in student achievement based on the calculation of individual growth trajectories. In this approach, student test scores are linked across time. A regression function is then fit to the outcome data obtained on each student. The resulting growth trajectories index the rate at which individual students acquire certain academic competencies. A measure of school performance, mean growth, follows from averaging the individual growth trajectories within each school.

As a measure of school performance, school mean growth has important advantages over other methods for evaluating schools. By tracking the gains that students make rather than indexing the level at which they perform, school mean growth is less confounded with the influence of socio-demographic factors that contaminate the unadjusted school mean (Stevens, 2000) and is a more transparent and practical than the adjusted school mean. The strength of tracking individual student gains is also evident when compared relative to the quasi-longitudinal approaches. A measure based on a longitudinally matched student cohort is not impacted by the year-to-year variations in student body composition that complicate the unmatched cohort approaches, thereby providing a more precise indication of the actual change in student achievement (Haug & Linn, 2002). Evidence that a measure based on the yearly change in student achievement is a less biased and a more precise indicator of school performance suggests that states and school districts would benefit from incorporating a measure that tracks individual student growth into their accountability systems. For school evaluation purposes, use of a methodology that provides control over the stable characteristics of students (e.g., socio-economic status, ethnic/cultural differences) that otherwise complicate the evaluation of school effectiveness facilitates a more valid comparison of schools that differ in the intake characteristics of their student bodies (Zvoch & Stevens, in press). Equally important, identification of variation in school growth rates enables investigation of the school-based factors that are effective at promoting student growth in achievement.

The Brazosport Independent School District in Texas applied the five effective school correlates coupled with Deming's Total Quality Management approach in an

effort to raise student achievement and close the achievement gap in a socio-economically polarized community where inequities in education were clearly evident (Anderson, 2002). Brazosport was a school district with two high schools that epitomized the achievement gap that exists between socio-economic groups. One high school was affluent and its achievement scores reflected a higher level of effectiveness whereas the other high school was failing and appeared to add strength to the *Coleman Report of 1966*. In 1994, Anderson, along with key educators in the system, developed an eight-step instructional process that is currently being replicated throughout the United States as a process to close the achievement gap between minorities and lower socio-economic groups. The key components of this process revolve around the use of student test data, effective lesson planning, a continual focus on basic skills, and tutorial and enrichment opportunities for individual learners (Anderson, 2002). Based upon Deming's cyclic nature of self-assessment, a great deal of responsibility is placed upon the school administrator to become the instructional leader of the campus rather than a manager of administrative affairs. This process developed in Brazosport began a nationwide movement towards data disaggregating of sub-populations and a no excuses approach to targeting the inequities in minority and low-socio-economic student achievement (Anderson, 2002).

Key Factors to Influence Positive Change

Research has clearly indicated that there are two factors that greatly influence student achievement, either negatively or positively. The overall effectiveness of the school leader (principal) and subsequently the effectiveness of the classroom teacher

dictate a child's academic success. Yes, other factors play a part such as parental support and involvement, adequate funding and resources, and school climate but they are ancillary in nature. The principal's leadership role in assessment then must become clearly defined and articulated.

Principals and their faculties face a staggering array of old and new assessment challenges. These include challenges revolving around standardized testing as well as day-to-day classroom assessment. Most troubling is the fact that because of a long-standing gap in their professional preparation, neither principals nor teachers possess the assessment "literacy" needed to meet these increasingly complex demands (Marzano, 2005). Although the typical teacher can spend as much as one third of their available professional time involved in assessment-related activities, teacher and administrator licensing requirements and training programs have for decades failed to provide essential assessment training. The implications of this lack of crucial training are numerous (Black & Wiliam, 1998).

Sound assessment represents one essential key to school effectiveness. If standardized tests are understood by their intended users, or if classroom assessments are of high quality, then sound instructional decisions may be made on the basis of the data such tests generate and student achievement may increase. But if standardized tests are misunderstood or poorly used or if classroom assessments are of poor quality, then poor decisions may be made on the basis of the test-generated data, instruction may be ineffective, and students may suffer (Stiggins, 2000). The problem is that because generations of teachers and administrators lack assessment training, educators cannot

assure their stake holders that standardized tests are being effectively used or that teachers are accurately assessing the achievement of their students. Principals have two crucial responsibilities regarding assessment literacy. First, they must become assessment literate themselves. Without this basis of professional expertise, principals will remain unable to bring the issue of effective assessment to the forefront as a school priority or provide the support teachers need to develop and use assessments effectively in their classrooms. Second, principals must remove all barriers to the development of teachers' assessment literacy. These include personal, institutional, and community barriers (Ingersoll, 1999). Personal barriers may include the anxiety that accompanies trying new assessments before one is certain that they will work. The principal needs to assure teachers that initial failure to assess dependably or to use assessment effectively will not lead to a directive to stop trying. Institutional barriers may include a lack of time to learn and to experiment with new assessment ideas. Teachers need to know that school resources will be allocated for these purposes—and the principal needs to make sure that they are. Community barriers may include parents who question changes in assessment and communication procedures. Principals need to be assessment literate to be able to ease community concerns and to support teachers in their relationships with parents during the process of change (Ingersoll, 1999). Leadership is needed to create an instructional environment that expects and supports competence in assessment, as well as the effective application of that competence in the service of students' academic achievement.

Standards-Based Curricula

A Brief History

The movement for education reform in the United States was an outcome of the public debate on social, economic and political issues ensuing from the release of a report by Peters and Waterman (1982). Extended to the education sector, this debate resulted in a spate of national studies on excellence in education, following the release of the report of the National Commission on Excellence in Education (1983). Generally, the reports of these studies were critical of the poor quality of public education, recommending a variety of strategies to reform education, particularly at the secondary level. Two waves of reforms during the 1980s affected improvements through small-scale school reform projects and by decentralizing decision-making authority to local communities, but failed to bring about national education reform.

Convened by President George H. W. Bush in September 1989, the Charlottesville Education Summit involved the President and the 50 state governors considering ways of bringing about changes in the education system that would make the United States internationally competitive by the year 2000. They reached agreement to establish a process for setting national education goals, seeking greater flexibility and accountability in using federal resources to meet the goals, undertaking a state-by-state effort to restructure the education system, and reporting annually on progress in achieving the goals (Vinovskis, 1999). Promulgated in February 1990, the six National Education Goals became the foundation for America 2000 and later Goals 2000, and provided the impetus for defining national standards based in academic disciplines.

A variety of trends in education had concurred by this time leading conservatives and liberals to forge a consensus about focusing on what students should know and be able to do. Policy-makers set nationally recognized groups in key disciplines the task of developing national standards consisting of content, performance, and opportunity-to-learn standards. Content standards refer to broad descriptions of knowledge and skills that students should achieve in particular subject areas. Performance standards are examples and definitions of knowledge and skills in which students need to demonstrate proficiency. Opportunity-to-learn standards, which address conditions necessary at each level of the education system to provide all students with opportunities to master content standards and meet performance standards, provide criteria covering six elements. These elements refer to the quality and availability of curricula, materials and technology, the capability of teachers to meet learning needs, the availability of professional development, the alignment of the curriculum to content standards, the adequacy of school facilities for learning, and the application of non-discriminatory policies.

(Symcox, 2002)

National Standards

The first effort to develop national standards preceded any initiative undertaken by the federal government. McLeod et al. (1996) reported that the national standards for mathematics originated from the work of the Commission on Pre-college Education in mathematics, science and technology, which released a plan of action for improving mathematics, science and technology education for all school students (National Science Foundation, 1983). Discussions at a series of conferences led the National Council of

Teachers of Mathematics (NCTM) to appoint a Commission on Standards for School Mathematics in 1986 to oversee the development of the national standards for mathematics. Four working groups consulted focus groups within the education community to develop the national standards for Mathematics, which were released in March 1989. Subsequently, NCTM published professional standards for teaching mathematics in March 1991 and assessment standards in May 1995. In 1995, NCTM initiated a project to revise and combine the three sets of standards into a single volume, which was released in April 2000 (Symcox, 2002).

In June 1991, the National Education Goals Panel created the National Council on Education Standards and Testing to examine the feasibility of national standards and a national system of assessments, and to recommend policies, structures and the mechanisms for setting them. In its report, the National Council on Education Standards and Testing (1992) recommended that voluntary and dynamic national standards should be developed initially for English, mathematics, science, history and geography, which reflected high expectations, focus and direction. In addition, multiple measures consisting of individual student and large-scale sample assessments aligned to the national standards should be set. This recommendation prompted the United States Department of Education to fund projects by nationally recognized groups to develop national standards for science, history, the arts, civics and government, geography, English language arts, and foreign languages in 1991 and 1992 (Ravitch, 1995). In addition, independently funded projects were initiated to develop national standards for social studies, health, physical education, and economics.

However, a controversy of national proportions arose between liberals and conservatives during the development of the national standards for history. The developmental process led to confrontation between minority groups seeking greater representation of their ethnic heritages and conservative groups seeking to represent democratic principles binding the United States together as a nation. In spite of resolving differences between these groups over the issue of multiculturalism and establishing criteria for world history during the standards-setting process, the national standards for history became controversial two weeks before their release. Lynne Cheney, the former chairperson of the National Endowment for the Humanities, published a criticism in the *Wall Street Journal* in October 1994. It argued that the national standards for history represented the effort of a small, radical group of academics, portrayed multicultural excess, and failed to depict the celebratory aspects of US history or emphasize Western civilization in world history. A few days after Cheney's attack, Rush Limbaugh, the popular conservative talk show host, told his audience that the national standards for history were part of the America-bashing multicultural agenda. Unleashed by Limbaugh's comments, conservative attacks were followed in December 1994 by adversarial debates on television between Cheney and prominent historians. The criticism then moved into the political arena, when the Senate passed a resolution in January 1995 condemning the national standards for History by a vote of 99 to 1. In the August 1995 issue of *Time*, Republican House Speaker Newt Gingrich wrote that the US history volume distorted and undermined US history. Senate Majority Leader and Republican presidential candidate, Robert Dole, speaking to the American Legion at a Labor Day ceremony in

Indianapolis in September 1995, said that the national standards for history disparaged America and its Western tradition. Soon afterwards, Secretary of Education Richard Riley responded by registering his own and President Clinton's opposition to using the existing standards as a basis for history curricula in schools. However, officials of several national standards projects had met with leading critics of the national standards for history in January 1995. In an effort to save the national standards for history, the National Center for History in the Schools agreed to the Council for Basic Education (CBE) convening two panels of historians, educators and public officials to determine whether they could be revised. In October 1995, both panels, one of which had examined the United States history standards while the other had reviewed the world history standards, announced that the national standards, though flawed, could be revised. They found that the overwhelming majority of criticisms were targeted at teaching examples in the documents, rather than the actual standards (Nash, 1998). The national standards were revised between November 1995 and February 1996 by staff of the National Center for History in the Schools, assisted by a small group of history educators. A newly formed Advisory Board to the National Center for History in the Schools appraised the revisions in December 1995, and the two review panels and CBE endorsed the revised edition, which had compressed the original edition's three volumes into a single document. In spite of this process, the opinions of conservatives, divided about whether the revisions overcame their objections, led to Republicans in the House of Representatives attempting to censure the revised national standards for history in September 1996. However, the press received them favorably, and the controversy died away, but it had been so divisive

that it led to numerous published interpretations, of which those by Nash et al. (1998) and Symcox (2002) best reflects the liberal viewpoint and Cheney (1995) the conservative standpoint. The bipartisan political support evident at the commencement of national standards' projects, however, dissipated following the controversy surrounding the national standards for history. Although the conservative Right's attacks undermined the consensus for developing national standards following the election of Republican majorities to both houses of Congress in November 1994, the movement for standards-based reform was reinvigorated by several events. These included the second National Education Summit convened in March 1996, the re-election of President Clinton in November 1996, the State of the Union address in February 1997, the third National Education Summit held in September 1999, and the fourth National Education Summit convened in October 2001 (Symcox, 2002).

Further developments within standards-based education also played an important part in its revival. The lack of consistency between the national standards developed by the different subject-based groups led national organizations to synthesize the work of these projects. Issues relating to state-level standards-based reforms led national organizations to design information services and evaluation models to assist states, school districts and schools implement state standards. Another important activity fostered by a national organization was the establishment of a national forum on standards-based education.

In July 1995, CBE initiated the Standards for Excellence in Education Project to synthesize the national standards' documents in the core subject areas into a more useful

form for educators, parents, business leaders and the public. A working group of CBE staff analyzed the documents produced by the mathematics, science, civics and government, history, geography, the arts, English language arts, and foreign language projects. After conducting a series of focus group meetings in 1996 to obtain responses regarding alternative formats for a single document, the working group established a common vocabulary for synthesizing the standards, and defined benchmarks for grades 4, 8 and 12 across the eight subject areas. The outcome of the project, a book presenting condensed, edited and commonly-formatted versions of the national standards in the eight subject areas, allowed users to trace the presentation of the material back to the original source documents (Council for Basic Education, 1998).

