2005

The Relationship Between Third and Fourth Grade Everyday Mathematics Assessments and Performance on the New Jersey Assessment of Skills and Knowledge in Fourth Grade (NJASK/4)

Mariann T. Helfant
Seton Hall University

THE RELATIONSHIP BETWEEN THIRD AND FOURTH GRADE *EVERYDAY MATHEMATICS* ASSESSMENTS AND PERFORMANCE ON THE NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE IN FOURTH GRADE (NJASK/4)

BY

MARIANN T. HELFANT

Dissertation Committee
Elaine Walker, Ph.D., Mentor
James Caulfield, Ph.D.
Rosemary Hrinko, Ed.D.
Tena Wright, Ed. D.

Submitted in partial fulfillment
of the requirements for the Degree
Doctor of Education.

Seton Hall University
2005

ABSTRACT

This study investigated the relationship between performance on 3rd and 4th grade *Everyday Mathematics* assessments and performance on the New Jersey Assessment of Skills and Knowledge for grade 4 (NJASK/4) for 623 students. The data used included test scores and proficiency levels on 3rd grade mid-year and end-of-year as well as 4th grade mid-year *Everyday Mathematics* assessments and the New Jersey state-mandated, 4th grade, standardized, criterion-referenced test. A series of statistical analyses including correlations, regression, cross tabulations and frequency distributions were performed to determine if *Everyday Mathematics* assessments had predictive validity for performance on the NJASK/4. The analyses also investigated whether predictive validity for performance existed for NCLB subgroups. Findings indicated that there was a significant, moderate relationship between student performance on *Everyday Mathematics* assessments and NJASK/4. Proficient performance on all *Everyday Mathematics* assessments accounted for 34% of the variance in relationship to NJASK/4 performance. Furthermore, 89% of students who performed at passing levels on the *Everyday Mathematics* assessments also passed NJASK/4. With regard to NCLB subgroups, relationships were also significant and moderate. Most subgroup populations had insufficient numbers to provide valid results. However, trends in performance for NCLB subgroups mirrored findings in other research studies. Overall, students who performed at proficient levels on *Everyday Mathematics* also performed proficiently on NJASK/4. Since 11% of the students did not perform proficiently on either assessment, other strategies and programs may be needed to assure student achievement. This study did not investigate the influence of teacher preparation or expectations on student performance nor did it examine other mathematical programs. All of these areas are appropriate subjects for further research.

ACKNOWLEDGEMENTS

I would like to express my appreciation to all who provided time, support, and encouragement while I strove to complete this lifelong goal.

I particularly want to thank my committee members who were always ready to give feedback and gentle prods when needed. I would like to thank my mentor, Dr. Elaine Walker, who not only guided me throughout this process but who made statistics understandable, relative and "user-friendly". Thank you also to Dr. James Caulfield, Dr. Rosemary Hrinko and Dr. Tena Wright who were always available for assistance.

Special thanks must go to my family, especially my husband Rex, who has lived through my obsession with reaching this goal. Matthew, thanks for leading the way, and Ann, thanks for listening. I also want to thank Barbara Kane and Dr. Walter Mahler for their willingness to support this study.

Finally, I would like to thank the faculty of Seton Hall's graduate School of Educational Leadership, Management and Administration who provided motivation and enthusiasm throughout my years of study.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

Introduction

Background

In 1965 Congress passed legislation entitled the Elementary and Secondary Education Act (ESEA). The purpose of this legislation, which was part of President Lyndon Johnson's "War on Poverty", was to provide public schools with additional federal dollars to improve the academic achievement levels of lower socio-economic level children. Under Title I of ESEA, public schools that received these federal funds had to monitor the progress of students through an assigned method of evaluation, generally standardized test scores. For the past thirty-five years, Congress has reauthorized this act on a five-year cycle. However, there has been mixed evidence that American students have increased their achievement level over that time.

In 1983, the National Commission on Excellence issued a document entitled *A Nation at Risk*. According to this document "...the educational foundations of our country are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people" (A Nation at Risk, 1983, p. 1). With regard to mathematics, the study also reported that only one-third of 17-year-olds could "solve a mathematical problem requiring several steps" (A Nation at Risk, 1983, p. 3). Furthermore, between 1975 and 1980 "remedial mathematics courses in public 4-year colleges increased by 72% and constituted one-quarter of all mathematics courses taught in those institutions" (A Nation at Risk, 1983, p. 3). Although the report called for reform in several areas, it also made clear that a public commitment to educational excellence and reform should not come at the expense of "equitable treatment of our diverse population" (A Nation at Risk, 1983, p. 5) Excellence and equity have long been the twin supports, and sometimes competing goals, for education in America.

In 1991 the United States Secretary of Labor appointed the Secretary's Commission on Achieving Necessary Skills (SCANS) to determine the skills young people would need to be successful in the workplace. The document outlined both the "fundamental skills' and the "workplace competencies" that students leaving high school should possess. The foundation skills included the basic skills of reading, writing, arithmetic/mathematics, listening, and speaking (Secretary's Commission on Achieving Necessary Skills [SCANS], 1991). Thinking skills, including creative thinking, decision-making, problem-solving, seeing things in the mind's eye (processes symbols, graphics, information), knowing how to learn, and reasoning, constituted the second leg of the foundation (SCANS, 1991). Personal qualities comprised the third section (SCANS, 1991). Personal qualities included responsibility, self-esteem, sociability, self-management, and integrity/honesty. The report also identified five "workplace readiness skills" (SCANS, 1991)

In 1989 and again in 1996 two major educational summits that included governors, educational and business leaders took place. As a result of the 1996 summit, governors from forty-one states agreed to support several initiatives including no social promotion, comparison of their individual states against others states through the National Assessment of Educational Progress (NAEP), and the introduction of standardized assessments for accountability purposes at grade 4 , 8 and high school exit (National Education Summit Policy Statement, 1996, p. 5-7).

In New Jersey the current tests of record are: the New Jersey Assessment of Skills and-Knowledge for grade 4(NJASK/4), the Grade Eight Proficiency Assessment (GEPA), and the High School Proficiency Assessment (HSPA) in grade 11.

Concurrent with the educational summits, President Clinton in his 1994 reauthorization of ESEA included the *Improving America's Schools Act* (IASA). This amendment to ESEA required that states receiving federal funds set up a system of standards and aligned assessments (Improving America's Schools Act [IASA] Public Law 103-382, 1994). By the year 2003, forty-nine states had met this goal (Mid Continental Research for Education and Learning, 2003). New Jersey produced its Core Curriculum Content Standards (NJCCCS) in 1996.

Despite this adherence to the federal directives, the debate among educators, legislators, and businessmen on how to improve America's schools and our students' achievement continued. This debate is not a new one. In the late 1800's to the early 1900's American educational conditions were quite similar to those that exist today (Payne, 2003). There was a change in the economic base as the country moved from an agrarian society to an industrial one and as technology moved from working with animals to working with machines. There was mass immigration, inclusion, as these immigrants as well as African–Americans joined the student bodies, and the development of private schools as some members of society sought to avoid inclusion. There was a strong involvement of business in education because business felt that schools were not producing effective workers. Finally, there was a shift in how schools delivered instruction as they moved from a one-room schoolhouse to a more structured factory model with students segregated by age and grade. As a result of these changes, emphasis was put on the development of testing instruments to determine student placement. Another focus of testing was to predict a student's ability to achieve in the public school environment. The original intent of such testing, such as that done by Alfred Binet, was to determine if students with mental disabilities could function in a public school. However, the tests were soon turned into "intelligence" tests for all students. Overall, the reason for testing was to provide a system of accountability (O'Neil & Tell, 1999).

Payne (2003) proposed that in today's educational environment, there is another shift in the economic base as society moves from an industrial society to a knowledge-based one and as technology moves from working with machines to working with computers and peripherals. There is mass immigration, inclusion of special education students in general education classes, and the development of public/private educational systems through charter schools, vouchers, and home schooling. Payne (2003) continues that since the 1980's, there has been a strong involvement of the business community in education because business once again feels that schools are not producing effective workers. Students (at least at the high school level and in higher education) also have a wider range of options for delivery of instruction. Once again society's answer to these changes is an accountability system based on testing. However, Payne (2003) contends that the accountability system has to be an assessment of the educational system rather than an assessment of the individual learner.

Consequently, when the United States Congress began the process of reauthorizing the ESEA in 2001, its members, and President Bush in particular, decided that drastic changes needed to be made in the legislation. The reauthorization of the Elementary and Secondary Education Act, in its new form as the No Child Left Behind Act (2001), embodies four key principles: stronger accountability for results, greater flexibility for states, school districts, and schools that use federal funds, particularly Title I funds, more choices for parents of children from disadvantaged backgrounds, and an emphasis on teaching methods that have been demonstrated to work (scientifically-researched methods). For the purposes of this study, particular attention will be paid to principles one and four.

Under NCLB (2001), schools have until the 2013-2014 school year to have all students perform at proficient levels on statewide assessments. To attain this goal states must develop and implement a single, statewide accountability system that will be effective in ensuring that all districts and schools make adequate yearly progress and hold accountable those schools that do not (United States Department of Education, 2002). Furthermore, schools must show that specific subgroups of students are also making progress and that at least 95% of each subgroup has taken the prescribed test. These subgroups include the economically disadvantaged, a variety of ethnic groups, special education students, and limited English proficient students (LEP). If even one subgroup does not meet the 95% standard or does not make adequate yearly progress as defined by state benchmarks then the entire school is considered in need of improvement (NCLB, 2001; USDOE, 2002).

To meet these demands public schools seek to develop curricula programs that align with state standards and also prepare students to perform at proficient levels on state tests. At the same time schools want to use curricula that not only support the test but also emulate best practice in the specific discipline. Schools place particular emphasis on scientifically-researched curricula in language arts and mathematics since these are the two disciplines for which NCLB currently holds states, districts, and specific schools accountable. The focus of this study will be on the mathematics curriculum, specifically the *Everyday Mathematics* program. Many school districts have selected the *Everyday Mathematics* program, also known as the *University of Chicago Math Program*, as the mathematics curriculum for their students.

Everyday Mathematics is a research-based curriculum developed by the University of Chicago School Mathematics Project (UCSMP). UCSMP was founded in 1983 during a time of growing consensus that our nation was failing to provide students with an adequate mathematical education (*Everyday Mathematics*, 1986). (Note that the UCSMP was founded in the same year as the publication of *A Nation at Risk, 1983.)*

Initial research for the *Everyday Mathematics* program began with a review of existing research on mathematical theory and instruction in America and abroad as well as interviews with hundreds of children. The program was developed one grade level at a time beginning with Kindergarten and involved the same team of authors for the K-6 curriculum. Because of the consistency of authorship, the curriculum offers a careful and selective sequence of instruction that builds on concept development in a spiral approach. Furthermore, the *Everyday Mathematics* program closely adheres to the standards and principles of the National Council of Teachers of Mathematics (NCTM) (*Everyday Mathematics*, 1986).

Beginning in 1989, the NCTM developed a vision for effective mathematics education. In 1989 the NCTM released its *Curriculum and Evaluation Standards for School Mathematics.* The NCTM followed this release in 1991 with the *Professional Teaching Standards for School Mathematics* and in 1995 with the *Assessment Standards for School Mathematics.* In 2000 the organization revised its initial work with the introduction of its *Principles and Standards for School Mathematics.* The revised principles and standards expect that all students will learn a common foundation of mathematics through high quality mathematical instruction (National Council of Teachers of Mathematics, 2000). It is the NCTM's belief that those who understand and perform mathematical operations and procedures have enhanced opportunities in the future. Those who lack an understanding of mathematical concepts will have doors closed to them (NCTM, 2000). This is a belief in "intellectual capital".

Intellectual capital is the "development of minds that can operate/manipulate/use the abstract representations of symbols and systems" (Payne, 2003, p. 11). It is difficult to measure intellectual capital with pencil and paper. However, Payne (2003) formulates a relationship between intellectual capital, standards, assessments, and accountability. She states:

> Standards are really an attempt to assign the intellectual capital that can be gained from a particular grade level or course. Assessments are an attempt to measure how much of the intellectual capital the student actually internalized. Accountability is an attempt to measure the success of the system developing intellectual capital. (p. 10)

The theory behind the development of intellectual capital is that communities that do not develop this ability in their students will ultimately become poorer. Hence, we see the relationship to accountability under No Child Left Behind and its emphasis on disaggregated data--accountability not just for the overall performance of a school but for each of its designated subgroups. Mathematics performance on statewide tests has been traditionally low (New Jersey Department of Education, 2002).

For the 4th grade ESPA (now NJASK), the starting point for AYP in Math is 53.1%. The State of New Jersey determined this AYP starting point by calculating the average score for students who passed the 2001-2002 ESPA in Mathematics. This indicates that students are beginning from a low level of understanding of mathematical concepts (53.1%) and will need to achieve substantial growth to meet the NCLB 100% proficiency mandate within twelve years. In order to achieve the goals of NCLB, considerable progress must be made in student acquisition of an understanding of mathematical concepts and in the ability to apply mathematical procedures and thinking. Schools will need to provide mathematics curricula that will assist them in achieving this end (Barton, 2003; Burns, 2003).

Statement of the Problem

Since schools will be held accountable for student success on state-wide assessments under No Child Left Behind (NCLB), it behooves schools to adopt curricula that will prepare students to achieve success. The question to be asked and answered then is: "How effective are the Everyday Mathematics assessments in predicting performance on the NJASK/4?"

The *Everyday Mathematics* program is based on research that parallels that of the NCTM standards and principles (*Everyday Mathematics*, 1986). The NJCCCS are similarly based on the NCTM standards. NJASK/4 is based on the NJCCCS. This would indicate that the program, national standards, state standards, and tests should have content validity. Content validity measures the degree to which a test represents the content, skills and/or behaviors of a domain of interest (Sattler, 1992). Usually, experts in the domain of interest determine a test's level of content validity. However, content validity is not sufficient for accountability. Content validity measures the degree to which items on a test are a representative sample of the content domain assessed. In order for content validity to be accurate, the domain of tasks must be clearly defined and the subject matter content and objectives including types of pupil performance should be identified (Sattler,1992). However, according to Popham (1975), "because content validity involves someone's inspecting the items and deciding whether they are sufficiently consonant with the content or learner behaviors to be measured, there is obviously a heavy reliance on human judgment" (p. 59). In other words, simply finding content validity between a domain of tasks and its assessment does not guarantee that the domain is a valid one. Popham (1975) goes on to state that there have been few "exemplary applications of this approach." Similarly, Messick (1995) determined that content validity only provides judgmental evidence of the domain but does not address test scores that are the core of the validity concept.

Maureen O'Sullivan Lally (2000), in her dissertation, found that external validity is necessary for purposes of accountability. She based this conclusion on the work of Anastasi (1959) in which she defined this aspect of validity as the extent to which test performance is related to another valued, independent and direct measure of that which the test is designed to assess. One form of external validity is predictive validity that measures whether achievement on one assessment can predict achievement on some future assessment of the same content and skills (Sattler, 1992). For the purposes of this paper, predictive validity will be used to determine the relationship between *Everyday Mathematics* assessments and performance on the NJASK/4 by correlating 3rd and 4th grade test results on the *Everyday Mathematics* mid-year and end-of-year assessments with performance on the NJASK/4. Furthermore, the study will investigate whether membership in an NCLB sub-group of ethnicity, socio-economic status, special education classification, or limited English proficiency also has an effect of NJASK/4 performance.

Purpose of the Study

The purpose of this study is to determine if there is a relationship between the *Everyday Mathematics* program as measured by its third and fourth grade assessments and performance on the 4th grade NJASK. The study will also investigate if this relationship extends to all subgroups identified by NCLB.

Research Question #1: What is the relationship between *Everyday Mathematics* assessments and performance on the NJASK/4 in mathematics?

Sub-question a: Are proficiency levels on individual *Everyday Mathematics* assessments, grade 3 mid-year, grade 3 end-of-year and grade 4 mid-year, valid predictors of performance on the NJASK/4?

Sub-question b: Are the scaled scores on individual *Everyday Mathematics* assessments, grade 3 mid-year, grade 3 end-of-year and grade 4 mid-year, valid predictors of performance on the NJASK/4?

Research Question #2: What relationship do the *Everyday Mathematics* assessments have with performance on the NJASK/4 for all NCLB subgroups: ethnicity, special education classification, economic disadvantage and limited English ability.

Significance of the Study

In our rapidly changing world, the ability to understand and use mathematics in everyday life is a necessity. Mathematical knowledge, tools and means of communication are constantly evolving and changing. Yet, classroom mathematics instruction, particularly at the elementary school level, has only recently begun to change. Michael Battista (1999) tells us that "to perform a reasonable analysis of the quality of mathematics teaching requires an understanding of not only the essence of mathematics but also of current research about how students learn mathematical ideas" (p. 425). He posits that traditional mathematics teaching where the same topics are taught and retaught each year with little retention by students constitutes a "massive miseducation" of American youth (Battista, 1999).

The National Council of Teachers of Mathematics (NCTM) also states that despite the efforts of educators and policymakers, the Council's vision of mathematical standards and principles is not "reality in the vast majority of classrooms, schools and districts" (NCTM, 2000). Consequently, students are not learning what they need to know in mathematics. Glenda Lappan (2000) quoting Carl Sagan delineates the need we have to educate future generations in mathematics, science and technology:

> We've arranged a civilization in which most critical elements profoundly depend on science (mathematics) and technology. We have also arranged things so that almost no one understands science (mathematics) and technology. This is a prescription for disaster. We might get away with it for a while, but sooner or later this combustible mixture of ignorance and power is going to blow up in our faces. (p. 329)

An understanding of mathematical concepts and applications is the key to opening doors for future generations. Those who do not possess the necessary knowledge, abilities, and critical thinking will find doors shut.