By analyzing different perspectives taken by subject-based groups involved in developing the national standards, researchers based at Mid-Continent Research for Education and Learning (McREL) concluded that analysis and synthesis of standards and benchmarks, specified in the national standards' documents, were needed. They classified the standards and benchmarks identified in 116 standards' documents published by national, state, and private groups, into three types of knowledge. Procedural knowledge involves processes critical to the content area. Declarative knowledge consists of information important to the content area, which is often acquired through understanding its component parts. Contextual knowledge includes information or skills that give particular meaning, because of the conditions that form part of their description. As well as coding standards according to these three categories, they classified standards and benchmarks into four bands: Level I for grades K to 2; Level II for grades 3 to 5; Level

III for grades 6 to 8; and Level IV for grades 9 to 12. Applying these concepts to the analysis, 256 standards and 3968 related benchmarks were identified across 14 subject areas. Based on this analysis, McREL designed a database of standards and benchmarks, known as McREL's compendium, linked by subject areas to various web sites providing lesson plans, activities and curriculum resources for school districts and schools to construct their own standards and benchmarks (Kendall and Marzano, 1997). The effort to analyze standards' documents led to a concern that the amount of classroom time available to teach the full range of standards may be inadequate, a presumption that further research showed to be correct. McREL then conducted a study to synthesize the standards and benchmarks from the five most highly rated state standards' documents as a means of reducing the subject area content coverage to a manageable level. The content found in the exemplary documents was classified by mapping the content against McREL's compendium. This process led to the production of a master document containing all the knowledge and skills identified in the five documents, and where it was located in each document. The common knowledge and skills was identified to produce sets of benchmarks organized around essential standards for language arts, mathematics, science, geography and history.

In November 2000, McREL convened a meeting of 35 national education leaders to design the National Dialogue on Standards-Based Education, which was launched in April 2001 in Kansas City, Missouri, where 130 participants distilled a list of topics related to standards-based education, and wrote collective statements on each topic. In collaboration with Public Agenda, McREL developed a process, based on the National

Issues Forums, for participants to conduct three-hour dialogues, and provided training for local facilitators.

Founded in October 1996 as an outcome of the second National Education Summit, the Achieve Resource Center on Standards, Assessment, Accountability and Technology for Governors (Achieve) played an important part in organizing the third and fourth National Education Summits. Achieve also designed a standards database, consisting of standards for mathematics, English language arts, science, and history and social studies organized into a consistent structure using McREL's compendium. Data on the standards of each state, territory and the Department of Defense Education Activity were collected from liaison officials appointed by each chief state school officer. Content area experts then reviewed the submitted materials, and tagged each standard using McREL's compendium.

In 1998, Achieve collaborated with CBE and the Learning Research and Development Centre at the University of Pittsburgh to develop a process for benchmarking state standards. Achieve provides four benchmarking services tailored to particular states' requirements. State standards may be benchmarked through brief or in-depth reviews. The brief review provides basic feedback on the content of standards as part of the developmental process. Achieve identifies the standards' strengths and weaknesses, and offers states action steps for improvement. The in-depth review involves thorough evaluation based on comparisons with exemplary standards from other states and countries, followed by detailed feedback and recommendations for improvements. Achieve also ensures that assessments a state is administering to students are aligned to

the state's standards by examining them as a package, based on comparisons with other states and countries. Benchmarking also includes institutes for policy-makers designed to build capacity in aligning standards and assessments. Comprehensive reviews of systemic reform policies are also provided for states. A team of prominent experts reviews various aspects of a state's education system, state policies and practices, interviews stakeholders, and makes recommendations to build on the reform strategy. Achieve's benchmarking services have been commissioned by Delaware, Florida, Illinois, Indiana, Maryland, Massachusetts, Michigan, Minnesota, Missouri, New Jersey, Ohio, Oklahoma, Oregon, Pennsylvania, Texas and Washington (Council for Basic Education, 1998).

Founded in September 1992 by 51 national business organizations, the America 2000 Coalition served as a resource to raise public awareness of the need to achieve the National Education Goals, and for local communities to implement America 2000 strategies. In 1994, the America 2000 Coalition changed its name to the Coalition for Goals 2000, and developed an information system, GoalLine, providing members with a bulletin board, updates, a conference area, electronic mail, and a database of promising programs, standards, and assessments available in the United States. GoalLine was launched nationally in September 1994, and made available on the GoalLine web site in August 1998. In collaboration with Denis P. Doyle and Associates, the Coalition for Goals 2000 commenced a two-year project in 1995 to define the standards-setting process. First published in 1997 as a book for local educators and citizens, Doyle and Pimental (1999) released a revised version outlining an eight-step plan for the standards-setting process, illustrated by case studies of standards-based reforms in five school

districts. In June 1997, the Coalition for Goals 2000 founded StandardsWork as a consulting project to provide technical assistance in standards' planning, drafting, benchmarking and alignment, student diagnostics, rapid response assistance to embattled school districts, and institutes for training. In January 2000, StandardsWork and the Education Leaders Council (ELC), formed by the chief state school officers of Arizona, Colorado, Florida, Georgia, Michigan, Pennsylvania and Virginia at a meeting at Burlington, Vermont, in July 1995, began collaborating on a project to create a results' card. Following agreement reached in the summer of 1999 on which data elements needed to be monitored, StandardsWork and ELC launched a pilot project in which the seven states used a prototype of more than 60 indicators to collect and analyze data to accurately determine school and student progress over time. In 2000, Maryland, Ohio and Pennsylvania joined the six remaining states to provide state data. In 2001, StandardsWorks published a prototype Results Card, an annual report analyzing the impact of each state's goals and policies on improving student performance, and a report examining the multiple measures of performance used in the Results Card. In March 2002, StandardsWork convened a conference to streamline the Results Card, simplify data collection procedures, and invite participating and new states to join. However, the project did not continue because of the perception that new regulations in the No Child Left Behind Act, making such data collection non-negotiable, would undermine a potential market for the Results Card.

State Standards

The *Goals 2000: Educate America Act*, passed by the Clinton administration in March 1994, required state education agencies to use the national standards as blueprints to develop and implement state standards and curriculum frameworks, which are aligned to state assessment systems. From July 1994, state education agencies applied to the United States Department of Education for Goals 2000 grants under Title III to develop and implement comprehensive, education improvement plans, which included establishing challenging state standards. The *Goals 2000: Educate America Act* required each state education agency to appoint a broadly representative panel to develop state improvement plans in consultation with the state governor and the chief state school officer. The *Improving America's School Act*, passed by the Clinton administration in October 1994, required each state to develop state content and performance standards for mathematics and reading by the 1997-1998 school year and assessments by the 2000-2001 school year appropriate for all students, including the disadvantaged (Symcox, 2002).

Following enactment in December 2001 of the *No Child Left Behind Act* by President George W. Bush, Secretary of Education Rod Paige convened a negotiating committee in March 2002, which received advice from 140 interested parties on developing new standards and assessment provisions. In July 2002, Secretary Paige issued new proposals, and invited public comments, to which 140 interested parties submitted over 700 comments. In November 2002, Secretary Paige released the final regulations, requiring that by the 2005-2006 school-year, each state must measure

students' progress in reading and mathematics in each of grades 3 to 8, and at least once during grades 10 to 12. By the 2007-2008 school-year, states must also administer assessments in science at least once each during grades 3 to 5, 6 to 9, and 10 to 12. At the beginning of 2003, each state was required to establish a definition of adequate yearly progress to use each year to determine the achievement of each school district and school. Definitions were required to meet 10 criteria. First, a single, statewide accountability system must be applied to all public schools. Second, all public school students must be included in the accountability system. Third, adequate yearly progress must be based on expectations for growth in student achievement that are continuous and substantial. Fourth, the state must make annual decisions about the achievement of all public schools. Fifth, all public schools must be accountable for the achievement of individual subgroups. Sixth, adequate yearly progress must be based primarily on the state's academic assessments. Seventh, adequate yearly progress must include graduation rates for high schools, and an additional indicator selected for middle and elementary schools. Eighth, adequate yearly progress must be based on reading and mathematics achievement. Ninth, the accountability system must be statistically valid and reliable. Tenth, the state must ensure that at least 95 percent of students in each subgroup enrolled in a school are assessed. In defining adequate yearly progress, each state sets the minimum levels of improvement that school districts and schools must achieve within time frames specified in the *No Child Left Behind* Act. Each state begins by setting a starting point that is based on the performance of its lowest achieving demographic group or the lowest achieving schools. The state then sets the level of student achievement that

a school must attain in order to make adequate yearly progress. Subsequent thresholds must increase at least once every three years until at the end of 12 years, all students in the state are achieving at the proficient level in state assessments of reading language arts, mathematics and science.

The standards-based education became entrenched in the states during the 1990s. This situation arose from the domination of the national education policy agenda with the concept that academic standards should provide direction for developing curricula and assessments, and should be linked to teacher development, accountability and other education policies. Both conservative and liberal policy-makers agreed on the merits and worth of this approach to reform, which persisted despite changes in political leadership and criticisms about the quality of particular standards and assessments. Policy-makers overcame these criticisms by adopting mixed-models that balanced newer and more traditional approaches to content, assessments, professional development and other aspects of education reform. The tendency to maintain vitality by shifting emphasis from content in the initial phase to assessment in the most recent phase suggests that standards-based education may not follow some other reforms by shining brightly for a few years and then fading. If standards-based education persists as the main driving force in the national education reform strategy in the United States, it is likely to have an even more profound influence in the future on curriculum reforms in other countries. As a nation, the United States does develop national educational standards emphasized to public school curricula and content domains. States, however, also have the ability to formulate their own educational standards for public schools. The national standards movement

can influence state-level standards under the auspices of NCLB. The national government sets what it considers reasonable AMOs (Annual Measurable Objectives), which terminate in the school-year 2013-14 when all students are required to reach a 100% mastery-level of academic achievement. States are required to develop their independent AMOs to reach the 100% mastery goal of 2014. If states, local education agencies, school districts, or even campuses consistently do not meet adequate yearly progress in relation to the AMOs, federal funding can be withdrawn or reallocated through the Title programs originally established under the *Elementary and Secondary Education Act of 1965*. This is why most states allow their standards to mirror the standards developed at the national level.

The Link to Performance and Formative Assessments

When one mentions standardized testing to a group of educators, controversy and debate begin as teachers begin to demand fairness while administrators echo the political sentiment of accountability for all. Sadly, neither side appears to be willing to shift the focus of the debate from an annual high-stakes testing event to the primary location for effective assessment—the classroom.

School leaders have an obligation to counter the prevailing myth surrounding standards and testing and communicate that standards and standards-based assessments are effective, fair, and essential. Advocacy for standards does not prevent school leaders from an open and honest acknowledgment of the flaws in many standards. Popham (1997) notes that many state standards lack descriptive rigor, and Marzano (2000) has noted the proliferation of standards exceeds the capacity of the typical time available in a

school year. These observations are a good rationale for the improvement of standards with respect to their clarity and focus. The state of Georgia has recently over-hauled its entire curriculum and instituted a standards-based curriculum with distinct performance measures. This is a radical departure from its former QCC (Quality Core Curriculum) that expressed content almost in the terms of what needs to be covered by the teacher. Georgia administrators, along with other public school leaders, must ask themselves, “If standards were abandoned, what then would *we* do?” The answer is simple. When student work is not assessed towards a standard, it is then compared to the work of other students. In other words, to reject standards-based assessments is comparably to advocating a bell curve. If NCLB has taught us anything, it is that the bell curve is no longer acceptable for statistically representing student achievement. By 2014, NCLB requires all students to be performing at 100% mastery in the content areas vis-à-vis the erosion of the two standard deviations above average (U.S. Dept of Education, 2007).

With this kind of pressure being exerted, school leaders and teachers are more and more resulting to a belief that the path to better performance on standardized tests is to drill students and work on test taking strategies rather than the development of reasoning, thinking, challenge, and communication inherent in performance assessment (Kohn, 2000). The Brazosport Model described earlier in this chapter is based upon that simple premise. Students are segregated into groups for remediation based upon what skills they do not possess. Then, the groups are drilled on these basic skills until they are mastered (Anderson, 2002). Moreover, the implication is that a teacher who is committed to the reasoning, thinking, and communication involved in performance assessment will engage

his or her students in such assessment at the risk of lower standardized test scores. Kohn (2000) states this fear:

The time, energy, and money that are being devoted to preparing students for standardized tests have to come from somewhere. Schools across the country are cutting back or even eliminating programs in the arts, recess for young children, electives for high schoolers, class meetings (and other activities intended to promote social and moral learning), discussions about current events (since that material will not appear on the test), the use of literature in the early grades (if the tests are focused narrowly on decoding skills), and entire subject areas such as science (if the test covers only language arts and math). Anyone who doubts the scope and significance of what is being sacrificed in the desperate quest to raise test scores has not been inside a school lately.

Although these presumptions are common, their popularity does not match the evidence. In fact, the opposite is true—students whose teachers focus on writing, thinking, and reasoning has not only more engaging and interesting classrooms but also have higher standardized test scores. Researchers at San Diego State University (Klentschy, Garrison, and Amaral, 2000) conducted a four year study for the National Science Foundation that compared students who wrote extensively in blank science journals to students whose science education came from the traditional textbook. The students who focused on fewer science topics but with extensive writing had scores on the Stanford 9 science test that were twice that of students in the regular curriculum. Not only did these students excel in the core skill of literacy but they also improved in science. Other

evidence from Missouri (*St. Louis Post-Dispatch*, 9/14/00) reports that on science and social studies exams, results soared not through mindless test drill but after a year-long writing program in which students were required to have monthly writing projects in every class and every subject. This is also consistent with earlier studies (Darling-Hammond 1997; Reeves 2000) that found strong relationships between writing and high multiple-choice scores in math, science, and social studies.