There have been some recent small gains on testing scores in mathematics, especially the National Assessment of Elementary Progress (NAEP). However, the overall results do not represent the achievement of all students. This is especially true for Hispanic and African-American students as well as special education students (Barton, 2003; NCES, 2003b). In the 1999 NAEP results, the average 17-year-old African-American and Hispanic student demonstrated a four-year lag behind white students in mathematics (NCES, 2003b). Furthermore, the same assessments indicate that only about 13% - 16% of twelfth graders, overall, are proficient in mathematics. According to the National Research Council 75% of Americans stop studying mathematics before they complete career or job prerequisites.

On the recently released 2003 NAEP results, Governor McGreevy and Commissioner Librera (NJDOE, 2003) were happy to announce that New Jersey students scored higher than students across the nation on 4^{th} and 8^{th} grade math. New Jersey students scored the fifth-highest scaled score (239) in 4^{th} grade mathematics and also scored above the national 4th grade math average (234). In addition, the NAEP reports indicated that a greater percentage of students scored proficient in 2003 versus 1996 (39% in 2003, 25 % in 1996). With respect to Hispanic and African-American 4th graders, the results were improving but still showed that students in these ethnic groups scored lower than their white counterparts. The gap between Hispanic and white 4th grade students in math lowered to 24 points from 32 points in 1996. Similarly, results between 4th grade African-American students and whites lowered from 38 points in 1992 to 31 points in 2003. However, a substantial gap still remained (NCES, 1994; NCES, 1998a, NCES, 1998b; NCES, 2003b).

On the 2003 results for NJASK/4, GEPA and HSPA, the scores for special education and limited English proficient students were the reason for more than 800 New Jersey elementary, middle and high schools being given "early warning" notices (NJDOE, 2004b). This is the focus of the No Child Left Behind legislation. It is neither acceptable nor possible for segments of the student population to be overlooked in the reporting of achievement. It is no longer sufficient to say that the overall population is improving when specific segments of the student population are not achieving at the same level.

The results of NAEP are significant because this is the external measure being used to determine if the New Jersey state standardized tests are measuring what they should be measuring. NAEP is New Jersey's form of construct validity. This study will investigate how effective the *Everyday Mathematics* program is in providing an additional measure of validity by predicting performance on a New Jersey state test not only for the general population but for all subgroups identified by NCLB.

Conceptual Framework

According to the New Jersey Core Curriculum Content Standards (NJDOE, 1996), the New Jersey Mathematics Standards and the respective benchmark assessments are aligned with the National Council of Teachers of Mathematics (NCTM) standards as refined for New Jersey's specific needs. Similarly, the *Everyday Mathematics* program as developed by the University of Chicago under the auspices of the National Science Foundation grant used the NCTM standards as its foundation (*Everyday Mathematics,* 1986).

If both the New Jersey Core Curriculum Standards and related state assessments and the *Everyday Mathematics* program and assessments use the NCTM standards as their basis, there should be congruent validity between the two assessments, NJASK and *Everyday Mathematics*. Furthermore, success on the *Everyday Mathematics* assessments should have predictive validity for performance on the New Jersey state tests due to this congruency. The need for this predictive validity becomes essential under the mandates of No Child Left Behind (NCLB) legislation.

The accountability provisions of NCLB require that every student, not just the aggregate of students for a school or district, make significant progress toward subject mastery over a specific time span. To this end, success on state standardized tests is no longer measured as a percentage of the total population but as a percentage of each disaggregated group present in the school community. Furthermore, the NCLB legislation (2001) requires that only scientifically researched and proven educational programs can be used, particularly in Title I schools, to provide students with instruction needed to demonstrate proficient performance on these state assessments. Failure to meet accountability standards may result in harsh penalties for schools and the educators who teach in those schools. In addition schools can no longer hide the lack of achievement by individual groups of students within aggregated reporting scores. This shift in educational philosophy (Schwartzbeck, 2004) has had a significant effect on the delivery of instruction in New Jersey. Schools need to know if the resources they invest in specific programs are worth the results as measured by state assessments.

Definition of Terms

Accountability: Under NCLB each state must develop and implement a single, statewide accountability system that will be effective in ensuring that all districts and schools make adequate yearly progress and will hold accountable those who do not.

Adequate Yearly Progress (AYP): Under NCLB each state must establish a definition of adequate yearly progress that each district and school is expected to meet. States must establish annual benchmarks to measure progress of the total population and nine specific subgroups so that all students reach proficient performance in twelve years. Benchmarks for AYP are based on students' scores on the 2001-2002 statewide assessments.

Correlation Coefficient: A number between –1 and 1 that describes the relationship between variables.

District Factor Group (DFG): The District Factor Group is the New Jersey Department of Education's classification of the socioeconomic status of its Citizens (NJDOE, 1974). The DFG is also used for comparative reporting of statewide test results. The measure was first developed in 1974 and has been updated after the 1980, 1990 and 2000United State census. There are eight categories of DFG ranging from A, the lowest socioeconomic group to J, the wealthiest group. The categorization is based on six factors: educational level, occupational status, population density, median family income, unemployment rate and poverty status. As of the 1990 census, the GH group has 78 districts. Although this information changed with the 2000 census, the 1990 information was still valid at the time of the study.

Elementary and Secondary Education Act (ESEA): Legislation passed by the United States Congress in 1965 based on the principle that children from lower socio-economic environments required more educational services than children from affluent environments (USDOE, 1965.

Educational Equity: All students have an equal opportunity to access and achieve success in education.

Educational Excellence: All students will receive a high standard of education that results in a high level of measurable achievement.

Improving America's Schools Act (IASA): This 1994 Congressional Amendment to the ESEA of 1965 recognized the fact that the achievement gap between disadvantaged children and others had decreased but was still sizable (USDOE, 1994).

Intellectual Capital: The knowledge (information and/or the ability to locate it) that can be used for economic gain or other useful purposes.

Limited English Proficient (LEP): Students who have a language other than English as a primary language and who have not yet achieved fluency (the ability to function appropriately and effectively) in English.

National Council of Teachers of Mathematics (NCTM): An association of mathematics educators whose goal is to ensure that all students receive a high-quality mathematics education through the implementation of standards viewed as necessary for mathematics education improvement.

NCTM Standards and Principles: Originally established in 1989, the standards and principles were revised in 1991, 1995, and 2001. They represent a comprehensive set of mathematical goals for all students, Pre-K through 12th grade. The principles are "statements reflecting basics precepts that are fundamental to high-quality mathematics education" (NCTM, 2000, p. 1). The principles include equity, curriculum, teaching, learning, assessment and technology.

New Jersey Assessment of Skills and Knowledge (NJASK): This test, established in 2003 by the Educational Testing service (ETS), is the successor to the New Jersey Elementary School Proficiency Assessment (ESPA). This test of reading, writing and mathematics is New Jersey's 3rd and 4th grade statewide assessment developed to meet the mandates of NCLB for yearly assessment in grade 3 through 8. The NJASK/4 is the test-of-record for all 4th graders in New Jersey schools and the basis for achieving AYP.

New Jersey Core Curriculum Standards (NJCCCS): Published in May, 1996 the NJCCCS are an attempt by New Jersey legislators to meet their constitutional obligation to provide a "thorough and efficient" education for all students. The standards encompass nine instructional areas: Language Arts Literacy, Mathematics, Science, Social Studies, World Languages, Visual and Performing Arts, Comprehensive Health and Physical Education, Technology Literacy and Career Education/Consumer, Family and Life Skills. The NJCCCS also include workplace readiness skills.

No Child Left Behind (NCLB): Federal legislation passed in 2001 by Congress and signed into law by President Bush on January 8, 2002. It replaces the ESEA of 1965 and the IASA of 1994 and focuses on individual school success as measured by student achievement data. NCLB also provides severe consequences for schools whose students do not meet prescribed achievement levels.

Predictive Validity: The degree to which the score on a test predicts the individual's score or performance in some other test or area.

Proficiency: New Jersey statewide assessments define proficiency as an achievement of a scale score of 200 or higher on the respective grade-level state test.

SCANS: The Secretary's Commission on Achieving Necessary Skills (SCANS): The U.S. Secretary of Labor appointed this committee in 1991 to determine the skills students would need to be successful in the world of work

Title I: The funding source for additional educational services under ESEA, IASA and NCLB. Title I funds are based on a school district's poverty level as well as an individual school's poverty level.

Test Validity: The degree to which a test measures what it claims to measure. Test validity is also the extent to which inferences, conclusions and decisions made on the basis of test scores are appropriate and meaningful.

Limitations of the Study

The study was conducted in one suburban school district that has six primary schools (K-3) and two intermediate schools (4-5) as well as a middle school and high school. For the purposes of this study, only the primary and intermediate schools were involved. Consequently, the conclusions of this study may not apply to other districts with different configurations.

The state of New Jersey categorizes the district as belonging to the district factor group (DFG) of GH. Therefore, the results of the study may not be applicable to other DFG's.

Of the eight schools under study, four qualify for Title I funds under No Child Left Behind. The district poverty level is 5%. The six primary schools have poverty levels that range from 0% to 12%. The two intermediate schools have poverty levels of 5% and 6%. The results of this study may not apply to districts with differing poverty levels.

One of the intermediate schools is a state of New Jersey Blue Ribbon School. The current fourth grade class (2003-2004) had 729 students. This is a large and growing district with low mobility. Students in this district have had a consistent *Everyday Mathematics* curriculum since Kindergarten. Most fourth grade teachers have been teaching the curriculum for two years and third grade teachers have taught the curriculum for three years. Only students who took all of the *Everyday Mathematics* assessments under consideration (grade 3 mid-year and end-of year and grade 4 mid-year) and the 2004 NJASK/4 test were included in the study.

The study is only investigating the relationship between the *Everyday*

Mathematics assessments and the NJASK/4. The conclusions of this study can not be generalized to other mathematical programs or other states' tests.

Organization of the Study

This study is organized into five chapters. Chapter I includes the background information, statement of problem, purpose of the study, significance of the study, conceptual framework, definition of terms, limitations of the study and the organization of the study.

Chapter II provides a review of literature related to the study including an examination of the No Child Left Behind legislation with particular emphasis on the accountability provisions, the history of standards-based reform, particularly national and New Jersey mathematics reform, and the curriculum response to both the legislation and the reform movement.

Chapter III describes the study's population data, the instruments used to collect data, and the procedures for collecting and analyzing data based on the study's research questions.

Chapter IV presents a statistical analysis of the quantitative data related to the research and a summary of the findings

Chapter V includes the summary, conclusions, and recommendations of the study.

CHAPTER II

Review of Literature

Introduction

The purpose of this study was to determine the relationship between the *Everyday Mathematics* third and fourth grade assessments and performance on the 4th grade NJASK. The study also investigated if this relationship extended to all subgroups identified by NCLB.

In order to understand the components of this study, the review of literature examined the No Child Left Behind legislation with particular emphasis on the accountability provisions, the history of standards-based reform, particularly national and New Jersey mathematics reform, and the curriculum response to both the legislation and the reform movement.

No Child Left Behind Legislation

The No Child Left Behind (NCLB) legislation is the reflection of a major change in American education policy (Schwartzbeck, 2004). For almost 170 years, federal education legislation has had universal access (equity) as its main focus. Ever since Horace Mann extended equal educational access to all elementary students in 1837, federal legislation has systematically opened the schoolhouse door wider and wider to provide equal opportunities to African-Americans, economically disadvantaged, minorities, migrants, women, the disabled, illegal immigrants and the homeless (Chart I, Appendix A). However, the NCLB legislation indicates that equity alone is insufficient.

Mathematics assessments and the NJASK/4. The conclusions of this study can not be generalized to other mathematical programs or other states' tests.

Organization of the Study

This study is organized into five chapters. Chapter I includes the background information, statement of problem, purpose of the study, significance of the study, conceptual framework, definition of terms, limitations of the study and the organization of the study.

Chapter II provides a review of literature related to the study including an examination of the No Child Left Behind legislation with particular emphasis on the accountability provisions, the history of standards-based reform, particularly national and New Jersey mathematics reform, and the curriculum response to both the legislation and the reform movement.

Chapter III describes the study's population data, the instruments used to collect data, and the procedures for collecting and analyzing data based on the study's research questions.

Chapter IV presents a statistical analysis of the quantitative data related to the research and a summary of the findings

Chapter V includes the summary, conclusions, and recommendations of the study.

CHAPTER II

Review of Literature

Introduction

The purpose of this study was to determine the relationship between the *Everyday Mathematics* third and fourth grade assessments and performance on the 4th grade NJASK. The study also investigated if this relationship extended to all subgroups identified by NCLB.

In order to understand the components of this study, the review of literature examined the No Child Left Behind legislation with particular emphasis on the accountability provisions, the history of standards-based reform, particularly national and New Jersey mathematics reform, and the curriculum response to both the legislation and the reform movement.

No Child Left Behind Legislation

The No Child Left Behind (NCLB) legislation is the reflection of a major change in American education policy (Schwartzbeck, 2004). For almost 170 years, federal education legislation has had universal access (equity) as its main focus. Ever since Horace Mann extended equal educational access to all elementary students in 1837, federal legislation has systematically opened the schoolhouse door wider and wider to provide equal opportunities to African-Americans, economically disadvantaged, minorities, migrants, women, the disabled, illegal immigrants and the homeless (Chart I, Appendix A). However, the NCLB legislation indicates that equity alone is insufficient.

Students and their parents should demand and expect excellence in the form of high achievement as measured by proficiency on state tests (NCLB, 2001).

Testing results from these state assessments are disaggregated by ethnicity, disability, limited English proficiency, and poverty level as well as total school performance to ensure that no child is left behind. The federal legislation also requires that annual school "report cards" indicate progress toward meeting state-mandated benchmarks established as a result of 2001-2002 baseline data. This progress should reflect movement toward meeting state standards and toward closing the achievement gap. Title I schools, those that receive federal funds based on poverty level, are immediately affected by this legislation and its accountability factors (NCLB, 2001; USDOE, 2002). Title I schools that do not make adequate yearly progress (AYP) toward these state-mandated goals will receive increasingly more severe sanctions including parental decisions regarding school choice, supplementary educational services, and ultimately, school restructuring. In exchange for this higher level of accountability, schools and districts have greater flexibility in spending Title I funds (NCLB, 2001). All other public schools are also required to meet the NCLB guidelines, but no specific sanctions have yet been set in place for these schools. However, every school's failure to meet the benchmarks is highly publicized in local and state newspapers.

The NCLB Act (2001) also requires that schools use educational programs and practices that are based on sound scientific research and that all students be taught by highly qualified teachers. A highly qualified teacher is one who has state certification, holds a bachelor's degree, and demonstrates subject area competency (NCLB, 2001; USDOE, 2002). Although NCLB legislation relates primarily to Title I schools, state reports as well as highly qualified teacher status relate to all schools. Title I school teachers had to meet the requirements by the start of the 2003-2004 school year. All other teachers have until June, 2006 to meet the requirements (NCLB, 2001)

These demanding tenets of NCLB result from many factors including the federal government's return on its investment in education. According to the United States Department of Education budget statistics (2000), the federal government spent more than 321 billion dollars between 1965 and 2000 through the Elementary and Secondary Education Act to educate disadvantaged children. Yet, the National Assessment of Educational Progress (NCES, 2003b) indicated that only 35% of fourth graders were performing at proficient levels in mathematics. The outcomes for African-American , Native American, and Hispanic students were even worse with only 4%, 12% and 10% of these respective groups having proficient performance (NCES, 2003b).

In New Jersey, Governor McGreevey and Education Commissioner Librera (NJDOE, 2003) were quick to point out that on the March, 2003 National Assessment of Elementary Progress (NAEP) for fourth grade, New Jersey ranked fifth in mathematics in the country. A closer look at that ranking shows that New Jersey fourth graders scored a 39% proficiency level in mathematics (Quality Counts, 2004; Stevens, 2003). This percentage represented an increase over the 1996 and 1992 mathematics results that showed only 25% of fourth graders had performed in the proficient range. However, the disaggregated data showed that this average came from performance by whites at 51% proficiency, African-Americans at 11% and Hispanics at 18%. The achievement gaps between these subgroups and whites were 40% and 33 % respectively. When NAEP added the "basic" level of achievement to the reports, there was a noticeable increase in mathematics achievement across all groups for the past decade (Quality Counts, 2004; Stevens, 2003). In 1992 and 1996, New Jersey fourth graders scored 68% in basic and proficient levels of mathematics whereas in 2003, they scored 80%. The major increase came at the proficient and advanced levels whereas the basic level remained relatively flat at 42% to 44% of the total population (Quality Counts, 2004; Stevens, 2003). In terms of the achievement gap, the 31% discrepancy between whites and African-Americans in mathematics in 2003 represented a smaller gap than in 1992 when the gap was 38%. Similarly, the 23% discrepancy between whites and Hispanics in 2003 represented a decline in performance gap from 32% in 1992. Low-income students, as defined by eligibility for free or reduced lunch, also saw their gap narrow from 32% in 1992 to 26% in 2003 (Quality Counts, 2004; Stevens, 2003). However, as the percentage of poverty students increased in any given school, the average scaled score in mathematics decreased for these students. Although New Jersey's overall data shows signs of improvement, the specific results, especially when disaggregated for mathematics performance, still reveal a serious achievement gap and substantial room for improvement (Stevens, 2003).