The critics of standards and assessment and the school leaders advocating test prep appear to be on opposite ends of a debate. Both sides, however, have one thing in common: they both believe that effective classroom assessment practice and higher student scores on standardized tests represent a negative correlation—that is, an increase in one will cause a decrease in the other. Also, the political importance of high-stakes test scores cannot be denied, school leaders must carefully consider the evidence when plotting their strategy for improved student achievement. Although it is tempting to use a test prep curriculum, the literature and pedagogical theory do not support the replacement of great teaching and rigorous requirements with model test items. Thinking, reasoning, writing, and rigor will put students on the path toward better test scores and are the characteristics of the best standard-based performance assessments (Reeves, 2000).

Formative Assessments

Current Research and the Elements of Implementation

Classroom assessment that supports student learning, or formative assessment, is strongly favored in current educational literature. Formative assessment has been championed by assessment specialists (e.g., McTighe & O'Connor, 2005; Stiggins &

Chappuis, 2005), and it is increasingly endorsed by professional organizations (e.g., Joint Committee on Standards for Educational Evaluation, 2003; Miller, 2005). Formative assessment is thought to have intrinsic acceptability (Black and Wiliam, 2003) to teachers, but system wide implementation has met with some resistance at the secondary level (Hayward & Hedge, 2005; Smith & Gorard, 2005). The Center for Educational Research and Innovation (CERI, 2005) notes that powerful bureaucratic constraints limit the implementation of formative assessment in secondary schools, even though its use with adolescent students is particularly justified. In their international case studies, CERI (2005) concludes that the benefits of formative assessment outweigh the barriers to its implementation. Although they are negative, these barriers are also a natural part of the transformation currently underway in education. Black & Harrison (2001a) point out that “the development of formative assessment has led to more radical changes in the ways of working of many of the teachers involved” and that “it takes time and patience to achieve changes of this type.” As part of that process, empirical research on formative assessment captures not only ways of working and learning in classrooms, but also the ongoing dialogue between researchers about the changes.

The concept of formative assessment evolved from the early definitions by Bloom, Hastings and Madaus (1971), Ramaprasad (1983), and Sadler (1989). The term formative assessment has been used in conjunction with the popular term, assessment-for-learning (Assessment Reform Group [ARG], 1999; Black, 2003; Earl, 2003), and the two are now considered conceptually identical (Threlfall, 2005). However, considerable confusion remains regarding the nature of the concept (Black, 2003). As Yorke (2003) observes,

“formative assessment is a concept that is more complex than it might appear at first sight” (p. 478). It is sometimes described as a linear sequence, involving teacher-directed instruction, feedback and correctives (Guskey, 2005). This definition fits well with behaviorist or early cognitive theories of learning (Allal & Ducrey, 2000; Yorke, 2003), but its limitations are seen in classroom-based research. In their analysis of formative assessment events, Pryor and Torrance (1998) found that a purely cognitive approach minimized the complexity of the situation (p. 170) by overlooking the social aspects of classroom learning. Extending beyond prescribed instruction, formative assessment is fundamentally a collaborative act (Yorke, 2003, p. 496) that necessitates interaction between teachers and students. Several recent definitions detail the characteristics and elements of formative assessment (ARG, 2002; Cowie & Bell, 1999, CERI, 2005; Crooks, 2001; Leahy, Lyon, Thompson & William, 2005; Shepard, 2005; Stiggins, 2002; Torrance & Pryor, 2001). In analyzing these sources, formative assessment is described as a composite practice, involving: a) clearly communicated learning goals and evaluative criteria, b) varied approaches to elicit information about learning, including questioning and observation, c) balanced and descriptive feedback in varied forms, d) the adjustment of teaching and learning as a result of the assessment, and e) the active involvement of students. First, assessment, teaching, and learning are ideally integrated within a safe learning environment. Second, a multitude of internal and external factors, including teachers’ knowledge and beliefs, impact the elements of formative assessment in practice. This model is presented, not as a static or permanent definition of formative assessment, but as a framework for further analysis and discussion around the concept, as it is presently understood (Roos & Hamilton, 2005).

Formative Assessment in the Secondary School Context

Since the study group in question for this research is a middle school in Georgia, the researcher explicitly focused upon secondary school research in regards to formative assessments. From the research, five distinct elements arose as the building blocks for successful implementation of formative assessments in the public school classroom. The five key elements are: student involvement, feedback to students, adjustment of teaching, explicit learning goals or assessment criteria, and varied approaches to elicit learning. Some elements of formative assessment do appear in the results and discussion sections of the articles more than others. Student involvement (80%), feedback to students (70%), and explicit goals or criteria (53%) are more frequently mentioned than the use of assessment to inform instruction (47%), or varied assessment approaches (40%). At least three elements are mentioned 67% of the time, supporting the notion that formative assessment is a complex activity (Cowie & Bell, 1999). What can be understood about formative assessment from the analysis of works is discussed below in relation to each of the five central elements.

Feedback to Students

There is considerable focus in this set of articles on giving feedback to secondary students. Different types of feedback are discussed, such as comment-only marking by teachers (Black & Harrison, 2001; Wiliam et al., 2004), oral feedback offered informally and responsively during classroom activities (Bell & Cowie, 2001), or computer-generated feedback that is tailored to specific errors (Thissen-Roe et al., 2004). Rubrics are used as a feedback tool, to direct student attention to specific dimensions of an assignment (Hermann & Lewis, 2004), or to guide feedback conversations that involve peers in discussion about

learning (Hickey & Zuicker, 2005, p. 297). While these studies do not give indication of the relative merits of these different methods of feedback, positive consequences are generally seen. Feedback is described as an effective means of scaffolding learning (Leat & Nichols, 2000; Hodgen & Marshall, 2006), and encouraging greater student autonomy (Kirkwood, 2000). However, feedback is rarely considered in isolation from other elements of formative assessment, and only one study attributes increased student achievement specifically to feedback. Hickey & Zuicker (2005) encouraged greater use of feedback by students through the design of their study in its second year, finding that “the improved learning outcomes over time appear to be mostly due to continued enhancement of participation in the feedback conversations (p. 298). To improve the length and quality of the feedback conversations, the teacher modeled the use of the feedback rubric for students, and the successful outcome of the study can, therefore, be associated with the teachers’ pedagogical skill. Several studies conclude that teachers need guidance in this area. For example, Yung (2001) argues that teachers in Hong Kong should be provided with professional development on the use of feedback to motivate students and support learning. In their survey of assessment practices in the Netherlands, Stokking and colleagues (2004) note a wide range in the type, form, and quality of feedback, and they observe that some of the reported practices are less than ideal, especially for learning purposes.

Assessment Informing Teaching

The teachers in many of these studies benefit from sustained support in learning how to use assessment to inform teaching (Doppelt, 2003; Wiliam et al., 2004). Even experienced teachers can be surprised by student’s misunderstandings (Bell & Cowie, 2001; Thissen-Roe

et al., 2004), and in learning formative strategies, teachers are better able to use assessment information (Bell & Cowie, 2001; Hand & Prain, 2002). Teachers in these studies draw on a variety of assessment sources to inform their teaching, from students' responses to oral questions (Bell & Cowie, 2001; Black & Harrison 2001), student interaction with computer simulations (Vendlinski & Stevens, 2002), discussion in group problem solving (Leat & Nichols, 2002), individual products (Hermann & Lewis, 2004), and portfolios (Barootchi & Keshavarz, 2002). The teachers in Bell and Cowie's (2001) study describe different types of action under the umbrella of formative assessment, from proactive and planned to reactive and spontaneous, and they list a host of ways in which assessment can support a range of teaching activities, from planning to reporting. The teachers' experiences in Bell and Cowie's (2001) study are reflected in many of the classrooms that are described in these articles.

Pedagogical change is emphasized, highlighting the possibility of responding to the needs of an individual learner (Nunes, 2004) or a group (Thissen-Roe, et al. 2004), adjust unit plans (Hand & Prain, 2002) or shift curricular goals (Barootchi & Keshavarz, 2002; Dori, 2003). Again, the consequences are generally portrayed as positive. Assessment information provided by students can be "invaluable" (Nunes, 2004, p. 333) for teachers, and it can be used intentionally to improve the relevance and effectiveness of instruction (Vendlinski & Stevens, 2002). Improvements in student learning are linked to greater use of assessment information by teachers (Barootchi & Keshavarz, 2002; Dori, 2003), and improvements in student engagement are also suggested as teachers are able to "design future instructional strategies, materials and activities that are more meaningful and valuable to the learners" (Nunes, 2004, p. 333).

Student Involvement in Assessment

The most frequently mentioned aspect of formative assessment in these texts is student involvement. The articles involve student formative assessment using portfolio assessments (Barootchi & Keshavarz, 2002; Clark et al., 2001; Nunes, 2004; Simon & Forgette-Giroux, 2000; Torres Pereira de Eca, 2005), or they focus more specifically on peer and student self-assessment (Black & Harrison, 2001; Davies et al., 2004; McDonald, 2002; McDonald & Boud, 2003; Noonan & Duncan, 2005). In some, peers support or mediate the learning and assessment process (Bell & Cowie, 2001; Cowie, 2005; Hickey & Zuicker, 2005; Kirkwood, 2000; Wiliam et al., 2004; Yung, 2001), and in others, student self-assessment plays a strong role in the learning and assessment activities (Brookhart, 2001; Hand & Prain, 2002; Hermann & Lewis, 2004). The importance of student involvement in assessment is also suggested in some of the texts that take a broader look at the context in which classroom assessment occurs (Dori, 2003; Hayward et al., 2004; Stokking et al., 2004). Student is involved in their own assessment in two different ways in these studies. The first is a reflective process with a retrospective orientation. By looking back and reflecting on past efforts, change becomes visible for learners (Bell & Cowie, 2001; Clark et al., 2001; Hand & Prain, 2002), and they become aware of themselves as learners (Barootchi & Keshavarz, 2002; Brookhart, 2001; Nunes, 2004). The second way students are involved in their own assessment can be “triggered” (Clark et al., 2001, p. 221) by the first, but it is more action oriented. Learners “take charge of their own learning” (Barootchi & Keshavarz, 2002, p.286), and take responsibility for learning (Davies et al., 2004; Hayward et al., 2004; Kirkwood, 2000; McDonald and Boud, 2003), especially when they are able to make

decisions about, or have choices in the process (Bell & Cowie, 2001; Clark et al., 2001; Dori, 2003; Nunes, 2004; Torres Pereira de Eca, 2005). Students may also be involved in the assessment of others, and many of these studies they are seen to play supportive roles in classroom assessment processes, either as individuals or in group activities. Although this role of peer interaction appears to influence classroom formative assessments, it is beyond the scope of this research project to adequately examine the many facets that peer interaction can play in a student's academic achievement. In some of the classrooms, a supportive peer culture seems to emerge as a consequence of the opportunity for interaction between students (Cowie, 2005; Kirkwood, 2000). In other studies, student activities are organized to encouraged students to give each other feedback in small groups. Although peer assessment does not appear to be used to a great extent (Noonan & Duncan, 2005; Stokking et al., 2004), it seems to be valued as a means of assessment for learning. For example, the teachers who are given voice by Black and Harrison (2001) described several strategies that involve students in small group discussion about their work, and collaborative groups are purposefully established for "feedback conversations" (p. 295) in the study by Hickey and Zuicker (2005). The importance of a supportive peer group for some students is clearly illustrated by the experience of one student in the study by Clark and colleagues (2001). They note that "without this group, she floundered and lost her motivation to write and even her skillfulness as a writer" (p. 230). Concern about the consequences of withdrawing peer support is also expressed by one of the teachers in a pilot project in Scotland. Although the results are ultimately positive in both these cases, the relationship between fostering autonomy and encouraging constructive collaboration is not well explored in this body of

work. Although the benefits of allowing students to “confer in groups” (Hand & Prain, 2002; p. 748; also Yung, 2001) is clearly appreciated by some teachers, and the merits of peers as an authentic audience are noted (Dori, 2003; Hand & Prain, 2002), the risks involved are not often addressed. The role of student disclosure, and the need for trust and respect in classroom dynamics, is repeatedly stressed by Cowie (2005), but in most of these articles there is very little mention of the potentially negative consequences of peer assessment. Our own experience as teachers tells us that classroom assessment is not a private endeavor, and that a supportive learning environment must be actively nurtured. Several of these studies capitalize on the public nature of classroom learning in their assessment strategies (Hand & Prain, 2002; Hickey & Zucker, 2005), but this is also an area that is not well illuminated. There are a host of contextual factors that can impede or facilitate the involvement of students in assessment for learning, such as time limitations due to curricular requirements (Hayward et al., 2004), the familiarity of students with formative practices (Hermann & Lewis, 2004; Nunes, 2005; Torres Pereira de Eca, 2005), and teachers’ ideas about its feasibility or value (Hayward et al., 2004; Noonan & Duncan, 2005; Stokking et al., 2004). While the importance of interaction between students, and between the teacher and students is highlighted through many of these articles (Barootchi & Keshavarz, 2002; Bell & Cowie, 2001; Cowie, 2005; Hermann & Lewis, 2004; Nunes, 2004; Wiliam et al., 2004), there is also strong evidence that training is needed for assessment to be constructive. It is suggested that students need to understand the rationale for self-reflection (Kirkwood, 2000), and that they need to be prompted to elaborate on their thoughts (Hermann & Lewis, 2004). In one study, the teachers use a variety of tools and strategies, from checklists to writing

prompts, to support students' self-reflection (Simon & Forgette-Giroux, 2000), but it is not clear which of these is more effective, and in what circumstances. Making the formative process explicit for students does; however, seem to be a key factor, even in the varied contexts of these studies. Torres Pereira de Eca (2005), for example, found that art students who had more experience with critical self-reflection were more successful with a new portfolio assessment system, and Hickey & Zuicker (2005) had more success when the use of feedback was modeled by the teacher in an introductory genetics class. A large-scale program to train students in self-assessment across the curriculum in Barbados had a strong positive impact, not only on academic achievement, but also on students' attitudes about self-assessment (McDonald & Boud, 2003). The students in this study reported feeling more independent and empowered (p.215), and they were more confident in preparing for examinations and setting goals for the future.