Ethnicity issues. For many years, such academic results did not surprise educators because they based their beliefs in student achievement on an extensive research study undertaken by Johns Hopkins University sociologist James Coleman. Coleman and his colleagues interpreted the data collected in their study as indicating that pupil achievement could not significantly increase until conditions "governed by race, class and socio-economic disadvantage could be overcome" (Coleman, Campbell, Hobson, McPartland, Mod, Weinfield, & York, 1966). In 1972, another distinguished researcher, Christopher Jencks, confirmed Coleman's findings (Jencks, Smith, Acland, Bone, Cohen, Ginits, Heyns, & Michelson, 1972). Both of these studies indicated that genetics and socio-economic environment shape children and determine their educational achievement (Fallon, 2003). In other words, poor and minority students could be expected to be poor achievers in school no matter what efforts the school made. An indirect consequence of this belief system was that teachers believed it was not their responsibility if children did not learn. Performance deficiencies could be explained by weak family structure, poverty, ethnicity, or poor nutrition (Jencks & Phillips, 1998; Ferguson, 1998).

However, more recent research and even subsequent data analyses by Jencks (2003) contradicted these findings. Numerous studies show that students achieve better in small schools with small classes taught by competent teachers implementing a challenging curriculum (Sanders & Rivers, 1996; Slavin,1998; Artzt & Amour-Thomas, 1999; Ferguson, 2002;). Of particular interest was the Tennessee Value-Added Assessment System developed by Professor William Sanders and June Rivers (1996). Their research showed that student achievement over a three-year period differed by more than 50% between groups due to the quality of teaching each group received. More importantly, the research found that poor and minority children demonstrated significant academic gains when they were taught by effective teachers (Sanders & Rivers, 1996). Furthermore, according to Fallon (2003), another study conducted in Texas showed "that the proportion of the variance in student achievement gain scores accounted for by teacher quality is twenty times greater than that from any variable, including class size and socio-economic status" (p. 3).

In addition, Campbell and Ramey (1995) in their study of the Abecedarian Project in North Carolina found that early intervention made a significant, lasting difference in achievement. In this project low-income, African-American children received extensive services from birth to age eight. These services included comprehensive infant health care, quality pre-school and kindergarten programs and family support services. Children in the project scored substantially higher than those in the control group on measures of intelligence, reading and mathematics at ages eight, twelve, and fifteen. Furthermore, children in this project were half as likely as their control group counterparts to be classified for special education or to be retained (Campbell & Ramey, 1995). This early intervention study is significant because according to Nancy Kober (2001) a disproportionate number of minority students come from low-income families and families where parents are not well-educated. Since both income and parental education are strong influences on student achievement, these factors affect minority students' achievement before they even attend school (Kober, 2001). Such disparity continues through high school where African-American and Hispanic students generally take fewer mathematics courses, and those that they do take are generally lower-level courses (Kober, 2001).

Socio-economic issues. Despite such research, poor and minority students still have unequal access to a variety of educational resources including skilled teachers and challenging curriculum (Darling-Hammond, 1998). Darling-Hammond (1998) reminds us that until forty years ago, African-American, Hispanic and Native American students received their education in segregated schools that received far less funding than their white, suburban counterparts. The school busing movement precipitated by the Coleman et. al. (1966) study did little to integrate schools and actually led to "white flight" in many urban areas. The result is that two-thirds of minority students still attend schools that are predominantly minority and poverty centered. Even in integrated schools, tracking systems tend to segregate low-income and minority students (Darling-Hammond, 1998). A careful examination of classrooms populated by minority students reveal an educational environment with fewer curriculum materials and texts, large class sizes, poorly qualified and inexperienced teachers, and limited access to a substantive curriculum that prepares students for higher education or productive careers (Darling-Hammond, 1998).

Monitoring of achievement gap issues by the Education Trust through a series of surveys has consistently shown that teachers and administrators in high poverty schools have low expectations for students (Haycock, 2001). The surveys show that these students receive fewer assignments than more affluent counterparts, and these assignments are generally low in quality. Furthermore, NAEP (NCES, 2003b) results indicate that students who take a full college preparatory sequence of courses in mathematics score higher than students who only take one or two mathematics courses. Yet, fewer than 12% of students in high poverty/high minority schools take the full sequence of mathematics (Haycock, 2001). One of the reasons cited for such a minimal mathematics curriculum is the lack of teacher expertise. According to the Education Trust surveys (Haycock, 2001), "only half of the mathematics teachers in schools with 90% minority enrollment meet their state's minimum requirement to teach" (p. 5).

Special education issues. In addition to ethnic/racial and economically disadvantaged subgroups, No Child Left Behind accountability requirements also include special education and Limited English Proficient (LEP) students who are now called English Language Learners (ELL). Until NCLB, special education students were generally excluded from state testing and accountability systems. Similarly, ELL students were generally not included in testing until they had completed an English as a Second Language (ESL) or bilingual program. This process usually took from three to four years. During that time span most ELL students would exit a district's program for English Language Learners, and consequently, not be counted in the accountability system under this heading. With NCLB (2001), both special education and ELL students were entitled to receive accommodations in order to take state tests, but they had to take the tests and be counted as subgroups for adequate yearly progress (AYP). The inclusion of these two subgroups has become highly controversial.

Until 1975 when Congress passed Public Law 94-142 (PL94-142), the Education for All Handicapped Children Act, which granted classified students a "free, appropriate public education," most special education students were excluded from mainstream public education (Olsen, 2004). Although special education students received access to public education through PL94-142, it was not until the 1997 reauthorization of this act, entitled the Individuals with Disabilities Education Act (IDEA) that public schools were required to set educational goals for classified students that aligned with regular education goals and standards. The 1997 IDEA also required that special education students be included in state testing. However, the Act had no requirements for proficiency on these tests, and there were no consequences for schools if students with disabilities had poor achievement. NCLB has radically changed this scenario through its accountability requirements for all students and its severe consequences for Title I schools that don't make progress.

One of the reasons that public educators consider NCLB's requirements to be unreasonable relates to the wide variety of students serviced under the umbrella of special education (Viadero, 2004). IDEA recognizes ten categories of disability ranging from profoundly cognitively impaired to learning disabled, a rather nebulous classification encompassing many different sorts of disabilities. Currently, there are six million students receiving special education and related services (Olsen, 2004). Of these students, approximately two-thirds are classified learning disabled or speech impaired, the two classifications that represent the least involved level of disability. Despite the arguments of some special educators that testing classified students with the same measures as "typical" students is unfair, framers of the NCLB legislation contend that given appropriate accommodations, all students, especially the least impaired, can achieve. Accommodations can include a small group test setting, extended test-taking time, scribes to write student responses, computers, calculators, and access to teachers who read and clarify directions among other accommodations (NCLB, 2001; NJDOE, 2004d). Students with disabilities have these accommodations written as part of their Individual Education Plans (IEP's). These plans are mutually composed and agreed upon by classroom teachers knowledgeable about the child's academic progress, child study team members who have assessed the students' strengths and weaknesses and the student's parents.

Part of the problem is that until 1997, special education students frequently did not receive the same curriculum as their regular education counterparts (Gingerich, 2004;

Viadero, 2003). Because of their disparity in background knowledge, Gingerich (2004) believes that special education students will require more time to achieve standards and benchmarks. Until then, classified students will lag behind. However, NCLB provides relief for students with disabilities as well as all other subgroups through its “safe harbor” provision (NCLB, 2001). As long as each subgroup has a 95% participation rate, subgroups that do not achieve AYP based on proficiency at benchmark levels, may use a 10% improvement rate in a specific subject area to attain “safe harbor”, and ultimately, achieve AYP. To determine the ten percent improvement, schools need to delineate the percentage of subgroup students who did not achieve proficiency on the last administration of the particular standardized test, in this case the New Jersey Assessment of Skills and Knowledge (NJASK/4). The school then calculates 10% of that figure and adds it to the percentage of students who did achieve proficiency on the test’s latest administration. If the percentage of students who pass the current administration equals or exceeds this total percentage, the school achieves “safe harbor”(NCLB, 2001; NJDOE, 2004b). This “loophole” in NCLB provides some relief for all subgroups. In addition, those students with the most severe disabilities are permitted to take an alternate assessment, the Alternative Proficiency Assessment (APA), to meet the NCLB requirements and AYP. The APA is a lengthy portfolio assessment based specifically on a student’s IEP. Therefore, the NCLB Act does provide some leeway for students with disabilities while still maintaining accountability to meet standards.

Complicating the special education issue is the over-representation of minority students in students with disabilities classifications (Commission on Behavioral and Social Science Education, 2002; Jerald, 2003; Viadero, 2004). According to the National

Research Council of the National Academies of Science (Viadero, 2004), African-American students were "over-represented in the special education categories with the most subjective criteria: mild mentally retarded, specific learning disabled, and emotionally disturbed classifications. In disability categories with a clearer medical or biological origin, group differences were practically nil" (p. 3). Native American and Hispanic students were found to be similarly over-represented. This disparity brings into question whether minority students are highly classified because they have inferior educational opportunities or because they truly have a higher level of learning disability. It is also interesting that the recommendations for improving achievement for special education students are the same as those offered for poorly achieving minority and low-income students. In a report issued by the Council of Chief School Officers (Blank, 2003), the guidelines for improving student achievement for students with disabilities include: high expectations, strong curriculum aligned to standards, competent teachers, early intervention, partnerships with families, and a focus on appropriate instructional delivery systems (Finn & Achilles, 1990; Smith & O'Day, 1991; Slavin & Madden, 1999; Borman, 2003; Blank, 2003; Knapp & Sields, 2004). The parallels with NCLB are obvious (Blank, 2003).

Limited English proficient issues. Of all the schools in New Jersey that have been labeled "in need of improvement," the vast majority have failed to make AYP based on either their special education or English Language Learner populations (NJDOE, 2004b). Just as the special education population has grown so has the ELL population. According to the Bureau of United States Citizenship and Immigration Services statistics (2001), the ELL population has grown by 105 % since 1990. Initially, NCLB required that these students take a state-mandated test in their first year in American public schools. Under the 2004 revisions ELL students may be tested in their native language for their first three years in public school and may also receive accommodations of extended testing time, small group testing situations, and translation dictionaries if the test is taken in English. In addition, schools may now code ELL students by their number of years in an English as a Second Language program (NCLB, 2004). It is still unclear how the coding will affect AYP.

Similar to special education students, ELL students have traditionally received programs that are less aligned with state standards (Mid-Continental Research for education and Learning [McREL], 2004). To assist states in meeting the needs of ELL students, the NCLB provides additional funding under Title III for underachieving non-English speaking students. Yet, testing data show that ELL students as well as special education students still perform at 30% to 40% below non-disabled, English speaking students on standardized tests (NJDOE, 2004b).

Although NCLB does not base its achievement data on gender, the standardized test results for males and females are reported. Gender inequity, especially in mathematics, has been the subject of many educational articles and research projects (Strauss, 1991; Li, 1999; Kenschaft, 2004). To no one's surprise, some of the same recommendations made to increase achievement for minorities, special education, and ELL students are also made for females in the area of mathematics. Primary among these are high expectations and challenging course work.

Considering the research, it is not a coincidence that the NCLB Act requirements stress quality, scientifically-researched curriculum and highly qualified teachers. Nor is it surprising that the Act holds schools accountable for specific subgroups' achievement. The legislation implies that schools are not intrinsically motivated to provide those attributes of effective education necessary to close the achievement gap.

Standards Movement and Math Reform

Educational reform was already in progress in the 1950's when the Soviet Union launched *Sputnik* into orbit. However, this historical turning point "played a significant role in educational reform"(Bybee,1997). Americans viewed *Sputnik* as a threat to national security, and consequently, supported efforts to improve the educational system so that Americans could meet the Soviet challenge. Public support came at a crucial time because educational reform had come under attack and critics called for a return to basics (Bybee, 1997). *Sputnik* tipped the scales in favor of higher academic standards especially in mathematics and science and paved the way for the first federal education legislation with the National Defense Education Act of 1958. The reform model replaced textbooks with materials and strategies designed to replicate procedures used by mathematicians and scientists. The reform movement continued through the 1950's and 1960's until America's social and political factors caused the public to lose interest in educational excellence. The Vietnam War, Civil Rights Movement, and Watergate overrode public concern with education. Most of the changes achieved in mathematics disappeared by 1976/77, and mathematics teachers returned to basic curriculum instruction (Bybee, 1997).

Standards-based reform. Educational reform would not become an issue again until the publication of *A Nation at Risk* in 1983. This report showed that American students' scores exhibited a steady decline in measures of student achievement such as standardized tests and the Scholastic Aptitude Test (SAT). Simultaneously, this report indicated the increased need for remedial instruction and training both by schools and industrial employees. *A Nation at Risk* (1983) called for a return to excellence but also to equity. It laid the foundation for setting high expectations for all students.

Eight years later, in 1991, the Department of Labor, through the Secretary's Commission on Achieving Necessary Skills (SCANS) reiterated employers' needs to find competent workers with basic skills in reading, writing, and mathematics. The report indicated that employers still needed to engage in extensive remedial instruction with new employees (Secretary's Commission on Achieving Necessary Skills [SCANS], 1991). Furthermore, the report documented the lack of critical thinking, problem solving, and decision making skills in recent high school and college graduates.

According to Anne Lewis (1995) the last half-century textbook standards have "dumbed down" education, and this gave rise to a number of attempts to vitalize a standards movement. These attempts came to a head in 1992 when a group of American governors met and formulated Goals 2000. They debated a national curriculum and a national assessment, but these ideas did not receive support. Instead, the governors' committee decided on a system of standards and assessments that would arise from individual state initiatives.

Lewis, referring to the work of Diane Ravitch (1995), then described the different types of standards that states might consider in formulating standards. These standards included content, performance, opportunity to learn, and world class. Content standards establish what should be learned in specific subject areas. Performance standards define levels of learning considered satisfactory. These standards focus on student ability to apply and demonstrate what they know. Opportunity to learn standards refer to the conditions and resources necessary to give students an equal chance to meet performance standards. When the governors discussed this type of standard they were thinking about high quality teaching and instructional resources. World-class standards are based on content presented to and expectations held for students in other countries. The governors determined that the choice of standards and their implementation would be voluntary.

In Goals 2000 the governors also set up an agency, the National Education Goals Panel (NEGP), to decide if a standard is "good". The National Education Standards and Improvement Council (NESIC) was designated as the source of expertise to help the NEGP judge standards presented to it. Goals 2000 also set up a system of grants to help individual states develop their standards. Almost immediately, some states resisted what they saw as federal government interference in states' rights. These states saw the NEGP as a "Big Brother" movement and refused to take federal aid. Today, the NEGP helps states provide high-quality standards but does not judge them

The standards movement gave rise to a monumental amount of educational dialogue that revolved around two major conversations: the positive and negative effects of standards and the similar effects caused by the assessments devised to measure them. Those who supported standards believed that they could significantly contribute to higher student achievement (Schmoker & Marzano, 1999). Marzano (2000) contended that the success of any organization depends upon clear, commonly defined goals. His problem with the standards was that there were just too many to be taught (Marzano, 2000). Research done at the McREL educational laboratories in Colorado showed that teachers K-12 would need 15,000 hours to complete all of the standards. The average student K-

12 received 9,000 hours of instruction. Schmoker & Marzano (1999) also referred to the results of the Third International Mathematics and Science Study (TIMSS). American students perform progressively worse on these studies as they move through the grades (Third International Mathematics and Science Study [TIMSS], 1999). One of the reasons given for this result is that American students cover excessively higher percentages of topics than do their counterparts in Germany, Japan and Sweden (Schmidt, McKnight & Raizen, 1996; NCES, 1997; TIMSS, 1999; Hiebert, 2003). Furthermore, the lack of standards in the teaching/learning environment resulted in too many options. Teachers' picking and choosing of content caused large gaps in a student's education and a lack of continuity of content. Consequently, Schmoker & Marzano (1999) felt that "less is more" and that if the excessive tendencies of standards were trimmed, standards could benefit student achievement.

Linda Darling-Hammond (1998), who believes that standards can support more ambitious teaching and greater levels of success for all students provided the standards are shaped and used well, also holds similar views. The consequence of poor standards is to create higher rates of failure for those who are already adversely served by the current educational system, the poor and minorities (Darling-Hammond, 1998). Traditionally, these groups have had an overabundance of marginal teachers, deteriorating or inadequate facilities, and inferior, antiquated materials in their learning environments. In other words, the opportunity to learn standards have been inadequate for these students to meet performance standards.

In order to overcome such obstacles, both Schmoker & Marzano(1999) as well as Darling-Hammond (1999) advocated analyzing curriculum to determine the essential skills and knowledge students should learn and ensuring that all teachers taught them. Common teaching could then be achieved by developing banks of proven, standards-referenced instructional materials such as lessons, units and assessments perfected through action research. Furthermore, provision of access to well-trained, fully-qualified teachers, reasonable class sizes, and materials and equipment would also be necessary for student learning (Finn & Achilles, 1990; Burns. 2003).

Conversely, the concept of educational standards has drawn criticism from many factions. David Labaree (2000) saw standards as an attempt by some to return to times when schools were "tough" and students had to struggle for grades.

He described the resistance to standards as an historical movement encompassing three major factors: preservation of local control, expanding educational opportunity, and form over substance. The scope of local control was evident. Local school boards wanted to be able to decide on their own curriculum without governmental interference.

Expanding educational opportunity referred to the specific American tendency over the last two hundred years to provide more students with more schooling. Standards might restrict students from getting educational opportunities. Form over substance related to the amount of time spent in class and credits received rather than student performance. Measures such as the Carnegie unit stressed attendance over performance.