Learning Goals and Assessment Criteria

More than half of the articles reviewed mention learning goals or assessment criteria. Goal setting by students is discussed as a necessary, or prerequisite part of assessment for learning (Barootchi & Keshavarz, 2002; Black & Harrison, 2001b), and teachers' awareness of students' goals can enhance student engagement (Clark et al., 2001). However, Brookhart (2001) found that students' goals are not necessarily tied to specific learning targets, and they relate more generally to the improvement that students believe they need for future achievement. The use of clearly specified goals or criteria may also be more appropriate in certain subjects than others. For example, in the English lesson described by Hodgen and Marshall (2005), the teacher has students compare two versions of a text (Henry V) as a

means of sharing the criteria (p. 161) in an assignment that ultimately asks students not only to critique existing literary productions, but also to generate their own ideas for production. Hodgen and Marshall (2005) write that unlike the mathematics lesson they also analyze, the learning process in the English lesson involves “heading more towards a horizon than a tightly defined goal (p. 166). Although this teacher exhibits considerable skill in apprenticing students into the guild knowledge (p. 172) of the discipline, the criteria for success with the assignment were implied, but sometimes unspoken (p. 163). Implicit criteria is also seen in Yung’s (2001) study of Biology teachers in Hong Kong, where one teacher repeatedly warns students that the amount of help they ask for during practical work may affect their evaluation. As Yung (2001) notes, the teacher persists in using autonomy as evaluative criteria, even though it conflicts with the learning purpose of the assessment, because of her desire to be ‘fair’ in the process. Similar conflicts or inconsistencies are also seen in a survey of teachers’ practices done by Stokking and colleagues (2004) in the Netherlands. Not only was there a mismatch between learning goals and criteria, in some cases the “assessment criteria were lacking or were not sufficiently explicit” (p. 109) for use by students. This is unfortunate as positive consequences are associated with the use of explicit criteria in several of the studies reviewed here (Barootchi & Keshavarz, 2002; Davies et al., 2004; Hermann & Lewis, 2004), and a variety of methods are used to focus student learning, including criteria in rubric format (Hand & Prain, 2002; Hermann & Lewis, 2004) and exemplars of student work (William et al., 2004).

Varied Approach to Elicit Learning

This final aspect of formative assessment is not one that is explicitly discussed with

great frequency in the articles reviewed. Those that do mention the use of varied approaches to elicit learning are often studies where the teachers were not strictly bound by the terms of the research, but were encouraged to develop a range of strategies following some form of professional development with the researchers (Bell & Cowie, 2001; Hand & Prain, 2002; Simon & Forgette-Giroux, 2000; Wiliam et al., 2004). There are also descriptions of classroom projects that include multiple methods of eliciting learning (Hermann & Lewis, 2004; Kirkwood, 2000), and some studies where different approaches are compared in more experimental-type research (Dori, 2003; Schwartz & Martin, 2004). In general, the use of varied methods has a positive effect on the students involved in these studies. For example, the alternative embedded assessments used in the experimental schools in the Matriculation 2000 project described by Dori (2003) included a wider variety of tasks and activities than the traditional assessments used in the control schools. Students in the experimental groups participated in laboratory activities, group projects, and individual self-assessment, and their exposure to higher-order activities had a positive impact on their attitudes, as well as on their choice of more challenging summative tasks. Teachers in some of the studies analyzed also benefit from the use of varied assessment methods. Hand and Prain (2002) note that the “participant teachers believed that their repertoire of assessment strategies had been greatly enhanced by their learning from the program, and that their students had reacted very positively to more diverse assignments and classroom procedures” (p. 752). While it is evident that some of the teachers in these studies are using a variety of strategies (e.g. Wiliam et al., 2004), it is hard to speculate from this body of work to what extent the assessment approaches in most classroom are varied. One also does not know a

great deal about which of these varied approaches might be more effective in which contexts. The different questioning strategies proposed by the teachers in Black & Harrison (2001), or the variety of self-reflection tools used by the teachers in Simon & Forgette-Giroux (2000) provide an array of choices. Additionally, innovative methods that go beyond traditional measures of achievement, such as the problem-solving assessment with an embedded learning resource (Schwartz and Martin, 2004), open possibilities for classroom assessment practice that could bear further empirical and comparative study.

Chapter 3

Methodology

This quantitative study examined to see whether comprehensive post-formative assessments can accurately predict student academic achievement on AYP indicators as measured by standard criterion referenced tests. The predictability of formative assessment scores were analyzed to express if a correlation existed between formative and summative achievement data. The purposes of chapter 3 are to describe the: (a) sample population selected for this study; (b) instruments that were administered for data collection; (c) methods, materials and procedures utilized to implement and collect the data for the study; and (d) selection and use of statistical procedures employed in the analysis of the collected data.

Basic Research Design

This study was designed to examine whether comprehensive post-formative assessments can accurately predict student academic achievement in math on AYP indicators as measured by standardized criterion-referenced tests. Encompassing a three year period of formative assessments, data was collected for over 2,900 middle school students. For the three years worth of scores, the following statistical processes were performed: means, standard deviations, frequencies, percentages of pass/fail ratios, and finally, a logistic regression was performed for every grade, academic year, and quarterly post-test event. The logistic regression examined the Beta co-efficient, the standard error of measure, the statistical significance, and the exponential value of predictability. The exponential value of predictability coupled with the statistical significance enabled the

research model to successfully predict 84.87% of student achievement outcomes as discussed in Chapter 5.

Research question 1

Do comprehensive post-formative assessments accurately predict student academic achievement in math on AYP indicators as measured by standardized criterion-referenced tests scores in each of three quarters (1st quarter Standard assessment post scores, 2nd quarter Standard assessment post scores, 3rd quarter Standard assessment post scores) predict CRCT (pass vs. fail) in year 1?

Research question 2

Do the comprehensive post-formative assessments in each of three quarters (1st quarter Standard assessment post, 2nd quarter Standard assessment post scores, 3rd quarter Standard assessment post scores) predict CRCT (pass vs. fail) in year 2?

Research question 3

Do the comprehensive post-formative assessments in each of three quarters (1st quarter Standard assessment post scores, 2nd quarter Standard assessment post scores, 3rd quarter Standard assessment post scores) predict CRCT (pass vs. fail) in year 3?

The null hypotheses to be examined are:

Null Hypothesis 1

The comprehensive post-formative assessments in each of three quarters (1st quarter Standard assessment post scores, 2nd quarter Standard assessment post scores, 3rd quarter Standard assessment post scores) do not predict CRCT (pass vs. fail) in year 1.

Null Hypothesis 2

The comprehensive post-formative assessments in each of three quarters (1st quarter Standard assessment post scores, 2nd quarter Standard assessment post scores, 3rd quarter Standard assessment post scores) do not predict CRCT (pass vs. fail) in year 2.

Null Hypothesis 3

The comprehensive post-formative assessments in each of three quarters (1st quarter Standard assessment post scores, 2nd quarter Standard assessment post scores, 3rd quarter Standard assessment post scores) do not predict CRCT (pass vs. fail) in year 3.

Preliminary Procedures

Prior to the implementation of this study, a thorough review of literature was completed. The review of literature focused on characteristics of school reform, the need to develop best practices, a brief history of standards based curricula, and current research on formative assessments.

The Gainesville City Board of Education granted the researcher permission to use the CRCT and post assessment data from 2004-2007 and to use Gainesville Middle School's and Gainesville City System's name in this dissertation. Since the researcher used no personal identifiers, getting parental permission on the subjects was not necessary.

Selection of the Sample

Sixth, seventh and eight graders who were administered the Georgia state mandated annual assessment, Criterion Referenced Competency Test (CRCT), and the quarterly post assessments at Gainesville Middle School from 2004-2007 were the

subjects of this study. Below is a chart illustrating the ethnic and gender composition of the subject population:

	White	Black	Hispanic	Asian	Multi
Female	148	235	437	27	15
Male	184	197	504	31	23

Historically, Gainesville Middle School has been a predominantly white and African-American school. However, over the last 13 years, there has been a tremendous influx of Hispanic students due to the employment opportunities provided by the poultry industry of the area. Hall County is considered “the poultry capitol of the world.”

Additionally, the Gainesville City School System presents a conspicuous dichotomy on a socioeconomic level. Within the city limits served by the system is the Chattahoochee Country Club, an area consisting of homes valued at one million dollars and up. Yet also within these city boundaries are some of the most economically disadvantaged residents of the entire area including a number of homeless families. Interestingly, the number of families within the system that would be considered middle income is comparatively low.

According to the U.S. Census Bureau, the City of Gainesville has a population of 32,444. Of these, 21.8% of those are below the poverty line, compared with 12.4% for surrounding Hall County, 13.4% for the State of Georgia, and 12.4% for the United States as a whole. The median household income for the City is \$36,605.

Instruments Used in the Data Collection

Criterion-Referenced Competency Test

This study utilized the Criterion Referenced Competency Test (CRCT) required by the Department of Education of the State of Georgia and the quarterly post assessments given by Gainesville Middle School.

Georgia's mandated annual assessment is called the CRCT (Criterion Referenced Competency Test) which measures how well students acquire the skills and knowledge described in the Georgia Performance Standards (GPS) and the Quality Core Curriculum (QCC). The assessments yield information on academic achievement at the student, class, school, system, and state levels. This information is used to diagnose the student's strengths and weaknesses as related to the GPS and QCC and to gauge the quality of education throughout Georgia. In the State of Georgia, all public school students in grades K-8 must take this yearly comprehensive criterion-referenced exam. For students in the 3rd, 5th, and 8th grades, they must pass the mathematics and reading portions of the test. If they do not pass the CRCT, they may be retained. School districts in Georgia also administer the CRCT to the other grade levels but only these three grade levels use the test as a gateway for promotion. Since only the mathematics and reading content domains are used by the United States Department of Education as the AYP indicators to evaluate school system effectiveness, the researcher examined only the mathematics scores on the CRCT.

The CRCT is designed to measure how well students acquire the skills and knowledge described in the Georgia Performance Standards (GPS) and the Quality Core

Curriculum (QCC). The assessments yield information on academic achievement at the student, class, school, system, and state levels. This information is used to diagnose individual student strengths and weaknesses as related to the instruction of the GPS/QCC, and to gauge the quality of education throughout Georgia.

Reliability is one of the two cornerstones of technical quality in testing and measurement. In simple terms, it asks will the same measurement give the same or comparable result for the same student every time. Reliability is evaluated by statistical methods. For the 2004 CRCT, total test reliabilities ranged from 0.79 to 0.86 for Reading, 0.85 to 0.89 for English / Language Arts, 0.87 to 0.91 for Mathematics, 0.89 to 0.90 for Science, and 0.88 to 0.91 for Social Studies. The other cornerstone of technical quality in testing is validity, which begins with the purpose of the assessment and continues through item writing and review. All CRCT items are written by qualified, professional content specialists specifically for the Georgia CRCT. After the items are written, curriculum specialists and committees of Georgia educators review the items. Items are evaluated for overall quality and clarity, content coverage and appropriateness, alignment to the curriculum, and grade appropriate stimuli with an emphasis on higher order thinking skills. In addition, there should be one clear correct answer with appropriate, relevant, and reasonable distractors. Items should be free from bias toward or against any particular group. Great care is taken throughout the item-development process to monitor items for potential bias and to ensure representation of all of Georgia's students. To ensure that the CRCT meet the highest standards of technical quality and defensibility, the Testing Division meets with an independent panel of experts – Georgia's Technical

Advisory Committee (TAC) – on a quarterly basis. TAC members are experts in the field of educational measurement who review all aspects of the test development and implementation process on a continual basis. The Georgia Department of Education is confident that the CRCT are both reliable and valid. Georgia stakeholders can have the highest confidence in the CRCT program. (Georgia Dept of Education, 2008)

Gainesville Middle School focuses its instructional process upon the Georgia Performance Standards and has developed pacing guides for each grade level in mathematics. These pacing guides are then divided into quarters and given to each teacher. The system administers quarterly pre and post-formative assessments in order to optimize instructional time and resources. Pre-tests are administered at the beginning of each quarter in order to determine the students' level of mastery in a given content area. With formalized summative assessments garnering most of the focus of the private and public sectors, formative assessments appear to be the vital link to student achievement that is often overlooked. They are overlooked because their diagnostic characteristics are not used to assess *for* instruction vis-à-vis curriculum mapping, pacing, and scope and sequencing but rather, they have been relegated to an assessment *of* post instruction.