Labaree (2000) then placed all of these resistances in the context of three historical purposes of education: democratic equality (citizen), social efficacy (taxpayer), and social mobility (consumer). Both democratic equality and social efficacy supporters saw education as a public good. In the former the focus was on higher levels of shared knowledge and skills that produce competent citizens. Supporters of this theory would stress general education, neighborhood schools social promotion, inclusion, and whole class instruction. Advocates for social efficacy saw education as the means for training productive workers. They would support vocational schools, tracking, and grouping. On the other hand, those who promoted the social mobility philosophy saw education as a private good that preserved advantages and increased distinctions. This philosophy relied on grading, sorting, and selecting. According to Labaree (2000) the standards movement was an attempt to push us back into this consumer approach by providing goals that fewer students could meet.

Such a belief was echoed by Alfie Kohn (2000) who stated that standards are incompatible with personalized learning, with the interests of marginal students, and with education in general. Although Kohn favored what he called "horizontal standards", those that involve changing teaching and learning in the classroom, he opposed "vertical standards" that just have children doing more of the same (O'Neil & Tell, 1999). Kohn felt that the idea that children just need to work harder to achieve the same goals at the end of each year flew in the face of human development.

Those who supported standards but were concerned that their lockstep implementation would harm some students were supported by Carol Ann Tomlinson's (2000) work on differentiated instruction. However, even Tomlinson did not see a contradiction between standards and differentiation. She stated, "The conflict between focusing on standards and focusing on individual learner's needs exists only if standards are used in ways that cause us to abandon what we know about effective curriculum and instruction" (Tomlinson, 2000, p. 6). She feared that as tasks became too complex, educators would retreat to a familiar standardized approach for teaching students.

Indeed, standards were seen as a "double edged sword" by some (Pasi, 2000). Examples, such as those found in Virginia and Kentucky gave rise to this trepidation. Virginia's Standards for Learning (SOL) resulted in specific instructional and curricular changes. There was more direct instruction with teachers keeping to a tight instructional schedule. Previously, teachers taught topics in different orders, provided more group work and taught for meaning even if this meant not reaching part of the year's curriculum (Pasi, 2000). In addition, teachers rarely collaborated on instruction; whereas, collaboration, under SOL, became a necessity. Obviously, there were pros and cons to both sides of this example. Teacher collaboration and completion of the same curriculum for all were positive. However, the reliance on direct and scheduled instruction was not.

Another example of standards gone wrong was exemplified by the Kentucky experience (Jones & Whitford, 1997). Kentucky was one of the first states to put extensive educational reform into place. The Kentucky Education Reform Act (KERA) led to many progressive changes such as performance based assessment, an ungraded primary program, and a new financial aid structure. Its basic premise was that all students could learn at high levels. The program began with six learning goals and 75 outcomes. Eventually, goals three and four (self-sufficiency and responsible community members) were removed because local educators felt they were too difficult to measure. Furthermore, educators wanted the remaining goals to be matched with specific rather than general content since teachers and schools were placed on a financial incentive award or consequence plan. Therefore, the Core Content for the Kentucky Instructional Results Information System (KIRIS) was developed. KIRIS was the state's assessment instrument related to KERA. As requests for specifics became more demanding, the

KIRIS that originally featured performance events, open-response questions, and writing and math portfolios became reliant on multiple-choice questions, limited performance tasks, and removal of the math portfolio. This was a direct result of the KIRIS being an assessment of school accountability rather than student accountability (Jones & Whitford, 1997).

The concept of performance standards raised further questions for educators. Nell Noddings (1997) pointed out that athletes and professionals who are judged by performance standards chose to take part in competitions or to enter certain fields of study. K-12 education is compulsory, and any set of standards designed to be rich for one student's interests would contain items that are unnecessary for many. Conversely, a standard designed realistically for all would be inadequate for an individual. Noddings (1997) asserted that standards gave the illusion that everyone had a fair chance

In order for large-scale reform to be successful, time is essential. According to Michael Fullan (2000) the main enemies of reform are overload and fragmentation. He stated that to achieve successful change in student performance in an elementary school requires three years of practice. Reform also needs a collaborative work culture where professionals focus on student achievement and change instructional practices to get better results. Reform needs to be aware of powerful external forces such as parents, technology, business and governmental policies and be willing to include all constituencies in reform projects.

To date there has been little empirical evidence supporting or refuting a causal link between standards and enhanced student learning. However, there has been anecdotal evidence in some small, classroom-based experiences (Nave, Miech &

Mosteller, 2000). For example, in a rural New York community, students were assigned to a difficult Regents level biology class rather than a less rigorous non-Regents class. For one unit of study, students were told that they had to achieve 100% on a unit test. Students who did not score 100% had to continue to work on their own time until they achieved a perfect score. Eventually, 71 out of 72 did so. The students who had previously attributed school failure to outside sources now realized that they had control over their own success. Furthermore, they were all able to reach the same standard but in variable amounts of time and with a variety of study aids.

This variety is the very essence of the argument against standards and the assessments that they have produced. In New Jersey the statewide fourth grade Elementary School Proficiency Assessment (ESPA) came under intense scrutiny because of the imprecise and subjective nature of its scoring as well as the in inconclusive nature of the information it returned to the districts. This was particularly true in the Language Arts portion and led to a revision of that section of the ESPA.

Many educators believed that standardized tests and the amount of time spent studying for them would get in the way of learning (Baresic & Gelman, 2002). Others questioned the very nature of standardized tests. W. James Popham (1999) asserted that the substantial size of the content domain that standardized achievement tests are supposed to represent posed genuine difficulties for the developers of such tests. First of all, the descriptors of standardized tests needed to be general to make the tests acceptable to a nation of educators with a wide range of curriculum preferences. Second, test items were chosen based on only 40%-60% of the students being able to answer the question correctly. In other words they were designed to make some kids fail. Finally, questions were heavily influenced by out of school experiences. Popham (1999) continued that the only good purposes for standardized tests were to compare a student's relative strengths and weaknesses across subject areas, show student growth over time, and compare students nationally. Otherwise, standardized tests gave us little information about student learning or educational quality.

Unfortunately, the high-stakes nature of these tests caused some teachers and school districts to make the tests their curriculum resulting in neglect of subjects, topics, and outcomes not tested. In districts that followed this pattern, there was usually an overuse of instructional material that mimicked test items. Some schools even went so far as to reduce the number of students taking tests by encouraging certain students to stay home during testing or by exempting others from the tests (McClosky & McMunn, 2000). However, the accountability aspects of NCLB with regard to subgroup participation and score reporting took away this practice.

There are those who feel that that the new forms of accountability and assessment are the best tools we have to ensure quality education for all children. Although some civil rights leaders believe that high-stakes tests will cause more students to drop out, others believe that high stakes are imposed not by a test but by educational neglect (Taylor, 2000). Taylor (2000) believes that students who do not receive a quality education are penalized whether or not they receive a diploma. He feels that if the minority community wages war on standardized tests, they will continue the old regime of low expectations for poor and minority children. Instead, he advocates that minorities use the tests as a lever to demand that districts align their curriculum with standards, increase resources, and employ the best teachers. He believes that accountability will drive equity.

Michael Schmoker (2000) believes that schools need to go beyond standardized tests and create their own accountability and assessment systems. Gradually, based on local criteria and collaboration, schools should supplement objective assessments with summative projects, essential questions, scientific experiments, and performance proposals.

Murphy & Doyle (1999), who assert that every district in America needs a systematic way to examine and analyze academic performance, second this position. They suggest that a district develop a culture that encourages and rewards high standards, while they audit instructional program and student outcomes and that they survey the community to see if their vision aligns with that of the school. The district then needs to develop long and short terms plans to develop high standards and effective plans by giving teachers ongoing, systematic training and support. Finally, schools must make time the variable. Historically, schools have held time constant and have had learning vary. Murphy & Doyle (1999) make the argument for allowing students whatever time they need to meet the standards and move on. In some ways this proposal is akin to Ted Sizer's experiences as headmaster of the Phillips Academy in Andover (Goldberg, 1996). Although Sizer strongly believes that there is no one model or template for American schools, students at Andover did have to meet specific standards in order to move on and graduate. The difference is that students can stay or move based on their own needs and without prejudice. Consequently, students might receive a diploma at age 16 or 19 depending on their own performance.

Considering the wide variety of opinions and the multiplicity of factors surrounding standards and assessments, the debate could have continued for some time. However, Congress ended the debate with the passage of NCLB and hastened the movement to establish rigorous, consistent standards. All states had to develop their standards by the year 2000. NCLB did provide flexibility for states to develop their own pathways to achieve AYP and, ultimately, 100% proficiency by 2014. NCLB allowed states to establish their own definition of proficiency, their own tests to determine proficiency, and their own benchmarks to identify movement toward proficiency (Protheroe, Shillard & Turner, 2003; Education Commission of the States, 2004).

National council of teachers of mathematics standards. Of all the disciplines, mathematics was the most prepared for the standards movement and the testing and accountability provisions of NCLB. This was mainly due to the work of the National Council of Teachers of Mathematics (NCTM). In 1989 the NCTM published its *Curriculum and Evaluation Standards for School Mathematics*. This document presented a vision for mathematics reform based on five curricular goals including the student's ability to : value mathematics, develop confidence in personal mathematical aptitude, become a mathematical problem-solver, learn mathematical communication skills, and develop mathematical reasoning skills (NCTM, 1989; National Assessment Governing Board, 2002). The standards were broken into three grade level ranges: K-4, 5-8, and 9-12. The K-4 standards included thirteen curriculum strands: Mathematics as Problem Solving, Mathematics as Communication, Mathematics as Reasoning, Mathematical Connections, Estimation, Number Sense and Operations, Whole Number Concepts, Geometry and Spatial Sense, Measurement, Statistics and Probability, Fractions and Decimals, and Patterns and Relationships (NCTM, 1989).

Furthermore, the NCTM standards recognized the unique characteristics of young learners and supported reliance on children's innate inquisitiveness (National Academy of Sciences, 2000). The standards encouraged adoption of a developmentally appropriate curriculum that integrated student curiosity and mathematics. This developmentally appropriate curriculum would "incorporate real-world contexts, children's experiences, and children's language in developing ideas" (NCTM, 1989, p. 2). The standards document (1989) also indicated that successful programs would depend on the quality of the mathematical foundation developed in the first five years of school. It stated that "programs that provide limited developmental work, emphasized symbol manipulation and computational skills, and relied heavily on paper and pencil worksheets" did not match children's natural learning patterns nor "contribute to important aspects of children's mathematical development" (NCTM, 1989, p. 4).

Finally, the 1989 NCTM standards called for: "a curriculum that emphasized mathematical concepts and their uses in today's world, active involvement of students in performing mathematical tasks, a stimulating elementary classroom environment that allows all students to reach their potential, and assessment practices based on a variety of evidence sources" (Burton, 2003, p. 9).

The NCTM understood that their vision of a developmentally appropriate curriculum could not be realized if a shift in teacher training and development did not also occur. To that end, the Mathematics Council, in 1991, published its *Professional Standards for Teaching Mathematics*, a complementary companion to the 1989 content standards. The teaching/learning environment envisioned in the *Professional Standards* differed significantly from what many educators had expressed in their own mathematics education (Lee, 2001). This learning environment shift was based on two key assumptions. First, the teacher is the change agent, and second, teachers need on-going, consistent professional development and adequate resources to affect change. The *Professional Standards* addressed both the pre-service instruction and the continuing education of practicing teachers. They attempted to provide guidance to colleges and universities, state departments of education, public and private schools, and organizations involved in continuing professional development of educators. Primary among the *Standard's* requirements was the need for teachers of mathematics to understand and model sound mathematical thinking and reasoning, to visualize and be able to make connections between concepts both within and outside of the discipline, and to be fully aware of how students' develop as mathematical learners (NCTM, 1991). Furthermore, the *Professional Standards* reiterated the curricular goals that all students are capable of learning mathematics and that teaching math is a "complex process that cannot be reduced to "recipes or prescriptions" (NCTM, 1991, p. 1).

New Jersey core curriculum content standards in mathematics. The New Jersey Department of Education heeded the admonitions of the NCTM standards. Through the 1992/1993 school year a committee of New Jersey educators (teachers, administrators, college educators and mathematicians), and businessmen met to develop a framework for New Jersey's Core Curriculum Content Standards (NJCCCS) in mathematics. Their intent was to build on the NCTM's 1989 and 1991 curriculum and professional standards to provide a vision of exemplary mathematics learning and articulate the standards needed to achieve this vision (NJDOE, 1996; Kunz, 2003). Once developed, the State Department sent the draft framework to educators across the state for study, comment, and revision. New Jersey published a revised framework representing this extended input in January, 1995. A new group of educators and businessmen, the Governor's Review Panel for the Mathematics Curriculum Standards, built upon the framework and preliminary standards to develop the final product. On May 1, 1996 New Jersey adopted standards in mathematics and six other content areas as well as cross-content workplace readiness standards (NJDOE, 1996).

The New Jersey Mathematics Curriculum Standards (NJDOE, 1995) included 16 content standards and two learning environment standards. The 16 content standards encompass: problem solving, communications, connections, reasoning, tools and technology, number sense, geometry and spatial sense, numerical operations, measurement, estimation, patterns, relationships and functions, probability and statistics, algebra, discrete math, building blocks of calculus, and excellence and equity for all (NJDOE, 1996).

The learning environment standards included keys to success in the classroom and assessment. Each chapter began with an overview of the standard K-12 and then broke the concept into age appropriate levels K-2, 3-4, 5-6, 7-8 and 9-12. The specific content chapters also included activities designed to help students achieve academic expectations.

In 2001 New Jersey began its review of all standards as previously determined by the State Board of Education. The Mathematics Curriculum review panel wanted to maintain two principles: first, the content of the mathematics standards and their assessments were to remain essentially the same; second, the standards were to be presented in a more teacher-friendly manner (NJDOE, 2004a).

The 16 standards have been grouped into five standards each of which has several strands. The first four standards, the content standards, have a series of indicators that correspond with state testing clusters. The fifth standard includes all of the mathematical processes and may serve as the basis for future state assessments (NJDOE, 2004a).

New Jersey standardized tests: NJASK/4. The current state assessment for grade 4 is the New Jersey Assessment of Skills and Knowledge (NJSK/4). The original grade 4 standardized assessment was the Elementary School Proficiency Assessment (ESPA). New Jersey schools administered the ESPA to fourth grade students from 1997 through 2002. Initially, the ESPA included language arts, mathematics, and science components. Both the language arts and science subtests had performance assessments as well as multiple choice, open-ended short answer, and longer responses (NJDOE, 1996). The language arts test soon came under fire from educators questioning whether the standard levels were appropriate for fourth grade students. A reexamination of the language arts assessment standards led to a scoring revision in 1999-2000. Furthermore, similar to their Kentucky counterparts, New Jersey educators raised concerns regarding the validity and reliability of the performance assessments. This concern resulted in the removal of these hands-on segments from the testing process. The mathematics portion of the ESPA was the one area that educators were able to support.

In the spring of 2003, New Jersey replaced the ESPA with NJASK which is intended to eventually become a comprehensive assessment system for grades three through eight. This system will meet the accountability requirements of the NCLB.

These yearly assessments are to serve as indicators of student progress in meeting the state's curriculum standards (NJDOE, 2002). Students who do not demonstrate sufficient progress are to receive additional instructional support (NJDOE, 2003). Currently, NJASK/4 has language arts and mathematics components. In 2004, fourth grade students took a science field test, and this area was added in 2005 as part of the official assessment. The mathematics segment of NJASK is in exact alignment with the newly revised Core Curriculum Content Standards for mathematics (NJDOE, 2004c). An examination of the New Jersey standards and the NCTM standards, both 1989 and 2000 versions, shows that the New Jersey standards encompass all of the NCTM standards (Chart II, Appendix A).

Curriculum Response to NCLB and the Standards Movement

The combination of the mathematics reform and standards movement along with the No Child Left Behind(NCLB) accountability requirements has made it necessary to evaluate not only content but process and product in mathematics teaching. Realizing this need, in 1983 the federal government began providing grant funds to the National Science Foundation (NSF) to support the development of scientifically researched mathematics programs (National Science Foundation, 1983).

Everyday Mathematics program. Working in concert with the national Teachers of Mathematics, the NSF eventually funded twelve such programs: five for high school, four for middle school and three for elementary school. *Everyday Mathematics* is one of the elementary programs. It is an outgrowth of the original University of Chicago School Math Project (UCSMP) that was designed to improve mathematics instruction in grades 7 through 12. UCSMP was formed in 1983 as a response to the concerns articulated in *A*

Nation at Risk. The success of UCMSP at the secondary level provided impetus for Max Bell, Jean Bell and their colleagues (1988) to conduct extensive research on the appropriateness of their curriculum concepts for elementary-aged students. The development of the *Everyday Mathematics* program fits the definition of scientifically-based research established by the federal government: "Scientifically-based research means research that involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to educational activities and programs (USDOE, 2002, p. 3).

The federal guidelines include four types of research: academic research, research concerning effective classroom practices, field-testing research, and learner verification research. Academic research refers to research using experimental design preferably with randomly selected subjects for both the control and experimental groups. Research for effective classroom practices uses empirical methods either qualitative or quantitative that can be replicated with similar results regarding the hypothesis. Field-testing involves measurement across multiple observations. Learner verification is based on peer-reviewed studies of students and teachers using the program under study. The *Everyday Mathematics* program used all four types of research in its development (Everyday Mathematics, 1986).