Formative Assessments

The formative pre and post assessments are developed collaboratively among content area teachers and content literacy coaches. A 70,000 question bank is accessed and items are chosen based upon performance standard correlation and a content validity measurement assigned by the providing vendor. Assessments are multiple-choice by

design and typically have 25-30 questions each but comprehensive in relation to the standards being measured for that particular quarter. Student performance is then compiled and analyzed via the vendor's software (Testgate), and teachers along with instructional leaders are able to plan instructional units in relation to the students' level of mastery.

To assure item validity in Testgate, a team of content experts led by a psychometrician has reviewed the correlation between each item and its designated curriculum standard. A description of the alignment process is provided in response to Question Two, below. As new items are added, they too are reviewed by the psychometrician and their team. Thinkgate also provides item difficulty data (p-values) for each item. If a value of .2 is assigned, fewer students correctly answered the item. If a value of .9 is assigned, more students correctly answered the item.

In order to meet client needs efficiently, and provide them with quality items from which to build assessments, Thinkgate is continually expanding its bank of items by utilizing its vast base of subscribers to add questions and then facilitating a content-based item mapping validation process. The following information details the review and mapping process and who are involved in the process.

At the beginning of the item review process, Thinkgate had over 35,000 items in its bank of items. Each of the items were aligned to one state content standard - the original standard to which it was written which in the majority of cases was a Georgia GPS or QCC standard. The purpose of the review process was to first evaluate the

question's face validity and second to verify authenticity of the original mapping. Thus, at the end of the 2-phase process, the 35,000+ items were individually reviewed for content validity and standards alignment.

Using Thinkgate's online system, a content expert first reviews an item to evaluate its quality and determine whether it should remain in the active bank of items for subscribers' use. A reviewer could select one determination per item - approved, reject, or reject for revision. Reviewers were allowed to make minor grammatical edits as part of their review. Items that needed involved revisions and edits were classified as 'reject for revision'. The reviewers also assured that stimuli or addenda (e.g., passage, table, graphic) appropriately matched its associated item(s). Grade level appropriateness and reading level were judged as either appropriate or inappropriate using the state content standard as a guide.

When a reviewer 'approved' the quality of an item, that item then moved to the second phase of the mapping process. The same reviewer then compared the content being measured by each item to the Georgia GPS state content standards. (The standards were uploaded to Thinkgate's system for electronic access by the reviewers. This access also assured that all of the reviewers were using the latest version of the state content standards.) Using the reviewers' professional judgment and content expertise, each item was individually approved or remapped to the appropriate GPS standard. None of the items were 'shoe-horned' into mapping to a state standard as the guiding factor throughout the process was the valid mapping of quality items to state content standards. If we could

not accurately map an item, then that item wasn't mapped for a particular set of state standards.

Because the process is dependent on the reviewers' understanding of the content measured by an item, each reviewer went through a strict hiring review evaluation. In addition to having the necessary content knowledge, grade level experience was also considered in the selection of the reviewers.

To date, approximately 30 content specialists have been involved in mapping the 43,000+ items across multiple state content standards. Several of the reviewers had been involved in mapping activities previously, most had experience in item writing or item reviewing, and all were knowledgeable with the specific grade level and content area to which they were assigned to map questions.

To further ensure equitable representation, the reviewers had teaching or education experience in nine different states. The diversity in perspective was important for a variety of reasons, 1) familiarity with different state content standards, 2) ability to quickly digest and apply state content standards, 3) diversity in teaching experience and perspective, and 4) exposure to different assessments and test questions.

Procedures

The researcher identified the sixth, seventh, and eighth grade students that were enrolled in Gainesville Middle School from 2004-2007. Using the district's student information database, the researcher filtered the data in order to extrapolate only students

who had valid test scores for the academic years pertinent to the study. The students' CRCT scores in mathematics as well as their three quarterly math post assessment scores were collected and entered into an excel spreadsheet.

The students' personal data was protected by deleting all identifying test identification numbers, names, and classroom assignments.

Analysis of Data

Data were entered into SPSS 15.0 for Windows. Descriptive statistics were conducted on the data. To examine the hypothesis, nine binary logistic regressions for grades 6th, 7th and 8th were conducted to assess if the school years 2004-2005, 2005-2006, and 2006-2007 post test scores can serve as independent predictor variables in relation to CRCT (pass vs. fail) as the dependant criterion variable.

To examine hypotheses 1a-c, three binary logistic regressions for grades (6th, 7th and 8th) will be conducted to assess if school year 2004-2005 post test scores as the independent predictor variables and CRCT in (pass vs. fail) as the dependant criterion variable.

To examine hypotheses 2a-c, three binary logistic regressions for grades (6th, 7th and 8th) will be conducted to assess if school year 2005-2006 post test scores as the independent predictor variables and CRCT in (pass vs. fail) as the dependant criterion variable.

To examine hypotheses 3a-c, three binary logistic regressions for grades (6th, 7th and 8th) will be conducted to assess if school year 2006-2007 post test score as the

independent predictor variables and CRCT in (pass vs. fail) as the dependant criterion variable.

Binary logistic regressions were chosen as the statistical tool of analysis because of the categorical, dichotomous nature of the criterion variable. In binary logistic regression we are testing the odds in which the predictors can predict the probability of the criterion variable (1=Pass, 0 =Fail). The assumptions of binary logistic regression, the criterion variable is dichotomous, as well as independent and mutually exclusive, were met. In addition, a large sample size, it is suggested a minimum of 20 cases per predictor, is required to obtain ideal and accurate results from binary logistic regression which met favorably with 2900 valid participants.

Summary

This chapter has explained methods used for this descriptive quantitative study in the researcher's attempt to examine whether comprehensive post-formative assessments can accurately predict student academic achievement on AYP indicators as measured by standard criterion referenced tests. The next chapter presents the results that were obtained from the methods used and the need for further research.

Chapter 4

Research Findings

This study was designed to examine if comprehensive post-formative assessments can accurately predict student academic achievement in math on AYP (Adequate Yearly Progress) indicators as measured by standardized criterion-referenced tests.

Research question 1a

For 6th grade students, do the post test scores in each of the three quarters (1st quarter standard assessment post scores, 2nd quarter standard assessment post scores, 3rd quarter Standard Assessment post scores) predict CRCT (pass vs. fail) in 2004-2005?

Research question 1b

For 7th grade students, do the post test scores in each of the three quarters (1st quarter standard assessment post scores, 2nd quarter standard assessment post scores, 3rd quarter standard assessment post scores) predict CRCT (pass vs. fail) in 2004-2005?

Research question 1c

For 8th grade students, do the post test scores in each of the three quarters (1st quarter standard assessment post scores, 2nd quarter standard assessment post scores, 3rd quarter standard assessment post scores) predict CRCT (pass vs. fail) in 2004-2005?

Research question 2a

For 6th grade students, do the post test scores in each of the three quarters (1st quarter standard assessment post scores, 2nd quarter standard assessment post scores, 3rd quarter standard assessment post scores) predict CRCT (pass vs. fail) in 2005-2006?

Research question 2b

For 7th grade students, do the post test scores in each of the three quarters (1st quarter standard assessment post scores, 2nd quarter standard assessment post scores, 3rd quarter standard assessment post scores) predict CRCT (pass vs. fail) in 2005-2006?

Research question 2c

For 8th grade students, do the post test scores in each of the three quarters (1st quarter standard assessment post scores, 2nd quarter standard assessment post scores, 3rd quarter standard assessment post scores) predict CRCT (pass vs. fail) in 2005-2006?

Research question 3a

For 6th grade students, do the post test scores in each of the three quarters (1st quarter standard assessment post scores, 2nd quarter standard assessment post scores, 3rd quarter standard assessment post scores) predict CRCT (pass vs. fail) in 2006-2007?

Research question 3b

For 7th grade students, do the post test scores in each of the three quarters (1st quarter standard assessment post scores, 2nd quarter standard assessment post scores, 3rd quarter standard assessment post scores) predict CRCT (pass vs. fail) in 2006-2007?

Research question 3c

For 8th grade students, do the post test scores in each of the three quarters (1st quarter standard assessment post scores, 2nd quarter standard assessment post scores, 3rd quarter standard assessment post scores) predict CRCT (pass vs. fail) in 2006-2007?

Null Hypotheses

Research question 1a

For 6th grade students, post test scores in each of the three quarters (1st quarter standard assessment post scores, 2nd quarter standard assessment post scores, 3rd quarter standard assessment post scores) do not predict CRCT (pass vs. fail) in 2004-2005?

Research question 1b

For 7th grade students, post test scores in each of the three quarters (1st quarter standard assessment post scores, 2nd quarter standard assessment post scores, 3rd quarter standard assessment post scores) do not predict CRCT (pass vs. fail) in 2004-2005?

Research question 1c

For 8th grade students, post test scores in each of the three quarters (1st quarter standard assessment post scores, 2nd quarter standard assessment post scores, 3rd quarter standard assessment post scores) do not predict CRCT (pass vs. fail) in 2004-2005?

Research question 2a

For 6th grade students, post test scores in each of the three quarters (1st quarter standard assessment post scores, 2nd quarter standard assessment post scores, 3rd quarter standard assessment post scores) do not predict CRCT (pass vs. fail) in 2005-2006?

Research question 2b

For 7th grade students, post test scores in each of the three quarters (1st quarter standard assessment post scores, 2nd quarter standard assessment post scores, 3rd quarter standard assessment post scores) do not predict CRCT (pass vs. fail) in 2005-2006?

Research question 2c

For 8th grade students, post test scores in each of the three quarters (1st quarter standard assessment post scores, 2nd quarter standard assessment post scores, 3rd quarter standard assessment post scores) do not predict CRCT (pass vs. fail) in 2005-2006?

Research question 3a

For 6th grade students, do the post test scores in each of the three quarters (1st quarter standard assessment post scores, 2nd quarter standard assessment post scores, 3rd quarter standard assessment post scores) predict CRCT (pass vs. fail) in 2006-2007?

Research question 3b

For 7th grade students, post test scores in each of the three quarters (1st quarter standard assessment post scores, 2nd quarter standard assessment post scores, 3rd quarter standard assessment post scores) do not predict CRCT (pass vs. fail) in 2006-2007?

Research question 3c

For 8th grade students, post test scores in each of the three quarters (1st quarter standard assessment post scores, 2nd quarter standard assessment post scores, 3rd quarter standard assessment post scores) do not predict CRCT (pass vs. fail) in 2006-2007?

Data Analysis

Data were entered into SPSS version 15.0 for Windows. Descriptive statistics were conducted on demographic data which included frequency and percentages for nominal (categorical/dichotomous) data and means/standard deviations for continuous (interval/ratio) data. Standard deviation measures statistical dispersion, or the spread of values in a data set. If the data points are all close to the mean, then the standard

deviation is close to zero. The arithmetic mean is defined as the sum of scores divided by the number of scores.

To examine hypotheses 1a-c, three binary logistic regressions for grades (6th, 7th and 8th) will be conducted to assess if school year 2004-2005 post test scores as the independent predictor variables and CRCT in (pass vs. fail) as the dependant criterion variable.

To examine hypotheses 2a-c, three binary logistic regressions for grades (6th, 7th and 8th) will be conducted to assess if school year 2005-2006 post test scores as the independent predictor variables and CRCT in (pass vs. fail) as the dependant criterion variable.

To examine hypotheses 3a-c, three binary logistic regressions for grades (6th, 7th and 8th) will be conducted to assess if school year 2006-2007 post test score as the independent predictor variables and CRCT in (pass vs. fail) as the dependant criterion variable.

Binary logistic regressions were chosen as the statistical tool of analysis because of the categorical, dichotomous nature of the criterion variable. In binary logistic regression we are testing the odds in which the predictors can predict the probability of the criterion variable (1=Pass, 0 =Fail). The assumptions of binary logistic regression, the criterion variable is dichotomous, as well as independent and mutually exclusive, were met. In addition, a large sample size, it is suggested a minimum of 20 cases per predictor, is required to obtain ideal and accurate results from binary logistic regression which met favorably with 2900 valid participants.

Results

Means, standard deviations and ranges for CRCT scores by grade in school year 2004-2005 are presented Table 1. Table 2 presents the frequencies and percents for pass vs. fail by grade in school year 2004-2005. Means, standard deviations and ranges for CRCT scores by grade 2005-2006 are presented Table 3. Table 4 presents the frequencies and percents for pass vs. fail by grade in 2005-2006. Means, standard deviations and ranges for CRCT scores by grade in school year 2006-2007 are presented Table 5. Table 6 presents the frequencies and percents for pass vs. fail by grade in school year 2006-2007.