Of particular interest was the systematic field-testing and revisions of the program conducted from 1986 to 1996. Each grade level of the program was drafted, field-tested under controlled conditions, analyzed using systematic, replicable procedures, and finally revised based on empirical findings. The same group of authors wrote the K-6 curriculum based on these researched findings. This provided a consistent voice to the program.

In addition to the field-testing, Northwestern University conducted a five-year longitudinal study of *Everyday Mathematics* (Carroll, 2000). This longitudinal study included student and teacher interviews, observations, written tests, and surveys. Some of the items on the written tests were drawn from the National Assessment of Educational Progress (NAEP) and from the TIMSS (Carroll, 2000). In addition, there have been numerous independent research studies of *Everyday Mathematics* as well as studies done by individual states and school districts that have adopted *Everyday Mathematics*. These include studies by the states of North Carolina, South Carolina, Kentucky, Arizona, Washington, Tennessee, Michigan, Florida, Illinois and Massachusetts. For example, in the Tri-State Student Achievement Study (Consortium for Mathematics and Its Applications [COMAP], 2000) for Illinois, Massachusetts and Washington, the Alternatives for Rebuilding Curricula (ARC) Center, a National Science Foundation (NSF) funded project located at the Consortium for Mathematics and Its Applications (COMAP), studied the effects of *Everyday Mathematics* on student performance on state-mandated standardized tests. The study included over 78,000 students, approximately half who used *Everyday Mathematics* and half who used other programs. Students in grades three through five were matched for reading level, socioeconomic status, mobility, and ethnicity. Students in the *Everyday Mathematics* program had to have used this mathematics program for at least two years. The study compared scores on all topics tested at each of the tested grade levels in each of the three states. The results showed that the average scores of students in the *Everyday Mathematics* program were year assessment as compared to performance on the NJASK/4 indicated that a significant number of students performed at higher levels on NJASK/4 than expected. For the grade four mid-year *Everyday Mathematics* assessment and NJASK/4, 546 students performed at the proficient level in both with an expected count of 537.4. This represented 88% of the entire sample and 89.5% of those who had a passing performance.

This performance data indicated that 89% of students who performed at a passing level on the *Everyday Mathematics* assessments in grades three and four also passed the NJASK/4. This information reinforced the finding that there is a positive relationship between performance on *Everyday Mathematics* assessments and performance on the NJASK/4.

Initially, the fact that 89% of the students performed proficiently on both the *Everyday Mathematics* assessments and the NJASK/4 seemed to be inconsistent with correlation levels of .4 to .6. However, scattergrams of students' scaled scores (Appendix C) indicated that the data showed a skewed population with most students scoring in the upper right-hand quadrant of Quadrant 1. The remainder of the graph was nearly empty except for a handful of outlier scores scattered throughout Quadrant 1.

Correlations look for a "line of best fit" on a diagonal from the origin. In this skewed data that line only existed in the upper right-hand quadrant of Quadrant 1 because 89% of the students scored in that area. This phenomenon resulted in a lower correlation statistic while maintaining a high level of frequency.

A series of linear regression analyses were also performed to determine the effectiveness of the relationship between the *Everyday Mathematics* assessments and performance on the NJASK/4.

The initial regressions were simple linear regressions in which the NJASK/4 results were predicted by the *Everyday Mathematics* assessment scores. The final regression was a multiple regression for all three *Everyday Mathematics* assessments combined and their respective effects on NJASK/4 performance. All of the regressions were significant at the .000 level of significance, and all models showed a positive direction in the ability of the variables to predict performance on the NJASK/4.

Table 6:
Effectiveness of the Relationship of Everyday Mathematics Scores (EMA) with Performance on the NJASK/4 for a Sample of 623

Time of Assessment	R^2	F	Beta	Sig	df
Grade 3 EMA mid-year scores	.173	129.875	.416	.000	1
Grade 3 EMA end-of-year scores	.189	144.875	.435	.000	1
Grade 4 EMA mid-year scores	.230	185.974	.480	.000	1
EMA combined	.338	105.527	.210 .207 .334	.000	3

In the regression between the *Everyday Mathematics* grade 3 mid-year assessment and NJASK/4 the model had an F value of 129.875 and was significant at the .000 level of significance with one degree of freedom. The R^2 value of .173 indicates that the *Everyday Mathematics* grade 3 mid-year assessment accounted for 17% of the variance related to performance on the NJASK/4 math section.

For the grade 3 end-of-year *Everyday Mathematics* assessment, the model had an F value of 144.875 and was significant at the .000 level of significance with one degree of freedom. The R^2 value of .189 indicated that the grade 3 end-of year *Everyday Mathematics* assessment accounted for almost 19% of the variance in performance on the NJASK/4. .

A review of the grade 4 mid-year results showed that this assessment with an F value of 185.974 was also significant at the .000 level of significance with one degree of freedom. The R^2 value of .230 indicated that the grade 4 *Everyday Mathematics* mid-year assessment accounted for 23% of the variance in performance on the NJASK/4.

As expected, the Beta values of the regression analysis were consistent with the correlation coefficients found in the correlation analysis of proficiency performance levels and showed that the effects of performance on the individual *Everyday Mathematics* assessment to performance on the NJASK/4 were relatively stable. Performance on each assessment tended to have the same predictive value on performance on NJASK/4. The regressions also demonstrated that as the assessments came closer to the time of taking the NJASK/4 their effectiveness in predicting proficient performance increased. This supported the findings of the correlations that were performed on the same data.

When all three assessments were combined, the resulting regression had an F value of 105.527 and was significant at the .000 level of significance with three degrees of freedom. The R^2 value of .338 indicated that the combined assessments accounted for almost 34% of the variance in performance on the NJASK/4. The Betas for individual *Everyday Mathematics* assessments decreased to .210, .207 and .334, indicating that proficient performance on all of the *Everyday Mathematics* assessments was a better indicator of successful performance on the NJASK/4 than individual performance on any

The initial regressions were simple linear regressions in which the NJASK/4 results were predicted by the *Everyday Mathematics* assessment scores. The final regression was a multiple regression for all three *Everyday Mathematics* assessments combined and their respective effects on NJASK/4 performance. All of the regressions were significant at the .000 level of significance, and all models showed a positive direction in the ability of the variables to predict performance on the NJASK/4.

Table 6:
Effectiveness of the Relationship of Everyday Mathematics Scores (EMA) with Performance on the NJASK/4 for a Sample of 623

Time of Assessment	R^2	F	Beta	Sig	df
Grade 3 EMA mid-year scores	.173	129.875	.416	.000	1
Grade 3 EMA end-of-year scores	.189	144.875	.435	.000	1
Grade 4 EMA mid-year scores	.230	185.974	.480	.000	1
EMA combined	.338	105.527	.210 .207 .334	.000	3

In the regression between the *Everyday Mathematics* grade 3 mid-year assessment and NJASK/4 the model had an F value of 129.875 and was significant at the .000 level of significance with one degree of freedom. The R^2 value of .173 indicates that the *Everyday Mathematics* grade 3 mid-year assessment accounted for 17% of the variance related to performance on the NJASK/4 math section.

For the grade 3 end-of-year *Everyday Mathematics* assessment, the model had an F value of 144.875 and was significant at the .000 level of significance with one degree of freedom. The R^2 value of .189 indicated that the grade 3 end-of year *Everyday Mathematics* assessment accounted for almost 19% of the variance in performance on the NJASK/4. .

A review of the grade 4 mid-year results showed that this assessment with an F value of 185.974 was also significant at the .000 level of significance with one degree of freedom. The R^2 value of .230 indicated that the grade 4 *Everyday Mathematics* mid-year assessment accounted for 23% of the variance in performance on the NJASK/4.

As expected, the Beta values of the regression analysis were consistent with the correlation coefficients found in the correlation analysis of proficiency performance levels and showed that the effects of performance on the individual *Everyday Mathematics* assessment to performance on the NJASK/4 were relatively stable. Performance on each assessment tended to have the same predictive value on performance on NJASK/4. The regressions also demonstrated that as the assessments came closer to the time of taking the NJASK/4 their effectiveness in predicting proficient performance increased. This supported the findings of the correlations that were performed on the same data.

When all three assessments were combined, the resulting regression had an F value of 105.527 and was significant at the .000 level of significance with three degrees of freedom. The R^2 value of .338 indicated that the combined assessments accounted for almost 34% of the variance in performance on the NJASK/4. The Betas for individual *Everyday Mathematics* assessments decreased to .210, .207 and .334, indicating that proficient performance on all of the *Everyday Mathematics* assessments was a better indicator of successful performance on the NJASK/4 than individual performance on any one of the assessments. This finding was also reinforced by the percentage of students who performed at proficient levels on all four assessments as indicated on the Cross Tabulation analysis.

Therefore, in response to Research Question #1, "What is the relationship between *Everyday Mathematics* assessments (EMA) and performance on the NJASK/4 in Mathematics?" the findings indicated that students who performed at proficient levels on the EMA tended to perform at proficient levels on the NJASK/4. Both the proficiency levels and the scaled scores were positive indicators of performance even though their relationships did not fulfill the hypothesis of a .7 correlation.

The study then turned to the second Research Question, "What relationship do the *Everyday Mathematics* assessments have with performance on the NJASK/4 for all NCLB subgroups: ethnicity, special education classification, limited English proficiency, and socioeconomic status?"

The study results showed that the sample school district had a relatively homogeneous population for most of the NCLB subgroups. Consequently, with the exception of special education classification, the numbers of students in these NCLB categories (limited English proficiency, economically disadvantaged and some ethnic groups) were too small to produce reliable correlation analyses.

However, there were sixty-three special education students considered in the study. This number represented about 10% of the sample population. The correlation coefficient for special education was .111 at the .005 level of significance, indicating that for this subgroup there was a minimal relationship between performance on the *Everyday Mathematics* assessments and performance on the NJASK/4.

To further understand the relationship between special education classification and NJASK/4 performance, a descriptive analysis using cross tabulation was run. Since all students whose data was used in the study had been instructed with the *Everyday Mathematics* program for at least two years before the study, the *Everyday Mathematics* assessments were considered a constant for this analysis. The intent in this analysis was to see the relationship between special education and regular education students' performance on NJASK/4 when both groups had been instructed with the same program.

Table 7:
Relationship Between Special Education Classification and Performance on the NJASK/4 in Mathematics

		NJASK/4—Mathematics			
		Partially Proficient	Proficient	Advanced Proficient	Total
Special Ed.	Count	20	31	12	63
	Expected	8	28.8	26.2	63
	Standardized Resid.	4.2	.4	-2.8	
Regular Ed.	Count	59	254	247	560
	Expected	71	256	232.8	560
	Standardized Resid.	−1.4	-.1	.9	

In this Cross Tabulation measuring the association between special education classification and performance on the NJASK/4, the results indicated a discrepancy between special education and regular education students at the upper and lower ends of the performance spectrum. Thirty-two percent of special education students performed at a partially proficient level as compared to 11% of their regular education counterparts. At the same time, 19% of special education performed at the advanced proficient level as compared to 44% of regular education students. This finding was further reinforced by the fact that two cells in the cross tabulation had standardized residuals higher than 2:

NJASK/4 partially proficient/special education and NJASK/4 advanced proficient/special education. A standardized residual of 2 or higher indicates that the specific cell results have a strong effect on the association being measured (Sattler, 1992). The NJASK/4 partially proficient/special education standardized residual of 4.2 indicated that special education students had a stronger likelihood of performing at a partially proficient level than regular education students. Similarly, the NJASK/4 advanced proficient/special education residual of –2.8 also indicated that special education students were less likely to perform at the advanced proficient level than regular education students.

Although the limited English proficient (LEP) and economically disadvantaged (ED) NCLB categories had small numbers, LEP =12 and ED =26, an overview of their performance percentages, showed similar results to that of special education students.

Table 8:
Comparison of Student Subgroups Performing at Passing Levels on the NJASK/4

Percentage of students performing at passing levels			
Special Ed.	68%	Regular Ed.	89%
LEP	58%	Non-LEP	88%
ED	77%	Non-ED	88%

Forty-two percent of limited English proficient students performed at the partially proficient level on the NJASK/4 as compared to 12% of non-LEP students. Sixteen percent of LEP students performed at the advanced proficient level versus 42% of non-LEP students. In all, 58% of limited English proficient students performed at a passing level on NJASK/4 compared to 88% of English speaking students.

Similarly, of the twenty-six economically disadvantaged students, 23% performed at the partially proficient level compared to 12% of non-economically disadvantaged students. Conversely, 12% of economically disadvantaged students performed at the advanced proficient level versus 43% of students who were not identified as economically disadvantaged. In all 77% of economically disadvantaged students performed at a passing level on the NJASK/4 as compared to 88% of non-economically disadvantaged students. Therefore, students in all three NCLB categories: special education classification, limited English proficiency, and economically disadvantaged tended to have higher percentages of students who were unable to pass the NJASK/4 as compared to their regular education, English speaking, non-economically disadvantaged counterparts.

It was interesting to note that the percentage of students performing at passing/proficient levels remained consistent for all non-NCLB subgroups whereas the percentages for these NCLB subgroups were not only lower but varied considerably by subgroup. It should also be noted that students in these NCLB subgroups are often double or triple counted as they may fit into two or more categories for NCLB accountability purposes (NCLB, 2004).

The cross tabulation results with regard to ethnicity were even more definitive. Only three subgroups were reported due to the homogeneous nature of the district and the small numbers of several ethnic subgroups other than Whites and Asians. The three subgroups investigated were: Whites, Asians and a combined group of all other ethnicities including African-American, Hispanic and Native American. This combined ethnic group consisted of only 48 students out of the 623 valid cases. Hispanic students comprised 75% of this group while African-Americans accounted for 24%. Native Americans were 1% of this combined ethnic group.

Table 9:
Cross Tabulation Representing the Relationship Between Ethnicity and Performance on the NJASK/4 for Mathematics

		NJASK/4—Mathematics			
Ethnicity		Partially Proficient	Proficient	Advanced Proficient	Total
White	Count	52	219	180	451
	Expected	57.1	206.7	187.2	451
	Standardized Resid.	-.7	.9	-.5	
Asian	Count	11	45	69	125
	Expected	15.8	57.3	51.9	125
	Standardized Resid.	-1.2	-1.6	2.4	
All Other Ethnic Groups	Count	16	22	10	48
	Expected	6.1	22	19.9	48
	Standardized Resid.	4.0	0	-2.2	

Three sets of standardized residuals indicated major effects on NJASK/4 performance. The standardized residual of 2.4 for Asians/advanced proficiency indicated that Asians were more likely to perform at the advanced proficient level than other groups. Conversely, the standardized residual of –2.2 for All Other Ethnic Groups/advanced proficient indicated that students from ethnic backgrounds other than White or Asian were least likely to perform at advanced proficient levels. This was confirmed by the fact that 55.2% of Asians score advanced proficient compared to 39.9% of Whites and only 20.8% of all other ethnic groups. The standardized residual of 4.0 for All Other Ethnic Groups/partially proficient indicated that students in this subgroup (African-Americans, Hispanics, Native Americans) were more likely to perform at the partially proficient level than Whites or Asians. This was confirmed by the fact that

33.3% of students in this group scored partially proficient as compared to 11.5% of Whites and 8.8% of Asians.

When considering performance at passing levels on the NJASK/4 an even greater discrepancy could be seen. Eighty-nine percent of Whites and 91% of Asians performed at passing levels as compared to 67% of students from all other ethnic groups. This finding indicated that students from African-American, Hispanic, and Native American ethnic groups were less likely to perform at passing levels on the NJASK/4 in mathematics than their White and Asian counterparts.

ANOVA results (Appendix C) reinforced the findings of the cross tabulations with regard to special education classification and ethnicity. Again, the numbers for limited English proficient and economically disadvantaged were too small to provide reliable ANOVA results. In the ANOVA the main effects of special education classification and ethnicity were found to significantly impact NJASK/4 performance. Special education classification with an F value of 15.041 and two degrees of freedom was significant at the .000 level of significance. Ethnicity with an F value of 4.504 with two degrees of freedom was significant at the .011 level of significance.

Finally, multiple comparisons of NCLB subgroups and performance on the NJASK/4 were performed (Appendix C). Once again, the numbers for limited English proficient and economically disadvantaged students were too small to provide reliable results. However, several comparisons for special education classification and ethnicity were found. The pairwise comparison of means, based on the LSD test, yielded the following significant between group pairs: for the dependent variable of special education, partially proficient/proficient, partially proficient/advanced proficient as well as proficient/advanced proficient and for the dependent variable of ethnicity, partially proficient/proficient as well as partially proficient/advanced proficient

Special education students were 14% more likely to score partially proficient than proficient on the NJASK/4. They were 21% more likely to score partially proficient than advanced proficient and 6% more likely to score proficient than advanced proficient.

Students in all ethnic groups were 23% more likely to score proficient than partially proficient and 20% more likely to score advanced proficient than partially proficient. At first this seemed like a contradiction to other analyses, but one must consider that the variable of ethnicity covered all ethnic groups including Whites and Asians. The performance of Whites and Asians at the proficient/advanced proficient levels as shown on the Cross tabulation analyses far outweighed the poorer performance of other ethnic groups which produced this positive result.