Table 1
Means, Standard Deviations and Ranges for CRCT Scores by Grade in 2004-2005

2004-2005	Min	Max	M	SD
6 th CRCT	234	450	326.49	39.45
7 th CRCT	261	411	317.92	23.53
8 th CRCT	236	373	300.80	24.19

Table 2
Frequencies and Percents for Pass vs. Fail by Grade in 2004-2005

2004-2005	Pass		Fail	
	Frequencies	Percents	Frequencies	Percents
6 th	232	73.2	85	26.8
7 th	182	80.9	43	19.1
8 th	97	52.7	87	47.3

Table 3
Means, Standard Deviations and Ranges for CRCT Scores by Grade in 2005-2006

2005-2006	Min	Max	M	SD
6 th CRCT	332	920.00	806.50	46.17
7 th CRCT	263	824.00	325.24	47.19
8 th CRCT	233	407.00	315.09	33.05

Table 4
Frequencies and Percents for Pass vs. Fail by Grade in 2005-2006

2005-2006	Pass		Fail	
	Frequencies	Percents	Frequencies	Percents
6 th	192	29.0	127	39.8
7 th	292	76.0	92	24.0
8 th	236	66.8	112	33.2

Table 5
Means, Standard Deviations and Ranges for CRCT Scores by Grade in 2006-2007

2006-2007	Min	Max	M	SD
6 th CRCT	748	898	806.69	28.69
7 th CRCT	278	950	800.10	118.45
8 th CRCT	258	437	327.25	30.88

Table 6
Frequencies and Percents for Pass vs. Fail by Grade in 2006-2007

2006-2007	Pass		Fail	
	Frequencies	Percents	Frequencies	Percents
6 th	235	55.0	192	45.0
7 th	251	69.9	108	30.1
8 th	282	83.7	55	16.3

Hypothesis 1a

A binary logistic regression was conducted to examine the three 6th grade post test quarter scores of 2004-2005 predicted CRCT (pass vs. fail). Results of the regression indicate a significant model such that scores do predict CRCT (pass vs. fail), $\chi^2(3) = 83.70$, $p < .001$. The model (i.e., the three predictors) fit the data well, shown by the Nagelkerke $R^2 = 45.0\%$ of CRCT (pass vs. fail) variance. Overall, the model correctly predicts 84.8% of the participant outcomes in the sample. Table 7 presents the beta coefficients where the first and second quarter predicted pass-fail outcomes: for every one unit increase in the quarter scores, participants were 136.242 and 62.4 times more likely to pass than fail the CRCT, respectively.

Table 7
Logistic Regression on 6th Grade Posttest Scores for 2004-2005 predicting CRCT (Pass vs. Fail)

Predictor	β	SE	Sig.	Exp(B)
First Quarter	4.914	1.639	.003	136.242
Second Quarter	4.134	1.658	.013	62.407
Third Quarter	1.944	1.296	.134	6.987

Hypothesis 1b

A binary logistic regression was conducted to examine the three 7th grade post test quarter scores of 2004-2005 predicted CRCT (pass vs. fail). Results of the regression indicate a significant model such that scores do predict CRCT (pass vs. fail), $\chi^2(3) = 41.44$, $p < .001$. The model (i.e., the three predictors) fit the data well, shown by the Nagelkerke $R^2 = 42.3\%$ of CRCT (pass vs. fail) variance. Overall, the model correctly predicts 89.5% of the participant outcomes in the sample. Table 8 presents the beta coefficients where the first quarter predicted pass-fail outcomes: for every one unit increase in the first quarter scores, participants were 808.95 times more likely to pass than fail the CRCT.

Table 8

Logistic Regression on 7th Grade Posttest Scores for 2004-2005 predicting CRCT (Pass vs. Fail)

Predictor	β	SE	Sig.	Exp(B)
First Quarter	6.696	2.224	.003	808.945
Second Quarter	2.170	1.950	.266	8.754
Third Quarter	2.869	1.739	.099	17.613

Hypothesis 1c

A binary logistic regression was conducted to examine the three 8th grade post test quarter scores of 2004-2005 predicted CRCT (pass vs. fail). Results of the regression indicate a significant model such that scores do predict CRCT (pass vs. fail), $\chi^2(3) = 80.48$, $p < .001$. The model (i.e., the three predictors) fit the data well, shown by the Nagelkerke $R^2 = 44.9\%$ of CRCT (pass vs. fail) variance. Overall, the model correctly

predicts 81.5% of the participant outcomes in the sample. Table 9 presents the beta coefficients where the first, second, and third quarters predicted pass-fail outcomes: for every one unit increase in the quarter scores, participants were 1230.34, 187.03 and 1295.10 times more likely to pass than fail the CRCT, respectively.

Table 9
Logistic Regression on 8th Grade Posttest Scores for 2004-2005 predicting CRCT (Pass vs. Fail)

Predictor	β	SE	Sig.	Exp(B)
First Quarter	5.440	2.170	.012	230.340
Second Quarter	5.231	2.285	.022	187.033
Third Quarter	7.166	1.913	.000	1295.102

Hypothesis 2a

A binary logistic regression was conducted to examine the three 6th grade post test quarter scores of 2005-2006 predicted CRCT (pass vs. fail). Results of the regression indicate a significant model such that scores do predict CRCT (pass vs. fail), $\chi^2(3) = 115.86, p < .001$. The model (i.e., the three predictors) fit the data well, shown by the Nagelkerke $R^2 = 36.5\%$ of CRCT (pass vs. fail) variance. Overall, the model correctly predicts 81.2% of the participant outcomes in the sample. Table 10 presents the beta coefficients where the first and third quarter predicted pass-fail outcomes: for every one unit increase in the quarter scores, participants were 1890.90 and 38.03 times more likely to pass than fail the CRCT, respectively.

Table 10
Logistic Regression on 6th Grade Posttest Scores for 2005-2006 predicting CRCT (Pass vs. Fail)

Predictor	β	SE	Sig.	Exp(B)
First Quarter	7.544	1.218	.000	1890.147

Second Quarter	1.068	1.275	.402	2.909
Third Quarter	3.638	1.528	.017	38.032

Hypothesis 2b

A binary logistic regression was conducted to examine the three 7th grade post test quarter scores of 2005-2006 predicted CRCT (pass vs. fail). Results of the regression indicate a significant model such that scores do predict CRCT (pass vs. fail), $\chi^2(3) = 69.60, p < .001$. The model (i.e., the three predictors) fit the data well, shown by the Nagelkerke $R^2 = 24.8\%$ of CRCT (Pass vs. Fail) variance. Overall, the model correctly predicts 88.5% of the participant outcomes in the sample. Table 11 presents the beta coefficients where the first and third quarters predicted pass-fail outcomes: for every one unit increase in the quarter scores, participants were 714.53 and 230.56 times more likely to pass than fail the CRCT, respectively.

Table 11
Logistic Regression on 7th Grade Posttest Scores for 2005-2006 predicting CRCT (Pass vs. Fail)

Predictor	β	SE	Sig.	Exp(B)
First Quarter	6.572	1.736	.000	714.534
Second Quarter	1.460	1.380	.290	4.306
Third Quarter	5.441	1.716	.002	230.561

Hypothesis 2c

A binary logistic regression was conducted to examine the three 8th grade post test quarter scores of 2005-2006 predicted CRCT (pass vs. fail). Results of the regression indicate a significant model such that scores do predict CRCT (pass vs. fail), $\chi^2(3) =$

42.00, $p < .001$. The model (i.e., the three predictors) fit the data well, shown by the Nagelkerke $R^2 = 19.0\%$ of CRCT (pass vs. fail) variance. Overall, the model correctly predicts 84.4% of the participant outcomes in the sample. Table 12 presents the beta coefficients where the third quarter predicted pass-fail outcomes: for every one unit increase in the quarter scores, participants were 30.93 times more likely to pass than fail the CRCT.

Table 12
Logistic Regression on 8th Grade Posttest Scores for 2005-2006 predicting CRCT (Pass vs. Fail)

Predictor	β	SE	Sig.	Exp(B)
First Quarter	2.210	1.468	.132	9.119
Second Quarter	1.518	1.309	.246	4.564
Third Quarter	3.432	1.357	.011	30.936

Hypothesis 3a

A binary logistic regression was conducted to examine the three 6th grade post test quarter scores of 2006-2007 predicted CRCT (pass vs. fail). Results of the regression indicate a significant model such that scores do predict CRCT (pass vs. fail), $\chi^2(3) = 178.15$, $p < .001$. The model (i.e., the three predictors) fit the data well, shown by the Nagelkerke $R^2 = 41.0\%$ of CRCT (pass vs. fail) variance. Overall, the model correctly predicts 82.0% of the participant outcomes in the sample. Table 13 presents the beta coefficients where the first, second, and third quarters predicted pass-fail outcomes: for every one unit increase in the quarter scores, participants were 698.27, 103.12, and 9.24 times more likely to pass than fail the CRCT, respectively.

Table 13
Logistic Regression on 6th Grade Posttest Scores for 2006-2007 predicting CRCT (Pass vs. Fail)

Predictor	β	SE	Sig.	Exp(B)
First Quarter	6.549	1.440	.000	698.286
Second Quarter	4.636	.990	.000	103.120
Third Quarter	2.224	.990	.025	9.240

Hypothesis 3b

A binary logistic regression was conducted to examine the three 7th grade post test quarter scores of 2006-2007 predicted CRCT (pass vs. fail). Results of the regression indicate a significant model such that scores do predict CRCT (pass vs. fail), $\chi^2(3) = 122.96$, $p < .001$. The model (i.e., the three predictors) fit the data well, shown by the Nagelkerke $R^2 = 64.2\%$ of CRCT (pass vs. fail) variance. Overall, the model correctly predicts 87.5% of the participant outcomes in the sample. Table 14 presents the beta coefficients where the first and second quarter predicted pass-fail outcomes: for every one unit increase in the quarter scores, participants were 1221.50 and 19833.77 times more likely to pass than fail the CRCT, respectively.

Table 14
Logistic Regression on 7th Grade Posttest Scores for 2006-2007 predicting CRCT (Pass vs. Fail)

Predictor	β	SE	Sig.	Exp(B)
First Quarter	7.108	1.821	.000	1221.504
Second Quarter	1.541	1.859	.407	4.669
Third Quarter	9.895	2.198	.000	19833.766

Hypothesis 3c

A binary logistic regression was conducted to examine the three 8th grade post test quarter scores of 2006-2007 predicted CRCT (pass vs. fail). Results of the regression indicate a significant model such that scores do predict CRCT (pass vs. fail), $\chi^2(3) = 42.00$, $p < .001$. The model (i.e., the three predictors) fit the data well, shown by the Nagelkerke $R^2 = 29.0\%$ of CRCT (pass vs. fail) variance. Overall, the model correctly predicts 84.4% of the participant outcomes in the sample. Table 15 presents the beta coefficients where the first quarter predicted pass-fail outcomes: for every one unit increase in the quarter scores, participants were 30.93 times more likely to pass than fail the CRCT.

Table 15
Logistic Regression on 8th Grade Posttest Scores for 2006-2007 predicting CRCT (Pass vs. Fail)

Predictor	β	SE	Sig.	Exp(B)
First Quarter	2.045	2.500	.413	7.729
Second Quarter	8.340	3.482	.017	4188.377
Third Quarter	9.156	3.177	.004	9470.640

Chapter 5

Conclusions, Summary, Discussion, and Recommendations

This chapter provides a summary of the conclusions, and a discussion of the findings related to the study. In addition, recommendations are provided for further research in the area.

Conclusions

The purpose of this quantitative descriptive research study was to determine the predictability of formative assessment data as an indicator for student academic achievement on standardized summative assessments. Specifically, this study examined frequent formative assessment measures that followed a pre-assessment instrument, instruction, remediation or enrichment as needed, and subsequent post-assessment data. Multiple middle school grade levels (6th, 7th and 8th), along with multiple assessment years (2004-2005, 2005-2006, and 2006-2007) were studied. Also evaluated were three academic quarters assessment for every academic year that formative assessments were utilized. Nine research questions examined the predictability of data spanning three grade levels, three assessment periods or quarters, and three academic years and examined if available. Likewise, there was an equivalent of nine null hypotheses that attempted to negate a significant predictability of formative assessment data in relation to summative assessment achievement.

Summary of Results

A logistic regression was performed for all research questions and null hypotheses. Invariably, for every one unit increase in quarter scores (post assessment

formative assessments), a positive gain in student achievement could be predicted.

Overall, the research model correctly predicted 84.87% of participant outcomes in the samples. Also, of the possible 27 group variations, only nine of the groups displayed a statistical significance greater than .05 and were thus deemed statistically insignificant.

For 6th grade students, regardless of the year the test data were compiled, the 1st quarter scores on post-assessments was determined to be the best predictor of increased academic performance on the CRCT. For 7th grade students, regardless of the year the test data were compiled, the 1st quarter and the 3rd quarter scores on post-assessments was determined to be the best predictors of increased academic performance on the CRCT.

For 8th grade students, regardless of the year the test data was compiled, the 3rd quarter scores on post-assessments was determined to be the best predictor of increased academic performance on the CRCT. There are several reasons why the data expressed a greater degree of predictability in quarter 1 and quarter 3 results. Often, student motivation is highest at the beginning of a school year and intrinsically, they want to garner positive feedback from their teacher. Likewise, as a high-stakes testing event approaches, students tend to refocus towards their studies and recognize the significance of these standardized tests—especially since many of these exams are used as “gateway” assessments for promotion to the next grade level. There is always a great level of diversity in relation to student motivation. Further studies might focus upon the learner’s desire to achieve, their ability to display mastery on a multiple-choice assessment, and their various learning styles.