A further informal examination of the subgroup population numbers showed that the already small subgroup sizes were further diminished by the high percentage of assessment scores that were missing. This information was determined by a simple percentage calculation. For example, of the 90 special education students in the study, 25 students representing 28% of the special education population were missing grade 3 mid-year or end-of year scores. Similarly, 29% of limited English proficient students, representing 5 out of 17 LEP students were missing grade three scores while 27% or 9 out of 33 economically disadvantaged students had grade three scores missing. When examining ethnic data, 5 out of 47 Hispanics (10.6%), 21 of 132 Asians (15.4%), 6 of 15 African-Americans (40%), and 54 of 456 Whites (11.8%) had missing scores. This finding indicated that more than one-quarter of the special education and limited English proficient students has missing scores that effectively eliminated them from participation in the study.

Summary of Findings

Based on the correlation analysis performed on the relationship between *Everyday Mathematics* assessments and performance on the NJASK/4, this study found that there was a significant, moderate, positive relationship between the assessments. The strength of the relationship increased as students moved through third grade into fourth grade. The strongest relationship was found between the scaled scores on the fourth grade, mid-year *Everyday Mathematics* assessment and the NJASK/4. This relationship was a moderate positive one. However, the hypothesis was to be accepted only if the correlation was .7 or higher which would have indicated a strong, positive relationship. Since this did not occur, the hypothesis was rejected.

However, the relationship between the *Everyday Mathematics* assessments and performance on the NJASK/4 could not be ignored. The regression analyses of the same information indicated that *Everyday Mathematics* assessments accounted for almost 34% of the variance in relationship to performance on the NJASK/4. The findings also showed that 89% of the students who performed at passing levels on the *Everyday Mathematics* assessments also passed the NJASK/4. This finding is supportive of the goals of a criterion-referenced test.

Similarly, correlations between NCLB subgroup variables and performance on the NJASK/4 indicated that although these variables had a significant relationship with performance, none had a strong correlation. However, further analysis of these same subgroup variables did show trends that would be valuable for the district to study.

Specifically, the percentages of special education, limited English proficient, economically disadvantaged, African American, Hispanic, and Native American students who performed at the partially proficient level on NJASK/4 should give rise to some questions and discussions. Particular attention should be given to those scores statistically considered to be outliers.

CHAPTER V

Summary, Conclusions and Recommendations

Introduction

The purpose of this study was to determine the effectiveness of the *Everyday Mathematics* third and fourth grade assessments in predicting performance on the 4th grade NJASK. The study also investigated if there was a relationship between performance on *Everyday Mathematics* assessments and performance on the NJASK/4 for NCLB subgroups.

This chapter contains a summary of the research findings, conclusions based on the findings and recommendations for further research on the topic of the effectiveness of *Everyday Mathematics* in predicting proficient performance on NJASK/4.

Summary

In this study based on assessment data for a sample population of 623 fourth grade students, the research found that third and fourth grade *Everyday Mathematics* assessments were moderately effective in predicting performance on the NJASK/4. Despite the high level of content validity between the *Everyday Mathematics* program and assessments and the New Jersey Core Curriculum Content Standards and the NJASK created by their mutual derivation from the National Council of Mathematics Standards, the analyses of data could not provide a strong relationship finding based on correlation data. However, the research did indicate that there were significant relationships between the two sets of assessments as well as between the proficiency levels on the NJASK and the NCLB subgroup populations. The frequency with which students performed at proficient levels on both the *Everyday Mathematics* assessments and the NJASK/4 was 89%. This frequency distribution indicated a skewed population data reflective of previous studies of the *Everyday Mathematics* program (Everyday Learning Corp., 1996; Everyday Learning Corp., 1998; Riordan & Noyes, 2001; Carroll & Isaacs, 2003; Everyday Learning Corp., 2004).

Furthermore, the research showed that a combination of the three *Everyday Mathematics* assessment results explained 34% of the relationship between the assessments and proficiency on the NJASK/4. The results were reflective of other research findings cited in the review of literature in that the specific subgroup populations of special education, limited English proficiency, economically disadvantaged, Hispanic, and African American were all more likely to score partially proficient on the NJASK/4 than their native English speaking, economically advantaged, White and Asian, regular education counterparts.

Specifically, on the Nations' Report Card in mathematics (NCES, 2003b), White students had average scores that were higher than African-American or Hispanic students' average scores. Although the gap was narrower than in the 1992 results, a gap still existed. Conversely, Asian students had higher average math scores and a larger percentage of proficient math students than any other ethnic group (NCES, 2003b). Economically disadvantaged students maintained their level of disparity that had existed in 1992 even though the actual percentage points had decreased (NCES, 2003b). The Nation's Report Card results do not address special education or limited English proficient students since these populations are excluded from their review.

Therefore, the ethnic and socioeconomic relationships found in this study reflect the findings of the Nation's Report Card which is based on the National Association of Elementary Progress's (NAEP) random testing of American students. This study's results also mirror the reports of Quality Counts (2004) and research done by Stevens (2003) and Kober (2001) who found that by the end of grade two, a greater proportion of African-American and Hispanic students had slipped below grade level as compared to their White and Asian counterparts.

Conclusions

There are several possible reasons why the seemingly high content validity between *Everyday Math* assessments and the NJASK did not result in a strong predictive relationship. First of all, there is the human factor. Perhaps one or both of the assessments did not accurately articulate the mathematics standards on which they purport to be based. The expert decision-makers who decided that the assessments reflected the NCTM standards may have used different criteria in reaching their conclusions. This is one of the difficulties with content validity. Content experts tend to take their knowledge for granted and forget that the test takers may not be as knowledgeable (Cronbach, 1971). Hence the test questions may reflect a degree of difficulty inappropriate for the level of the test taker. If this was the case then one or both of the assessments might contain questions that reflected different perspectives on the same topic.

Another problem might be that content experts failed to identify the learning objectives of a subject. In his study Popham (1975) gave examples of such questions that were clearly an evaluation of students' language and comprehension abilities rather than their knowledge of mathematics, science, or social studies content. A common problem found on tests is that students are asked to answer questions that require "memorized facts" rather than the content knowledge. For example a question that asks for Euclid's birth date does not measure mathematics content. Furthermore, sampling knowledge from a larger domain involves choices (Cronbach, 1971). Value judgments on the relative importance of each subcategory may differ from assessment to assessment.

In addition to the actual test questions, the basis for the levels of proficiency between *Everyday Mathematics* scores and the NJASK/4 scores may not be in harmony. Since the *Everyday Mathematics* proficiency levels were based on scores for specific students in one specific school district rather than a cross-section of schools, the score ranges may not be universally acceptable. The district under study has a mildly diverse ethnic and economic population. Proficiency levels were set based on the mean scores of the district's population. Camilli & Monfils (2002) tells us that test scores reflect the community and family resources available to support students' learning inside and outside of school. Therefore, the data basis between the *Everyday Mathematics* assessments and NJASK/4 may differ.

Finally, despite seemingly strong content validity between *Everyday Mathematics* and NJASK/4 assessments and research supporting the effectiveness of *Everyday Mathematics*, one program should not be considered the answer to all students' mathematical needs. Similarly one test such as NJASK/4 cannot reflect the total knowledge and understanding of all students. The combined effect of the three *Everyday Mathematics* assessments in this study accounted for 34% of the variance in NJASK/4 scores. Other factors accounted for the balance of the variance. Obviously, the NCLB subgroups are some of the factors that influenced test results. According to research done by Firestone, Monfils, & Schorr (2003), other factors that influence ESPA mathematics scores are pressure and district factor grouping (DFG). Pressure is associated with short-term responses to the test and, to some extent, to more didactic instruction. This study did not include teacher interviews so the extent of the influence of pressure on instruction and testing was not measured, but could be a focus for further study.

According to Firestone et al. (2003), the poorer the DFG, the more teaching to the test occurs. Again, teacher surveys or observations would be needed to determine the extent of this influence. However, since the district DFG is I, the second highest category, it would seem, based on Firestone's study, that teaching to the test would not be a major influence in this case. However, whatever factors are studied, the *Everyday Mathematics* assessments still describe one-third of the relationship between *Everyday Mathematics* assessments and NJASK/4 results.

The reality in this study was that such a large percentage of students (89%) performed proficiently on both the *Everyday Mathematics* assessments and the NJASK/4 that the sample data was skewed. With a skewed population of this magnitude, the correlation would be low since the outcomes would not be evenly distributed along a "line of best fit."

Such a finding is not inconsistent with the goals of a criterion referenced test such as NJASK/4. Criterion-referenced tests are intended to measure how well a person has learned a specific body of knowledge and skills (Sattler, 1992). Standards-referenced tests are a form of criterion-referenced test based on "content standards" or curriculum frameworks. In a criterion-referenced test, it is possible for all students to earn a passing score if they have learned the body of material tested. In contrast, norm-referenced tests are made to compare test takers to each other (Sattler, 1992). Test takers would be compared to each other based on who knew most or least about the subject matter. The principle of a norm-referenced test implies that half of the test takers would score below a midpoint and half would score above.

In drawing further conclusions from the statistical findings, the limitations of the study need to be considered. Although the sample size of 623 was ample, the sample numbers were skewed toward White and Asian, native English speaking, regular education students from an economically advantaged environment. Numbers for the African-American, Hispanic, limited English proficient, and economically disadvantaged subgroups were low. Only the initial special education population of 90 students provided a reasonable sample investigation. The smaller the number in a sample, the more likely that the statistical analysis does not give an accurate result (Sattler, 1992). Therefore, the overall relationship between *Everyday Mathematics* assessments and NJASK/4 in this study was influenced by ethnic and socioeconomic subgroups that traditionally do well on standardized tests (NCES, 2003b). Again, teacher interviews would be needed to determine why subgroups, that were already few in number, were further reduced by missing test information. These scores may be missing because the assessments were not given. On the other hand, the assessments may have been given but in a different time frame. Perhaps some third grade teachers felt that students had not learned sufficient material to take the assessment at the prescribed time and administered the test at a later date. Some students were obviously new to the district in the fourth grade. Whatever the reasons, approximately one-quarter of the at-risk populations of special education, limited English proficient, and economically disadvantaged, and 40%

of African-Americans had missing data. Such lack of information could have affected data analysis. Furthermore, if the reason for missing grade three data was that students were not prepared to take the *Everyday Mathematics* assessments at the prescribed times then these students might be at a disadvantage when taking the NJASK/4 unless their fourth grade teachers made up the gap in teaching and learning.

Another factor of the research that needs to be considered is that the study only investigated the relationship between one mathematical program, *Everyday Mathematics*, and the New Jersey state assessment, NJASK/4. The results of the study did not indicate that *Everyday Mathematics* assessments were more or less effective than any other mathematics program assessment in predicting proficient performance on NJASK/4. The study only indicated that for this specific research population over a specific two-year interval, *Everyday Mathematics* assessments were moderately effective in predicting proficient performance on NJASK/4. Within the moderate level of effectiveness, the findings indicated that a higher percentage of Whites and Asians scored proficient on the mathematics portion of the NJASK/4 than African-Americans or Hispanics. Similarly, more special education, limited English proficient, and economically disadvantaged students scored partially proficient than their economically advantaged, English speaking, regular education counterparts. Since these findings were also true for the *Everyday Mathematics* assessments, it would be reasonable for the district to assume that if a student, especially one in an at-risk NCLB subgroup, scores partially proficient on an *Everyday Mathematics* assessment, there is a 60-40 chance that the same student will score partially proficient on the NJASK/4 based on the correlation levels of actual scores.

Recommendations

Based on the findings and conclusions of the research study, several recommendations can be made. First, to determine the validity and reliability of these findings, further studies should be done with other schools/districts with similar and dissimilar populations to see if this study's findings can be replicated. In this way researchers can determine if the findings of this study are an anomaly or if they are representative of other studies' findings. There are no studies that specifically investigate the relationship between *Everyday Mathematics* assessments and the NJASK/4. Such a study would be of interest because an increasing number of school districts are using *Everyday Mathematics* as their scientifically-based mathematics program. In a recent informal survey, conducted by the County Board of Education, of 15 districts polled in Somerset County, 12 or 80% were using *Everyday Mathematics*.

District superintendents need to be aware of such trends and question them since they are accountable for how they manage a district's limited financial resources. Superintendents need to know that the curriculum materials and the staff development used in the district result in high levels of student achievement. The superintendents and district curriculum directors/supervisors must consider whether any one nationally developed and marketed mathematical program/product can have a tight fit with each specific state's standards' requirements. Furthermore, they need to assess whether products and programs meet the needs of all students in their districts. No matter which mathematical program districts choose, there will always be a need to augment the program with proven instructional strategies and solid support for students who may struggle with a particular program's methodology. Therefore, this study's results should be used to investigate the needs of low–performing students and provide additional and/or differentiated instruction and materials as needed. Considering the skewed distribution of test data, particular attention should be paid to students whose scores were in the "outlier" areas.

A second recommendation would be to perform this study in other New Jersey schools/districts with mathematical programs other than *Everyday Mathematics*. Researchers could then compare the findings of the districts to see which mathematical assessments were more effective in predicting proficient performance on the NJASK/4. Although there have been a number of studies that demonstrate that students taught with *Everyday Mathematics* do better on standardized tests than students taught with other programs (Everyday Learning Corp., 1996; Everyday Learning Corp., 1998; COMAP, 2000; Riordan & Noyes, 2001; Carroll & Isaacs, 2003; Everyday Learning Corp., 2004), none of these studies have been conducted in New Jersey or related to NJASK/4.

Another recommendation would be that researchers investigate other variables that might influence performance on the NJASK/4. Specific attention might be given to teacher training and implementation of the mathematics program.

Some questions to be asked might be--Is training ongoing or is it only provided during initial implementation of the program? If teachers provide supplemental instruction, is it supportive of the primary program's goals and students' needs? Do teachers and supervisors use student performance data to inform instruction? In addition, researchers might investigate if the written curriculum and the taught curriculum are congruent. They might also investigate whether teachers' expectations are equally high for all students especially NCLB subgroup populations.

[redacted]

"safe harbor" provisions for NCLB subgroups (NCLB, 2001; NJDOE, 2004b), advocates for special education students still feel that more consideration needs to be given to classified students' needs. The subject district should be aware if this perspective is influencing the mathematics instruction in its schools.

If teachers of special education students are not maintaining the same pace of instruction as their regular education counterparts then they are placing special education students at further risk for failure on New Jersey's standardized assessments. Until laws are changed and special education students may attain standards' requirements according to their own IEP limitations, they must be adequately prepared to meet state assessment standards.

With respect to establishing proficiency levels for the *Everyday Mathematics* assessments at the district level, the district might consider establishing an appropriate tuning protocol for examining student work on the various *Everyday Mathematics* assessments. Such a practice might provide a more valid scaled score range that other districts could replicate.

References

Anastasi, A. (1959). *Psychological testing.* New York: The Macmillan Company.

Artzt, A. & Armour-Thomas, E. (1999). A cognitive model for examining teachers' instructional practices in mathematics: A guide for facilitating teacher reflection. *Educational Studies in Mathematics*, *40*, 211-23.

Barsenic, J., & Gelman, D. (2002). Standardized testing: How will the pendulum swing. *Streamlined Seminar*, *19*, 1-4.

Barton, P. (2003). Parsing the achievement gap. Report for the Policy Information Center of the Educational Testing Service. Princeton: ETS

Battista, M. T. (1994). *The mathematical mis-education of America's youth Phi Delta Kappan, 80*(6), 424-433.

Bell, M.S., & Bell, J. B. (1988). Assessing and enhancing the counting and numeration capabilities and basic operation concepts of primary school children. Chicago: UCSMP.

Blank, R. (2003). NCLB: The challenge to states. Report to the Council of Chief State School Officers. Philadelphia: CCSSO.

Borman, G. D. (2003) How can title I improve achievement. *Educational Leadership*, *60*, 49-53.

Bureau of U.S. Citizenship and Immigration Services. (2001*). Immigrants, fiscal year 2001*. Washington, DC: Author.

Burns, J. J. (2003). Standards and assessment—the AYP challenge. *School Leader, 33* (3), 14-16.

Burton, C. (2003). *Comprehensive Assessment Study.* Unpublished manuscript, Seton Hall University: South Orange, NJ.

Bybee, R.W. (1997). *The sputnik era: Why is this educational reform different from all other reforms.* Paper presented to the Center for Science, Mathematics and Engineering Education of the National Academy of Sciences. Washington, DC.

Camilli, G., & Monfils, L. (2002). *Test scores and equity in New Jersey.* New Orleans, LA: Paper presented at AHERA conference

Campbell, F. A., & Ramey, C.T. (1995). Cognitive and school outcomes for high-risk African-American students in middle adolescence: Positive effects of early intervention. *American Educational Research Journal, 32*(4), 743-772.

Carroll, W. M. (2000). *A longitudinal study of children in the Everyday Mathematics Curriculum.* Chicago: UCSMP.

Carroll, W., Bell, M., & Isaacs, A. (2001). *A research based curriculum: The research basis of the UCSMP Everyday Mathematics curriculum (UCSMP Elementary Component).* Chicago: Everyday Learning Corporation.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mod, A., Weinfield, F.D. & York, R. (1966). *Equality of educational opportunity.* Washington, DC: Department of Health, Education and Welfare.

Commission on Behavioral and Social Science and Education (CBASSE). (2002). *Minority students in special and gifted education.* Washington, DC: National Academies Press.

Consortium for Mathematics and its Applications (COMAP). (2000). *The ARC center tri-state student achievement study: use of everyday mathematics, investigations in number, data and space and math trailblazers by elementary schools in the U.S.* Lexington, KY: Author

Council of Chief School Officers (2002). *Expecting success: self-assessment and resources guide for improving service delivery.* Washington, DC, author.

Cronbach, L.J. (1971). Test Validation. In R. L. Thorndike (ed.) *Educational Measurement.* Washington, DC: American Council of Education.

Darling-Hammond, L., & Falk, B. (1997). Using standards and assessments to support student learning. *Phi Delta Kappan*, *79*(3), 190-199.

Darling-Hammond, L. (1998). Unequal opportunity: Race and education. *The Brookings Review, 16* (2), 28-32.

Education Commission of the States (2004). State implementation of the no child left behind act: Respecting diversity among states. ECS report to the nation. Denver, CO.

Elmore, R.F. (2003). A plea for strong practice. *Educational Leadership, 61* (3), 6-11.

Everyday Learning Corporation. (1996). *Everyday Mathematics: Student achievement studies.* Chicago: Author.

Everyday Learning Corporation. (1998). *Everyday Mathematics gets results: Student achievement studies,* Volume 2. Chicago: Author.

Everyday Learning Corporation, (2004). *Everyday Mathematics: Student achievement studies, Volume 3.* Chicago: Author.

Fallon, D. (2003). Case study of a paradigm shift. Presentation to the Education Commission of the States. Fall Steering Committee meeting, November 12, 2003, Richmond, VA.

Ferguson, R. (1998). Teachers' perceptions and expectations and the black-white test score gap. In C. Jencks and M. Phillips (Eds). *The black-white test score gap.* Washington, DC: Brookings Institute.

Ferguson, R. (2002).*What doesn't meet the eye: Understanding and addressing racial disparities in high-achieving suburban schools.* Naperville, IL: North Central Regional Educational Laboratory (NCREL), Learning Point Associates.

Finn, J, & Achilles, C. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal, 27*, 555-577.

Firestone, W., Monfils, L., & Schorr, R. (2003). *Test preparation in New Jersey: inquiry-oriented and didactic responses.* Paper funded by the National Science Foundation grants 9804925 and 9980458. New Brunswick, NJ: Rutgers University.

Fullan, M. (2000). The three stories of education reform. *Phi Delta Kappan, 81*(8), 581-584.

Gingerich, D. (2004). The special education challenge. *Research for Better Schools: Currents, 7*(2), 1-4.

Goals 2000: Educate America Act, Title III, Sec.302.

Goldberg, M. F. (1996). Here for the long haul: An interview with Ted Sizer. *Phi Delta Kappan, 77(*10), 685-687.

Haycock, K. (2001). Closing the achievement gap. *Educational Leadership, 58*, 6-11.

Hiebert, J. (1999). Relationship between research and the NCTM standards. *Journal for Research in Mathematics Education, 30* (1), 3-19.

Hiebert, J. (2003). Teaching mathematics in seven countries: Results from the TIMSS 1999 video study. Washington, DC: National Center for Educational Statistics: United States Department of Education.

Jencks, C. S. (2003). Educational research and educational policy: An historical perspective. In *Analytic Issues in Assessment of Student Achievement,* 167, Section III, 279-297. Boston, MA: Harvard University.

Jencks, C. S. & Phillips, M. (1998). The black-white test score gap: Why it persists and what can be done. *The Brookings Review, 60* (2), 24-27.

Jencks, C. S., Smith, M., Acland, H., Bone, M. J., Cohen, D., Ginits, H., Heyns, B., & Michelson, S. (1972). *Inequality: A reassessment of the effect of family and schooling in America.* New York, NY: Basic Books.

Jerald, C. (2003). Beyond the rock and the hard place. *Educational Leadership, 61*(3), 12-17.

Jones, K., & Whitford, B. L. (1997). Kentucky's conflicting reform principles. *Phi Delta Kappan,79* (94), 276-287.

Kenschaft, P.C. (2004). *Fifty-five cultural reasons why too few women win at mathematics.* Retrieved May 28, 2004 from the World Wide Web: http://www.enc.org/topics/equity/articles/document.shtm

Knapp, M. S., & Sields, P. M. (2004). *Reconceiving academic instruction for the children of poverty.* Retrieved on May 28, 2004 from the World Wide Web: http://www.enc.org/topics/equity/articles/document.shtm

Kober, N. (2001) *It takes more than testing: Closing the achievement gap.* Washington, DC: Center on Educational Policy.

Kohn,A. (2000, December). *The case against tougher standards.* Remarks presented at the winter meeting of the Morris County Elementary and Middle School Administrators, Madison, NJ.

Kunz, M. (2003). New Jersey's standards and assessments. *School Leader*, *33*(3) 17-27.

Labaree, D. F. (2000). Resisting educational standards. *Phi Delta Kappan, 82(*1), 28-33.

Lally, M. O. (2000). *A construct validity study of the New Jersey assessment (EWT), grade eight proficiency assessment (GEPA) and elementary school proficiency assessment (ESPA).* Unpublished dissertation. Seton Hall University, South Orange, NJ.

Lappan, G. (2000). A vision of learning to teach for the 21st century. *School Science and Mathematics, 100*(6), 319-326.

Lee, H. J. (2001). Enriching the professional development of teachers. *ERIC Digest* ED465495. ERIC Clearinghouse for Science, Mathematics and Environmental Education.

Lewis, A. C. (1995). An overview of the standards movement. *Phi Delta Kappan, 76*(10), 744-750.

Li, Q. (1999). Teachers' beliefs and gender difference in mathematics. *Educational Research, 41*(1), 63-76.

Marzano, R. (2000, December). *Awash in a sea of standards.* Remarks presented at the annual conference of the New Jersey Association for Supervision and Curriculum Development, East Windsor, NJ.

McCloskey, W., & McMunn, N. (2000). Strategies for dealing with high-stakes state tests. *Phi Delta Kappan, 82*(2), 115-120.

Messick, S. (1995). Validity of psychological assessment: validity of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

Mid-Continental Research for Education and Learning (McREL). (2003, Fall). *Changing Schools,* Aurora, CO: Author.

Murphy, J., & Doyle, D. (1999). No finish line. *Journal of Staff Development, 20*(1), 14-18.

A Nation at Risk. (1983). Retrieved December 4, 2003, from the World Wide Web: http://www.ed.gov/pubs/NatAtRisk/risk.html

National Academy of Sciences. (2000). *Final report of the coordinating committee for the national council of teachers of mathematics.* Project Scope Statement, NRC Project No. MSEB-S-99-02-A. Washington, DC.

National Assessment Governing Board. (2002). *Mathematics framework for the. 2003 national assessment of educational progress.* Washington, DC: United States Department of Education

National Center for Educational Statistics (NCES). (1994). *America's mathematics problem: rising student achievement, a synthesis of findings from NAEP's 1992 mathematics assessment. (Report No. 23-FR-03).* Washington, DC: United States Department of Education, Office of Educational Research and Improvement.

National Center for Educational Statistics. (NCES). (1997). *Pursuing excellence: a study of U.S. fourth grade mathematics and science achievement in international context. (Report No. 97-255).* Washington, DC: United States Department of Education.

National Center for Educational Statistics (NCES). (1998a). *Long term trends in student mathematical performance. (Report No. NCES 98-462).* Washington, DC: United States Department of Education.

National Center for Educational Statistics (NCES). (1998b). *Report in brief: NAEP trends in academic progress (revised). (Report No. NCES 98-530).* Washington, DC: United States Department of Education, Office of Educational Research and Improvement.

National Center for Educational Statistics (NCES). (2003a). *The condition of education 2003. (NCES 2003-067).* Washington, DC: United States Department of Education, Institute of Education Sciences.

National Center for Educational Statistics (NCES) (2003b). *The nation's report card.* Retrieved August 16, 2004 from the World Wide Web: http://nces.ed.gov/nationsreportcard/mathematicsresults2003

National Council of Teachers of Mathematics (NCTM). (1989). *Principles and Standards for School Mathematics: Curriculum and Evaluation.* Retrieved October 18, 2003 from the World Wide Web: http://standards.nctm.org.

National Council of Teachers of Mathematics (NCTM). 1991. *Professional Standards.* Retrieved October 18, 2003 from the World Wide Web: http://standards.nctm.org.

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and Standards.* Retrieved October 18, 2003 from the World Wide Web: http://standards.nctm.org.

National Education Summit Policy Statement. (1996). Retrieved on September 1, 2004 from the World Wide Web: www.edweek.com/ew/ewstory.cfm

National Science Foundation (NSF). (1983). *Educating Americans for the 21st century: A report to the American people and the national science board.* Washington, DC: National Science Board Commission on Precollege Education in Mathematics, Science and Technology.

Nave, W., Miech, E., & Mosteller, F. (2000). A lapse in standards: Linking standards based reform with student achievement. *Phi Delta Kappan, 82*(2), 128-133.

New Jersey Department of Education (NJDOE). (1974). District Factor Groups. Retrieved on July 5, 2004 from the World Wide Web: www.state.nj.us/njded/schools/achievement/dfg.htm.

New Jersey Department of Education (NJDOE). (1995). *The New Jersey Mathematics Curriculum Framework.* Trenton, NJ: Author.

New Jersey Department of Education (NJDOE). (1996). *New Jersey Core Curriculum Content Standards.* Trenton, NJ: Author.

New Jersey Department of Education (NJDOE). (2002) NCLB State report card. Trenton, NJ: Author.

New Jersey Department of Education (NJDOE). (2003). *Report on NAEP results for 2003 sent to all public school districts on November 14, 2003.* Trenton, NJ: Author.

New Jersey Department of Education (NJDOE). (2004a). *New Jersey Core Curriculum Content Standards* (revised). Trenton, NJ: Author.

New Jersey Department of Education (NJDOE). (2004b). *Final 2003 AYP data,* Trenton, NJ: Author.

New Jersey Department of Education (NJDOE). (2004c). *Linking the 1996 New Jersey core curriculum content standards and cumulative progress indicators in mathematics with the corresponding 2002 indicators.* Trenton, NJ: Author.

New Jersey Department of Education (NJDOE). (2004d). *New Jersey Assessment of Skills and Knowledge* (NJASK). Trenton, NJ: Author.

No Child Left Behind Act. (2001). *PL107-110.* Washington, DC: United States Department of Education.

No Child Left Behind Act. (2004). *PL107-100, revised.* Washington, DC: United States Department of Education.

Noddings, N., (1997). Thinking about standards. Phi Delta Kappan, *79*(3), 184-189.

Olsen, L. (2004). Enveloping expectations. *Education Week*, *23*(17), 8-20.

O'Neil, J., & Tell, C. (1999). Why students lose when tougher standards win: A conversation with Alfie Kohn. *Educational Leadership 57*(1). Available: www.ascd.org.

Pasi, R. (2000). No easy answers. *Educational Leadership, 57*(5), 75-76.

Payne, R. (2003). No child left behind what's really behind it? *Educational Viewpoints, 23*(1), 10-12.

Popham, W. J. (1975). *Educational Evaluation.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Popham,W. J. (1999). Why standardized tests don't measure educational quality. *Educational Leadership*, *57*(5), 9-15.

Protheroe, N., Shellard, E. & Turner, J. (2003). *A practical guide to school improvement: meeting the challenges of NCLB.* Arlington: Educational Research Service.

Quality Counts. (2004). Count me in: Special education in an era of standards. Retrieved on March 14, 2004 from the World Wide Web: http://counts.edweek.org

Ravitch, D. (1995). *National Standards in American Education.* Washington, DC: Brookings Institute Press

Riordan, J. E. & Noyes, P. E. (2001). The impact of two standards based curricula on student achievement in Massachusetts. *Journal for Research in Mathematics Education, 32*, 368-398.

Sanders, W. L. and Rivers, J. C. (1996). Teacher quality and equity in educational opportunity: Findings and policy implications. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Sattler, J. M. (1992). *Assessment of Children.* San Diego, CA: Jerome. M. Sattler Publications, Inc.

Schmidt, W. H., McKnight, C. C., & Raizen, S.A. (1996). Splintered vision: An investigation of United States science and mathematics education: Executive summary. Lansing, MI: U.S. National Research Center for the Third International Mathematics and Science Study.

Schmoker, M. (2000). The results we want. *Educational Leadership*, *57*(5), 62-65.

Schmoker, M. & Marzano, R. J. (1999). Realizing the promise of standards-based education. *Educational Leadership, 56*(6), 17-21.

Schwartzbeck, T. D. (2004). Implementing NCLB: Measuring subgroup performance. Philadelphia PA: P.resentation to the New Jersey Association of School Administrators.

Secretary's Commission on Achieving Necessary Skills (SCANS). (1991). *What work requires of schools.* Report to the Secretary of Labor. Springfield, VA: National Technical Information Service NTIS PB92-1467111NZ.

Slavin, R. (1998). Can education reduce social inequity. *Educational Leadership, 55* (4), 6-10.

Slavin, R. E. & Madden, N. A. (1999). *Success for all/roots and wings: 1999 summary of research on achievement outcomes*. Retrieved August 14, 2004 from the World Wide Web: http://www.successforall.net/resource/resesarch/report41entire.pdf

Smith, M. & O'Day, J. (1991). Systemic school reform. In S. Fuhrman and B. Malen (Eds.)., *The politics of curriculum and testing* (pp. 233-265). Philadelphia PA: Falmer Press.

Stevens, J.H. (2003). Statement of NAEP 2003 mathematics and reading results. Released for publication on November 13, 2003. Washington, DC: National Assessment Governing Board.

Strauss, M.. (1991). *Gender bias in mathematics, science and technology: the report card #3.* Washington, DC: Mid-Atlantic Equity Center.

Taylor, W. L. (2000). Standards, tests and civil rights. *Education Week,* 20, 56.

Third International Mathematics and Science Study (TIMSS). (1995). Lansing MI: International Research Center at Michigan State University.

Third International Mathematics and Science Study (TIMSS). (1999).Lansing MI: International Research Center at Michigan State University.

Tomlinson,C.A. (2000). Reconcilable differences? Standards-based teaching and differentiation. Educational Leadership, *58*(1), 6-11.

United States Department of Education (USDOE). (1965). The Elementary and Secondary Education Act of 1965. Retrieved on May 8, 2004 from the World Wide Web: http://cwx.prenhall.com.

United States Department of Education (USDOE). 1994. Improving America's Schools Act, Public Law 103-382. Retrieved on April 4, 2004 from the World Wide Web: www.ed.gov/legislation/ESEA/toc.html.

United States Department of Education (USDOE). (2002). *No Child Left Behind: A Desktop Reference.* Jessup, MD: Education Publications Center.

United States Department of Education (USDOE) (2003). *Why no child left behind. is important to America: Federal spending on K-12 education.* Retrieved March 29. 2004 from the World Wide Web: http://www.ed.gov./nclb/overview/importance/edlite-index.html

United States Department of Education (USDOE). (2004). *History of the IDEA..* Retrieved on August 12, 2004 from the World Wide Web: www.ed.gov/policy/speced/leg/idea/history/html

University of Chicago School Math Project. (1986). *Everyday Mathematics.* Chicago IL: Author.

Viadero, D. (2003). *Study probes factors fueling achievement gaps.* Retrieved on May 8, 2004 from the World Wide Web: http://www.edweek.org.

Viadero, D. (2004). *Disparately disabled.* Retrieved on August 8, 2004 from the World Wide Web: http://www.edweek.org/sreports/qc04/article.cfm

Appendices

Appendix A

Charts

Chart 1: Timeline of Events and Legislation Relating to Equity and Excellence in Education

DATE	EVENT/LEGISLATION	OUTCOME
1837	Horace Mann selected Secretary of Massachusetts Board of Education	Sets goal of universal education through elementary grades
1918-1954	Compulsory Education Legislation	All states eventually enact statues requiring children of certain ages to attend school
1954	Brown v. Board of Education	Goal: Educational opportunity and access for African-American students
1965	Elementary and Secondary School Act (ESEA)	Goal: Equal access and treatment of poor and minority students
1966	Migrant Education Act	Goal: Equal access and service for children of migratory farm workers and fishermen
1967	Title IX of ESEA	Goal: Equal treatment of girls in education and sports
1974	Individuals with Disabilities Education Act (IDEA)	Goal: Equal access and treatment of students with disabilities
1982	Plyer v. Doe	Goal: Access to public schools for the children of illegal immigrants
1987	McKinney Homeless Act	Goal: Assures access to public schools for homeless children
2001	No Child Left Behind Act	Goal: Universal Proficiency

Chart 11: Comparison of 2004 Revised New Jersey Core Curriculum Content Standards (NCCCS) for Mathematics and the *Everyday Mathematics* (EM) Curriculum Strands

NCCCS-2004	EM Strands
4.1-Number and Numerical Operations	*1. Numeration and Order*
A. Number Sense	A. Number Sense and Numeration
B. Numerical Operations	B. Numerical and Algebraic Operations & Analytical Thinking
C. Estimation	
4.2-Geometry and Measurement	*2. Geometry and Measurement*
A. Geometry Properties	A. Geometry and Spatial Sense
B. Transforming Shapes	B. Measures and Measurement
C. Coordinate Geometry	
D. Units of Measurement	
E. Measuring Geometric Objects	
4.3-Patterns and Algebra	*3. Patterns, Functions and Sequences*
A. Patterns and Relationships	A. Patterns, Relationships, Functions
B. Functions	B. Algebra and Uses of Variables
C. Modeling	
D. Procedures	
4.4-Data Analysis, Probability and Discrete Mathematics	*4. Data and Chance*
A. Data Analysis (statistics)	A. Data Analysis and Statistics
B. Probability	B. Probability and Discrete Math
C. Discrete Math-systematic listing & counting	
D. Discrete math-vertex-edge graphs & algorithms	
4.5-Mathematical Processes	*5. Curriculum Features*
A. Problem Solving	A. Real-life Problem Solving
B. Communication	B. Emphasis on Communication
C. Connections	C. Balanced Instruction
D. Reasoning	D. Multiple Methods for Basic Skills Math
E. Representations	E. Enhanced Home School Partnerships
F. Technology	F. Appropriate Use of Technology