Implications

For the Gainesville City School System in Gainesville, Georgia, formative assessments represent the foundation upon which the district prides itself on raising student achievement across all student ethnic and socio-economic populations.

Gainesville's investment of both personnel and monetary resources to administer, evaluate, and maintain their complex system of pre and post assessment data represents a similar commitment that school districts across the southeastern United States have invested. The "Gainesville Model" has been replicated and modified by multiple school districts in Georgia, Florida, North Carolina, South Carolina, and Virginia. In essence, this study was an attempt to determine whether the system of frequent formative assessments could accurately predict student achievement on high-stakes assessments that determine a school system's Adequate Yearly Progress (AYP) under the provisions of *No Child Left Behind*.

Delimitations/Limitations of the Study

Only students from Gainesville Middle School composed the sample groups and only three years worth of data was utilized. Therefore, the results may not be generalized to other grade levels or to other school systems.

Recommendations for Future Practice

The purpose of the study was to determine if the formative assessment data could accurately predict student achievement on high-stakes assessments. Since there was significant statistical data reported, the school district should continue to utilize the

system of pre and post test assessment analysis in order to plan their curricular content mapping and pacing sequences for instruction.

The researcher also recommends that the school system strive to ensure that its pre and post formative assessments exhibit a high degree of test validity and maintain a strong correlation to the Georgia Performance Standards (GPS). The vendor that provides the formative assessment software has implemented a team of professionals to ensure item validity on the assessments. Test constructors should use items only proven to be valid during test composition—as explained in Chapter 3. Moreover, the school system should monitor and ensure that instructors are utilizing research-proven instructional best-practices in their classrooms in order to limit the amount of variance that students may receive between different instructors.

Recommendations for Future Research

The current study is significant because it examined the relationship between formative assessments and the predictability of student success on these assessments when related to standardized formal assessments. The study concluded that student success on standardized formal assessments can be accurately predicted when one considers the student's unit of increase and frequency of formative assessment.

However, further research in this area is still needed. Additional research should evaluate a larger sample population that includes students of various grade levels (K-5, 9-12), students from assorted school districts and/or states, and students that represent a different demographic background than those found in the Gainesville City School

System. In particular, sub-populations as measured by AYP might include special education students, English language learners, and students of poverty.

Qualitative studies are also recommended in the areas of student morale and motivation, student grades as predictors of summative assessment achievement levels, and also the importance of school administrators as quality instructional leaders. Although countless studies have already occurred in these areas, they ultimately culminate in the relationship between student and teacher which is the true epitome of teaching and learning. This is particularly significant as the development of national and state standards continues, federal and state funding levels are continually linked to assessment achievement levels, and the role of formative assessments as instructional indicators continues to expand.

As school systems, administrators, and other educational leaders seek to improve student achievement on standardized tests, these professionals realize that simply hoping students do well on these high stakes test is not a plan. Moreover, as financial and human resources focus on improving scores, these professionals want to maximize the positive impact of these resources. The results of this study exhibit the need for a more predictive way to plan for student success and to incorporate ample formative assessments that enable teachers and school administrators to leave no child behind.

References

- Allal, L. & Ducrey, G. P. (2000). Assessment of - or in- the zone of proximal development. *Learning and Instruction*, 10, 137-152.
- Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box*. Retrieved July 28, 2007 from the Assessment Reform Group Web site: <http://www.arg.educ.cam.ac.uk/publications.html>
- Assessment Reform Group.(2002). *Assessment for learning: 10 principles*. Retrieved June 2, 2007 from the Assessment Reform Group Web site: <http://www.arg.educ.cam.ac.uk/publications.html>
- Ballowe, S., & Sullivan, C. (2006). *The Gainesville model for accountability and achievement*. Gainesville, GA: Gainesville School System.
- Bambino, D. (2006). *Teaching Outloud: A middle grades diary*. Retrieved from <http://www.middleweb.com/mgvoices1.html>
- Barbanell, P., Falco, J., & Newman, D. L. (2004). *Transforming educational societies through interactive video conferencing*. Retrieved October 1, 2007
- Barootchi, N. & Keshavarz, M. H. (2002). Assessment of achievement through portfolios and teacher-made tests. *Educational Research*, 44 (3), 279-288.
- Bell, B. & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education*, 85 (5), 536-553.
- Black, P. (2003, April). *The nature and value of formative assessment for learning*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Black, P. & Harrison, C. (2001a). Feedback in questioning and marking: The science teacher's role in formative assessment. *School Science Review*, 82 (301), 55-61.
- Black, P. & Harrison, C. (2001b). Self and peer-assessment and taking responsibility: The science teachers' role in formative assessment. *School Science Review*, 83 (302), 43-49.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5 (1), 7-74.
- Black, P. & Wiliam, D. (2003). 'In praise of educational research': Formative

assessment. *British Educational Research Journal*, 29 (5), 623-637.

- Bloom, B.S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on Formative and Summative Evaluation of Student Learning*. New York; McGraw-Hill.
- Borko, H. , Davinroy, K. H., Bliem, C. L., & Cumbo, K. B. (2000). Exploring and supporting teacher change: Two third-grade teachers' experiences in a mathematics and literacy staff development project. *Elementary School Journal*, 100 (4).
- Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education*, 8 (2), 153-169.
- Brophy, J., & Good, T. (1986). *Teacher behavior and student achievement*. New York: Macmillian.
- Census.gov (n.d.). Retrieved October 9, 2007, from <http://census.gov>
- Center for Educational Research and Innovation. (2005). *Formative assessment: Improving learning in secondary classrooms*. Retrieved March 2007 from the Organization for Economic Co-operation and Development Web site: <http://www.oecdbookshop.org/oecd/display.asp?sf1=identifiers&st1=962005021P>
- Cheney, L.V. (1995). *Telling the Truth: Why our Culture and our Country have Stopped making Sense - and What we can Do about it*. New York, NY: Simon & Schuster.
- Clark, C. Chow-Hoy, T. K., Herter, R. J., Moss, P. A. (2001). Portfolios as sites of learning: Reconceptualizing the connections to motivation and engagement. *Journal of Literacy Research*, 33 (2), 221-241.
- Cogshall, J. G. (2004). *Educational administration, policy and reform: Research and measurement*. Alexandria, VA: Age Publishing.
- Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weindfeld, F., et al. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.
- Commission on Excellence in Education. (1983). *A Nation at Risk: The Imperative for Educational Reform*. Washington, DC: United States Government Printing Office. National Council on Education Standards and Testing. (1992). *Raising Standards for American Education*. Washington, DC: National Council on Education Standards and Testing.

- Cooper, S. (2005). *Criterion referenced instruction (CRI)*. Retrieved October 1, 2007, from <http://www.lifecircles-inc.com/mager.htm>
- Council for Basic Education. (1998). *Standards for Excellence in Education: A Guide for Parents, Teachers and Principals for Evaluating and Implementing Standards for Education*. Washington, DC: Council for Basic Education.
- Cowie, B. & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education*, 6 (1), 101-116.
- Cowie, B. (2005). Student commentary on classroom assessment in science: A sociocultural interpretation. *International Journal of Science Education*, 27 (2), 199-214.
- Crooks, T. (2001, September). *The validity of formative assessments*. Paper presented at the annual meeting of the British Educational Research Association, Leeds, England.
- Davies, P. (2000). The relevance of systematic reviews to educational policy and practice. *Oxford Review of Education*, 26 (3&4), 365-378.
- Davies, P., Durbin, C., Clarke, J. & Dale, J. (2004). Developing students' conceptions of quality in geography. *The Curriculum Journal*, 15 (1), 19-34.
- Doppelt, Y. (2003). Implementation and assessment of project-based learning in a flexible environment. *International Journal of Technology and Design Education*, 13, 255-272.
- Dori, Y. J. (2003). From nationwide standardized testing to school-based alternative embedded assessment in Israel: Students' performance in the Matriculation 2000 project. *Journal of Research in Science Teaching*, 40 (1), 34-52.
- Doyle, D.P. and Pimental, S. (1999). *Raising the Standard: An Eight-Step Action Guide for Schools and Communities*. Thousand Oaks, CA: Corwin Press.
- Earl, L. M. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin Press, Inc.
- Ed.gov (n.d.). Retrieved October 11, 2007, from <http://www.ed.gov/NCLB>
- Evans, J. & Benefield, P. (2001). Systematic reviews of educational research: Does the medical model fit? *British Educational Research Journal*, 27 (5), 527-541.

- Georgia Department of Education (2008). Retrieved March 23, 2008, from www.doe.k12.ga.us
- Gipps, C. V. (1994). *Beyond testing: Toward a theory of educational assessment*. London: The Falmer Press.
- Gipps, C. V. (2002). Sociocultural perspectives on assessment. In G. Wells & G. Claxton (Eds.), *Learning for life in the 21st century: Sociocultural perspectives on the future of education* (pp. 73-83). Oxford: Blackwell Publishers, Ltd.
- Gregory, G. H., & Chapman, C. (2002). *Differentiated instructional strategies*. Thousand Oaks, CA: Corwin Press.
- Guskey, T. R. (2005, April). Formative classroom assessment and Benjamin S. Bloom: Theory, research, and implications. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada..
- Hand, B. & Prain, V. (2002). Teachers implementing writing-to-learn strategies in junior secondary science: A case study. *Science Education*, 86, 737-755.
- Haug, & Linn, . (2002). *Statistical Misunderstandings in School Accountability*. Retrieved September 23, 2007, from <http://www-stat.stanford.edu/~rag/nclb>
- Hayward, L. & Hedge, N. (2005). Travelling towards change in assessment: Policy, practice and research in education. *Assessment in Education*, 12 (1), 55-75.
- Hayward, L., Priestley, M. & Young, M. (2004). Ruffling the calm of the ocean floor: Merging practice, policy and research in assessment in Scotland. *Oxford Review of Education*, 30 (3), 397-415.
- Hermann, R. & Lewis, B. (2004). A formative assessment of geologic time for high school Earth Science students. *Journal of Geoscience Education*, 52(3), 231-235.
- Hickey, D. T. & Zuiker, S. J. (2005). Engaged participation: A sociocultural model of motivation with implications for educational assessment. *Educational Assessment*, 10 (3), 277-305.
- Hodgen, J. & Marshall, B. (2005). Assessment for learning in English and mathematics: A comparison. *The Curriculum Journal*, 16 (2), 153-176.
- Hsieh, H.-F. & Shannon, S.E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15 (9), 1277-1288.

- Hunter, J. E., & Schmidt, F. L. (1990). *Correcting error and bias in research findings*. Newbury Park, CA: Sage.
- Jencks, C., Smith, M. S., Ackland, H., Ban, J. J., Cohen, D., & Grintlis, H. (1972). *Inequality: A reassessment of the effects of family and schools in America*. New York: Basic Books.
- Joint Committee on Standards for Educational Evaluation. (2003). *The Student Evaluation Standards*. Thousand Oaks, CA: Corwin Press.
- Kearsley. (2006). *Instructional Design Models*. Retrieved October 5, 2007, from http://www.carbon.cudenver.edu/~mryder/itc_data/idmodels.html
- Kendall, J.S. and Marzano, R.J. (1997). *Content Knowledge: A Compendium of Standards and Benchmarks for K-12 Education*. Aurora, CO: Mid-Continent Regional Educational Laboratory.
- Kirkwood, M. (2000). Infusing higher-order thinking and learning to learn into content instruction: A case study of secondary computing studies in Scotland. *Journal of Curriculum Studies*, 32 (4), 509-535.
- Leahy, S., Lyon, C., Thompson, M. & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership*, 63 (3), 18-24.
- Leat, D. & Nichols, A. (2000). Brains on the table: Diagnostic and formative assessment through observation. *Assessment in Education*, 7 (1), 103-121.
- Los Angeles unified school district. (2006, October). *LAUSDnet*. Retrieved September 10, 2007, from <http://www.lausd.k12.ca.us>
- Mabry, L., Poole, J., Redmond, L., & Schultz, A. (2003). Local impact of state testing in Southwest Washington. *Education Policy Analysis Archives*, 11 (22). Retrieved July, 2007 from <http://epaa.asu.edu/epaa/v11n22>
- Marzano, R. (2000). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Marzano, R. J., Waters, T., & McNulty, B. A. (2005). *School leadership that works*. Alexandria, VA: Association for Supervision and Curriculum Development.

- Mayring, P. (2000). Qualitative content analysis. *Forum: Qualitative Social Research* [Online], 1 (2), Available at <http://www.qualitative-research.net/fqs-texte/2-00/2-00mayring-e.html>
- McDonald, B. (2002). Self assessment skills used by high school students without formal training. *School Psychology International*, 23 (4), 416-424.
- McDonald, B. & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education*, 10 (2), 209-220.
- McLeod, D.B., Stake, R.E., Schappelle, B.P., Mellissinos, M. and Gierl, M.J. (1996). 'Setting the standards: NCTM's role in the reform of mathematics education'. In: Raizen, S.A. and Britton, E.D. (eds.). *Bold Ventures*. (Volume 3: Case Studies of U.S. Innovations in Mathematics Education). Dordrecht, Netherlands: Kluwer Academic Publishers, 13-132.
- McReal Laboratory. (1998). *content knowledge standards and benchmarks data base*. Retrieved October 1, 2007, from <http://www.mcrel.org>
- McTighe, J. & O'Connor, K. (2005). Seven practices for effective learning. *Educational Leadership*, 63 (3), 10 -17.
- Mid-continent research for education and learning*. (1998). Retrieved October 1, 2007, from <http://www.mcrel.org>
- Miller, L. (2005). Assessing assessment. *Professionally Speaking*, March, 37-40.
- Nash, G.B., Crabtree, C. and Dunn, R.E. (1998). *History on Trial: Culture Wars and the Teaching of the Past*. New York, NY: Alfred A. Knopf.
- National commission on excellence in education. (1983, April). *A nation at risk: The imperative for education Reform*. Retrieved October 4, 2007, from <http://www.ed.gov/pubs/NatAtRisk/index.html>
- National Science Foundation. (1983). *Educating Americans for the 21st Century: A Plan of Action for Improving Mathematics, Science and Technology Education for all American Elementary and Secondary Students so that their Achievement is the Best in the World by 1995*. Washington, DC: National Science Foundation.
- Noonan, B. & Duncan, R. (2005). Peer and self-assessment in high schools. *Practical Assessment, Research & Evaluation*, 10 (17). Available online: <http://pareonline.net/getvn.asp?v=10&n=17>

- Nunes, A. 2004. Portfolios in the EFL classroom: Disclosing an informed practice. *ELT Journal*, 58 (4), 327-335.
- Peters, T.J. and Waterman, R.H. (1982). *In Search of Excellence: Lessons from America's BestRun Companies*. New York, NY: Harper and Row.
- Principles for Fair Student Assessment Practices for Education in Canada*. (1993). Edmonton, Alberta: Joint Advisory Committee.
- Pryor, J. & Torrance, H. (1998). Formative assessment in the classroom: Where psychological theory meets social practice. *Social Psychology of Education*, 2, 151-176.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28 (1), 4-13.
- Ravitch, D. (1995). *National Standards in American Education: A Citizen's Guide*. Washington, DC: Brookings Institution.
- Reeves, D. B. (2001, January). Standards make a difference: The influence of standards on Classroom Assessment. *Bulletin*
- Roos, B. & Hamilton, D. (2005). Formative assessment: A cybernetic viewpoint. *Assessment in Education*, 12 (1), 7-20.
- Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18 (2): 119-144.
- Sandelowski, M., Docherty, S. & Emden, C. (1997). Qualitative metasynthesis: Issues and techniques. *Research in Nursing & Health*, 20, 365-371.
- Sanders, W. L., & Horn, S. P. (1994). The Tennessee value-added assessment system (TVAAS): Mixed -model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, 8, 299-311.
- School Reform*. (n.d.). Retrieved October 3, 2007, from <http://www.lab.brown.edu/topics/reform.shtml>
- School reforms model overview*. (2006). Retrieved October 4, 2007, from http://www.edsource.org.edu_refmod.cfm

- Schwartz, D. L. & Matin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, 22 (2), 129-184.
- Service-learning and standards tool kit. (2001). *Education commission of the states*, p. 175. Retrieved October 12, 2007, from <http://www.ecs.org>
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29 (7),1-14.
- Shepard, L. A. (2005). Linking formative assessment to scaffolding. *Educational Leadership*, (3), 66-71
- Simon, M. & Forgette-Giroux, R. (2000). Impact of a content selection framework on portfolio assessment at the classroom level. *Assessment in Education*, 7 (1), 83-101.
- Smith, E. & Gorard, S. (2005). “They don’t give us our marks”: The role of formative feedback in student progress. *Assessment in Education*, 12 (1), 21-38.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*,3 (10), 75-78.
- Stiggins, R. J. & Chappuis, J. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory into Practice*, 44 (1), 11-18.
- Stokking, K., van der Schaaf, M., Jaspers, J., Erkens, G. (2004). Teachers’ assessment of students’ research skills. *British Educational Research Journal*, 30 (1), 93-116.
- Swanson, J. (2001, October). *Six years of standards based middle school reform*. Retrieved September 6, 2007, from <http://www.edmatters.org/webreports/longbeach/20016>
- Symcox, L. (2002). *Whose History? The Struggle for National Standards in American Classrooms*. Williston, VT: Teachers College Press.
- Thissen-Roe, A., Hunt, E. & Minstrell, J. (2004). The DIAGNOSER project: Combining assessment and learning. *Behavior Research Methods, Instruments, & Computers 2004*, 36 (2), 234-240.
- Threlfall, J. (2005). The Formative Use of Assessment Information in Planning - the Notion Of Contingent Planning. *British Journal of Educational Studies*, 53 (1), 54-65.
- Tierney, R. D. (2006). Changing practices: Influences on classroom assessment.

Assessment in education: Principles, Policy & Practice, 13 (3), 239-264.

- Torrance, H. & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal*, 27 (5), 615-631.
- Tomlinson, C. A. (2001). *How to Differentiate Instruction in Mixed-Ability Classrooms* (2nd). (Brophy & Good, 1986) (Coleman, et al., 1966) (Jencks, et al., 1972) (Hunter & Schmidt, 1990) (Rosenthal, 1991) (Sanders & Horn, 1994) (Wright, Horn, & Sanders, 1997) Alexandria, VA: Association for Supervision and Curriculum Development.
- Torres Pereira de Eca, M. T. (2005). Using portfolios for external assessment: An experiment in Portugal. *International Journal of Art & Design Education*, 24 (2), 209-218.
- Vendlinski, T. & Stevens, R. (2002). Assessing student problem-solving skills with complex computer-based tasks. *The Journal of Technology, Learning, and Assessment*, 1 (3). Retrieved November, 2007 from <http://www.jtla.org>
- Verhoeven, J. C. & Devos, G. (2005). School assessment policy and practice in Belgian secondary education with specific reference to vocational education and training. *Assessment in Education*, 12 (3), 255-274.
- Vinovskis, M. A. (1999) *The Road to Charlottesville: The 1989 Education Summit*. Washington, DC: National Education Goals Panel.
- William, D., Lee, C., Harrison, C., Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education*, 11 (1), 49-65.
- Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11, 57-67.
- Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, 45, 477-501.
- Yung, B. H. W. (2001). Three views of fairness in a school-based assessment scheme of practical work in biology. *International Journal of Science Education*, 23 (10), 985-1005.
- Zvoch, K. (2003, June). *Technical report on APS middle school mathematics*. Retrieved October 3, 2007, from <http://www.unm.edu/~keithz/midschtechreport.pdf>

Zvoch, K., & Stevens, J. J. (2003, July). A multilevel, longitudinal analysis of middle school math and language achievement. *Education Policy Analysis Archives*
Retrieved September 12, 2007, from <http://epaa.asu.edu/epaa/v11n20/>

APPENDIX A

GMS 6th Grade Math
last!!) _____
Post Test #1
1st Nine Weeks

Student Name (first and
Teacher Name _____
Course Name _____

Directions. Please make sure the above information is written on your answer sheet. Please **DO NOT WRITE ON THIS TEST**. All work may be done on scratch paper. Carefully bubble all of your responses in on your answer sheet. If you do not know how to solve a problem, please leave it blank. Your teacher will use the scores on this pretest to help plan lessons so do the best you can. You may not use a calculator. If you have any questions, raise your hand.

1. Which number is composite?

- A. 11 B. 23 C. 72 D. 37

2. Which number is prime?

- A. 53 B. 56 C. 57 D. 51

3. Which number is a factor of 44?

- A. 8 B. 11 C. 14 D. 24

4. List all the factors for the number 60

- A. 1, 2, 3, 4, 5, 6, 10, 12, 15, 20, 30, 60
 B. 1, 2, 4, 5, 10, 20
 C. 1, 2, 3, 4, 7, 14, 15, 20, 30, 60
 D. 1, 2, 3, 4, 5, 10

5. Which answer is *always* even?

- A. odd + odd B. odd \times odd C. even + odd D. none of these

6. Alejandro and Jean are distributing erasers and pencils to the art class. There are 42 erasers and 49 pencils. Each student receives the same number of pencils and the same number of erasers, and no supplies are left over. Using what you know about GCF, what is the greatest number of students in the class?

- A. 7 students B. 91 students C. 294 students D. 14 students

7. Which number is a multiple of 15?

- A. 1 B. 5 C. 10 D. 30

8. Which of the following numbers are multiples of 9?

0, 7, 9, 16, 18, 19, 27, 54, 90, 96

- A. all of the above
 B. 0, 7, 16, 19, 27, 96
 C. 9, 18, 27, 54, 90
 D. 0, 9, 19, 90, 96

9. Which number is a common multiple of 7 and 4?

- A. 11 B. 24 C. 48 D. 56

10. Find the LCM (least common multiple) of 30 and 9.

- A. 90 B. 180 C. 3 D. 270

11. The number 84 is the Least Common Multiple (LCM) of which set of numbers?

- A. 2, 6, and 7 B. 2, 6 and 21 C. 6, 7 and 42 D. 6, 14, and 42

12. The picture graph below shows the amount of money each fourth-grade class raised for an animal shelter.

Amount Raised by Each Class

Class	Amount Raised
Ms. Smith	\$ \$ \$ \$ \$ \$
Mr. Powell	\$ \$ \$ \$
Ms. Carly	\$ \$ \$ \$ \$ \$ \$ \$
Mr. Roper	\$ \$ \$ \$ \$ \$

. If Mr. Powell’s class raised \$20 and Mr. Roper’s class raised \$30, how much money does one of the symbols represent ?

- A. \$1 B. \$4 C. \$5 D. \$20

13. Which is the greatest common factor of 20 and 36?

- A. 1 B. 2 C. 4 D. 6

14. Find the Greatest Common Factor (GCF) of the numbers: 20 , 48

- A. 8 B. 68 C. 4 D. 240

15. Which number satisfies all of the following clues? (n is a number.)

$$n > 6$$

$$n < 21$$

n is a multiple of 4

n is divisible by 3

- A. 8 B. 9 C. 12 D. 24

16. Find the prime factorization of the number 360

A. $2 \times 3^2 \times 5^2$

B. $2^3 \times 3^2 \times 10$

C. $2^3 \times 3^2 \times 5$

D. $2^3 \times 3^3 \times 5$

17. What is the prime factorization of 160?

A. $1 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 5$

B. $2 \times 2 \times 2 \times 10$

C. 24×5

D. $2 \times 2 \times 2 \times 2 \times 2 \times 5$

18. Which string of factors is not a factorization of 180?

A. $2 \times 3 \times 10 \times 3$

B. $23 \times 32 \times 5$

C. $22 \times 3 \times 3 \times 5$

D. $2 \times 15 \times 6$

19. Which of the following numbers has exactly two odd numbers in its prime factorization?

A. 24

B. 27

C. 35

D. 45

6th grade post test #1 – Last page

20. Which How many $\frac{1}{10}$'s are in $\frac{3}{5}$?

- A. 6
- B. 50
- C. $\frac{1}{6}$
- D. 30

21. Which of the following fractions is not equivalent to $\frac{1}{3}$?

- A. $\frac{4}{12}$
- B. $\frac{2}{6}$
- C. $\frac{3}{9}$
- D. $\frac{6}{3}$

22. Order these fractions from smallest to largest: $\frac{7}{10}$, $\frac{1}{2}$, $\frac{1}{4}$

- A. $\frac{7}{10}$, $\frac{1}{2}$, $\frac{1}{4}$
- B. $\frac{1}{4}$, $\frac{1}{2}$, $\frac{7}{10}$
- C. $\frac{1}{2}$, $\frac{7}{10}$, $\frac{1}{4}$
- D. $\frac{1}{2}$, $\frac{1}{4}$, $\frac{7}{10}$

23. Which symbol below makes the following statement true: $\frac{2}{5}$ _____ $\frac{2}{3}$

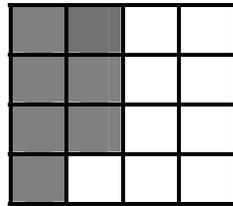
- A. >
- B. <

- C. =
- D. +

24. Choose the correct definition for Numerator

- A. The number written *below* the line in a fraction
- B. The number written *above* the line in a fraction
- C. A number of the form a/b , where a and b are whole numbers
- D. None of the above

25. What fraction of the grid below is shaded?



- A. $\frac{7}{10}$
- B. $\frac{7}{16}$
- C. $\frac{9}{10}$
- D. $\frac{9}{16}$

26. What fraction of the strip below is NOT shaded?



- A. $\frac{7}{10}$
- B. $\frac{3}{10}$
- C. $\frac{3}{7}$
- D. $\frac{7}{3}$

27. What is the least common denominator (LCD) of $\frac{5}{6}$ and $\frac{3}{10}$?

- A. 30
- B. 4
- C. 16
- D. 60

28. What is the least common denominator (LCD) of $\frac{1}{3}$, $\frac{1}{4}$ and $\frac{1}{2}$?

- A. 2
- B. 9
- C. 12
- D. 60

29. Convert $\frac{22}{5}$ to a mixed number.

A. 4

B. $4\frac{4}{5}$

C. $4\frac{2}{5}$

D. $1\frac{17}{5}$

30. Convert $2\frac{1}{4}$ to an improper fraction.

A. $\frac{9}{4}$

B. $\frac{7}{4}$

C. $\frac{3}{4}$

D. $3\frac{8}{4}$