Appendix B

Explanation of Variables

Variables

id.....identification number

gender.....1= males, 2 = females

ethnicity.....1 = Whites, 2=Asians, 3=African- Americans, 4=Hispanics, 5=Native Americans

sped.....special education classification, 1=special ed, 2=regular ed

lep.....limited English proficient, 1=limited English, 2=native American speaker

ses.....socioeconomic status, 1=economically disadvantaged, 2=economically advantaged

em3mysc.....*Everyday Mathematics*, grade 3 mid-year test scores

proflev.....proficiency levels for *Everyday Math* grade 3 mid year test, 1=partially proficient, 2=proficient, 3=advanced proficient

em3eysc.....*Everyday Mathematics,* grade 3 end-of-year test scores

prflev.....proficiency levels for *Everyday Math* grade 3 end-of-year test, 1=partially proficient, 2=proficient, 3=advanced proficient

em4mysc.....*Everyday Mathematics* grade 4 mid-year test scores

prolev.....proficiency levels for *Everyday Math* grade 4 mid-year test, 1=partially proficient, 2=proficient, 3=advanced proficient

njask4sc.....NJASK/4 test scores

profflev.....proficiency levels for NJASK/4, 1=partially proficient, 2=proficient, 3=advanced proficient

tp1.....*Everyday Mathematics*, grade 3 mid-year proficient and advanced proficient combined

tp2..... *Everyday Mathematics*, grade 3 end-of-year proficient and advanced proficient combined

tp3......*Everyday Mathematics*, grade 4 mid-year proficient and advanced proficient combined

tp4.....NJASK/4 proficient and advanced proficient combined

ethnic2.....ethnicity, 1=Whites, 2=Asians, 3="nonwhas" (African-Americans, Hispanics, Native Americans)

Appendix C

Data Analyses

Correlation for Table 2: *Correlation Between Proficiency Levels on Everyday Mathematics Assessments (EMA) and NJASK/4 Math (page 66)*

Nonparametric Correlations

Correlations

			PROFLEV	PROFFLEV
Spearman's rho	PROFLEV	Correlation Coefficient	1.000	.408(**)
		Sig. (2-tailed)	.	.000
		N	628	624
	PROFFLEV	Correlation Coefficient	.408(**)	1.000
		Sig. (2-tailed)	.000	.
		N	624	624

** Correlation is significant at the 0.01 level (2-tailed).

Correlations

			PRFLEV	PROFFLEV
Spearman's rho	PRFLEV	Correlation Coefficient	1.000	.433(**)
		Sig. (2-tailed)	.	.000
		N	624	623
	PROFFLEV	Correlation Coefficient	.433(**)	1.000
		Sig. (2-tailed)	.000	.
		N	623	624

** Correlation is significant at the 0.01 level (2-tailed).

Correlations

			PROLEV	PROFFLEV
Spearman's rho	PROLEV	Correlation Coefficient	1.000	.480(**)
		Sig. (2-tailed)	.	.000
		N	625	624
	PROFFLEV	Correlation Coefficient	.480(**)	1.000
		Sig. (2-tailed)	.000	.
		N	624	624

** Correlation is significant at the 0.01 level (2-tailed).

PROFLEV----- Proficiency levels for grade 3 mid-year EMA
PRFLEV--------Proficiency levels for grade 3 end-of-year EMA
PROLEV------- Proficiency levels for grade 4 mid-year EMA
PROFFLEV----Proficiency levels for NJASK/4

Correlation for Table 3*: Correlation Between Scaled Scores on Everyday Mathematics Assessments (EMA) and NJASK/4 Math Scores (page 68)*

Nonparametric Correlations

Correlations

			EM3MYSC	NJASK4SC
Spearman's rho	EM3MYSC	Correlation Coefficient	1.000	.537(**)
		Sig. (2-tailed)	.	.000
		N	729	728
	NJASK4SC	Correlation Coefficient	.537(**)	1.000
		Sig. (2-tailed)	.000	.
		N	728	728

** Correlation is significant at the 0.01 level (2-tailed).

Correlations

			EM3EYSC	NJASK4SC
Spearman's rho	EM3EYSC	Correlation Coefficient	1.000	.574(**)
		Sig. (2-tailed)	.	.000
		N	729	728
	NJASK4SC	Correlation Coefficient	.574(**)	1.000
		Sig. (2-tailed)	.000	.
		N	728	728

** Correlation is significant at the 0.01 level (2-tailed).

Correlations

			EM4MYSC	NJASK4SC
Spearman's rho	EM4MYSC	Correlation Coefficient	1.000	.594(**)
		Sig. (2-tailed)	.	.000
		N	729	728
	NJASK4SC	Correlation Coefficient	.594(**)	1.000
		Sig. (2-tailed)	.000	.
		N	728	728

** Correlation is significant at the 0.01 level (2-tailed).

EM3MYSC------*Everyday Mathematics* grade 3 mid-year scores
EM3EYSC-------*Everyday Mathematics* grade 3 end-of-year scores
EM4MYSC------*Everyday Mathematics* grade 4 mid-year scores
NJASKSC-------NJASK/4 scores

Frequency Distribution for Table 4: *Frequency of Students Performing at Passing Levels on Everyday Mathematics Assessments and NJASK/4 Math (page 69)*

Frequencies

Statistics

		TP1	TP2	TP3	TP4
N	Valid	623	623	622	623
	Missing	107	107	108	107

Frequency Table

TP1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	proficient	602	82.5	96.6	96.6
	not proficient	21	2.9	3.4	100.0
	Total	623	85.3	100.0	
Missing	0	105	14.4		
	System	2	.3		
	Total	107	14.7		
Total		730	100.0		

TP2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	proficient	581	79.6	93.3	93.3
	not proficient	42	5.8	6.7	100.0
	Total	623	85.3	100.0	
Missing	0	105	14.4		
	System	2	.3		
	Total	107	14.7		
Total		730	100.0		

TP3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	proficient	610	83.6	98.1	98.1
	not proficient	12	1.6	1.9	100.0
	Total	622	85.2	100.0	
Missing	0	105	14.4		
	System	3	.4		
	Total	108	14.8		
Total		730	100.0		

TP4

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	proficient	548	75.1	88.0	88.0
	not proficient	75	10.3	12.0	100.0
	Total	623	85.3	100.0	
Missing	0	103	14.1		
	System	4	.5		
	Total	107	14.7		
Total		730	100.0		

The TP categories combine students who performed at proficient and advanced proficient levels into one overall passing (proficient) level.

TP1--------grade 3 mid-year *Everyday Mathematics* assessments

TP2--------grade 3 end-of year *Everyday Mathematics* assessments

TP3--------grade 4 mid-year *Everyday Mathematics* assessments

TP4--------NJASK/4

Cross Tabulation for Table 5: *Relationship Between Students Who Performed at Proficient levels on Both Everyday Mathematics Assessments and NJASK/4* *(Page 70)*

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
TP1 * TP4	622	85.2%	108	14.8%	730	100.0%
TP2 * TP4	622	85.2%	108	14.8%	730	100.0%
TP3 * TP4	622	85.2%	108	14.8%	730	100.0%

TP1 * TP4

Crosstab

			TP4		Total
			proficient	not proficient	
TP1	proficient	Count	543[1]	58	601
		Expected Count	529.5	71.5	601.0
		Std. Residual	.6	-1.6	
	not proficient	Count	5	16	21
		Expected Count	18.5	2.5	21.0
		Std. Residual	-3.1	8.5	
Total		Count	548	74	622
		Expected Count	548.0	74.0	622.0

1. 543 = 87% ---students who were proficient (passed) both the grade 3 mid-year EMA and NJASK/4

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	85.711(b)	1	.000		
Continuity Correction(a)	79.480	1	.000		
Likelihood Ratio	49.406	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	85.573	1	.000		
N of Valid Cases	622				

a Computed only for a 2x2 table
b 1 cells (25.0%) have expected count less than 5. The minimum expected count is 2.50.

TP2 * TP4

Crosstab

			TP4		Total
			proficient	not proficient	
TP2	proficient	Count	530[2]	50	580
		Expected Count	511.0	69.0	580.0
		Std. Residual	.8	-2.3	
	not proficient	Count	18	24	42
		Expected Count	37.0	5.0	42.0
		Std. Residual	-3.1	8.5	
Total		Count	548	74	622
		Expected Count	548.0	74.0	622.0

2. 530 = 85%--students who were proficient (passed both the grade 3 end-of-year EMA and NJASK/4

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	87.970(b)	1	.000		
Continuity Correction(a)	83.402	1	.000		
Likelihood Ratio	55.873	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	87.829	1	.000		
N of Valid Cases	622				

a Computed only for a 2x2 table
b 1 cells (25.0%) have expected count less than 5. The minimum expected count is 5.00.

TP3 * TP4

Crosstab

			TP4		Total
			proficient	not proficient	
TP3	proficient	Count	546[3]	64	610
		Expected Count	537.4	72.6	610.0
		Std. Residual	.4	-1.0	
	not proficient	Count	2	10	12
		Expected Count	10.6	1.4	12.0
		Std. Residual	-2.6	7.2	
Total		Count	548	74	622
		Expected Count	548.0	74.0	622.0

3. 546=87.8% students who were proficient (passed) both the grade 4 mid-year EMA and NJASK/4

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	59.573(b)	1	.000		
Continuity Correction(a)	52.826	1	.000		
Likelihood Ratio	33.462	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	59.477	1	.000		
N of Valid Cases	622				

a Computed only for a 2x2 table
b 1 cells (25.0%) have expected count less than 5. The minimum expected count is 1.43.

The TP categories combine students who performed at proficient and advanced proficient levels into one overall passing (proficient) level.

TP1--------grade 3 mid-year *Everyday Mathematics* assessments

TP2--------grade 3 end-of year *Everyday Mathematics* assessments

TP3--------grade 4 mid-year *Everyday Mathematics* assessments

TP4--------NJASK/4

Regressions for Table 6: *Effectiveness of the Relationship of Everyday Mathematics Scores (EMA) with Performance on the NJASK/4 for a Sample of 623 (page 72)*

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
PROFFLEV	2.29	.678	624
PROFLEV	2.22	.495	624

Correlations

		PROFFLEV	PROFLEV
Pearson Correlation	PROFFLEV	1.000	.416
	PROFLEV	.416	1.000
Sig. (1-tailed)	PROFFLEV	.	.000
	PROFLEV	.000	.
N	PROFFLEV	624	624
	PROFLEV	624	624

Variables Entered/Removed(b)

Model	Variables Entered	Variables Removed	Method
1	PROFLEV(a)	.	Enter

a All requested variables entered.
b Dependent Variable: PROFFLEV

Model Summary
a Predictors: (Constant), PROFLEV

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.416(a)	.173	.171	.617

Change Statistics				
R Square Change	F Change	df1	df2	Sig. F Change
.173	129.875	1	622	.000

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	49.416	1	49.416	129.875	.000(a)
	Residual	236.661	622	.380		
	Total	286.077	623			

a Predictors: (Constant), PROFLEV
b Dependent Variable: PROFFLEV

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.025	.114		9.029	.000
	PROFLEV	.570	.050	.416	11.396	.000

a Dependent Variable: PROFFLEV

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
PROFFLEV	2.29	.678	623
PRFLEV	2.12	.506	623

Correlations

		PROFFLEV	PRFLEV
Pearson Correlation	PROFFLEV	1.000	.435
	PRFLEV	.435	1.000
Sig. (1-tailed)	PROFFLEV	.	.000
	PRFLEV	.000	.
N	PROFFLEV	623	623
	PRFLEV	623	623

Variables Entered/Removed(b)

Model	Variables Entered	Variables Removed	Method
1	PRFLEV(a)	.	Enter

a All requested variables entered.
b Dependent Variable: PROFFLEV

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.435(a)	.189	.188	.611

Change Statistics				
R Square Change	F Change	df1	df2	Sig. F Change
.189	144.875	1	621	.000

a Predictors: (Constant), PRFLEV

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	54.019	1	54.019	144.875	.000(a)
	Residual	231.551	621	.373		
	Total	285.570	622			

a Predictors: (Constant), PRFLEV
b Dependent Variable: PROFFLEV

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.053	.105		9.991	.000
	PRFLEV	.582	.048	.435	12.036	.000

a Dependent Variable: PROFFLEV

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
PROFFLEV	2.29	.678	624
PROLEV	2.35	.515	624

Correlations

		PROFFLEV	PROLEV
Pearson Correlation	PROFFLEV	1.000	.480
	PROLEV	.480	1.000
Sig. (1-tailed)	PROFFLEV	.	.000
	PROLEV	.000	.
N	PROFFLEV	624	624
	PROLEV	624	624

Variables Entered/Removed(b)

Model	Variables Entered	Variables Removed	Method
1	PROLEV(a)	.	Enter

a All requested variables entered.
b Dependent Variable: PROFFLEV

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.480(a)	.230	.229	.595

Change Statistics				
R Square Change	F Change	df1	df2	Sig. F Change
.230	185.974	1	622	.000

a Predictors: (Constant), PROLEV

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	65.847	1	65.847	185.974	.000(a)
	Residual	220.230	622	.354		
	Total	286.077	623			

a Predictors: (Constant), PROLEV
b Dependent Variable: PROFFLEV

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.808	.111		7.266	.000
	PROLEV	.631	.046	.480	13.637	.000

a Dependent Variable: PROFFLEV

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
PROFFLEV	2.29	.678	623
PROFLEV	2.22	.495	623
PRFLEV	2.12	.506	623
PROLEV	2.35	.516	623

Correlations

		PROFFLEV	PROFLEV	PRFLEV	PROLEV
Pearson Correlation	PROFFLEV	1.000	.417	.435	.482
	PROFLEV	.417	1.000	.474	.325
	PRFLEV	.435	.474	1.000	.383
	PROLEV	.482	.325	.383	1.000
Sig. (1-tailed)	PROFFLEV	.	.000	.000	.000
	PROFLEV	.000	.	.000	.000
	PRFLEV	.000	.000	.	.000
	PROLEV	.000	.000	.000	.
N	PROFFLEV	623	623	623	623
	PROFLEV	623	623	623	623
	PRFLEV	623	623	623	623
	PROLEV	623	623	623	623

Variables Entered/Removed(b)

Model	Variables Entered	Variables Removed	Method
1	PROLEV, PROFLEV, PRFLEV(a)	.	Enter

a All requested variables entered.
b Dependent Variable: PROFFLEV

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.582(a)	.338	.335	.552

Change Statistics				
R Square Change	F Change	df1	df2	Sig. F Change
.338	105.527	3	619	.000

a Predictors: (Constant), PROLEV, PROFLEV, PRFLEV

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	96.631	3	32.210	105.527	.000(a)
	Residual	188.939	619	.305		
	Total	285.570	622			

a Predictors: (Constant), PROLEV, PROFLEV, PRFLEV
b Dependent Variable: PROFFLEV

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.032	.129		.244	.807
	PROFLEV	.287	.052	.210	5.563	.000
	PRFLEV	.278	.052	.207	5.371	.000
	PROLEV	.438	.047	.334	9.279	.000

a Dependent Variable: PROFFLEV

PROFLEV----- Proficiency levels for grade 3 mid-year EMA
PRFLEV--------Proficiency levels for grade 3 end-of-year EMA
PROLEV------- Proficiency levels for grade 4 mid-year EMA
PROFFLEV----Proficiency levels for NJASK/4

Cross Tabulations for Table 7: *Relationship Between Special Education Classification and Performance on NJASK/4 in Mathematics (page 75)*

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
SPED * PROFFLEV	623	85.3%	107	14.7%	730	100.0%

SPED * PROFFLEV Crosstabulation

			PROFFLEV			Total
			partially proficient	proficient	advanced proficient	
SPED	special ed	Count	20	31	12	63
		Expected Count	8.0	28.8	26.2	63.0
		Std. Residual	4.2	.4	-2.8	
	regular ed	Count	59	254	247	560
		Expected Count	71.0	256.2	232.8	560.0
		Std. Residual	-1.4	-.1	.9	
Total		Count	79	285	259	623
		Expected Count	79.0	285.0	259.0	623.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	28.828(a)	2	.000
Likelihood Ratio	25.520	2	.000
Linear-by-Linear Association	26.368	1	.000
N of Valid Cases	623		

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 7.99.

SPED-------special education students

PROFFLEV—proficiency levels on NJASK/4

Cross Tabulations for Table 8: *Comparison of Student Subgroups Performing at Passing Levels (Proficient and Advanced Proficient) on the NJASK/4 (page 76)*

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
SPED * PROFFLEV	623	85.3%	107	14.7%	730	100.0%

SPED * PROFFLEV Crosstabulation

			PROFFLEV			Total
			partially proficient	proficient	advanced proficient	
SPED	special ed	Count	20	31[1]	12[1]	63
		Expected Count	8.0	28.8	26.2	63.0
		Std. Residual	4.2	.4	-2.8	
	regular ed	Count	59	254[2]	247[2]	560
		Expected Count	71.0	256.2	232.8	560.0
		Std. Residual	-1.4	-.1	.9	
Total		Count	79	285	259	623
		Expected Count	79.0	285.0	259.0	623.0

1. special education students passing NJASK/4 = 31 + 12 =43/63 = 68%
2. regular education students passing NJASK/4 = 501 = 560 = 89%

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
LEP * PROFFLEV	624	85.5%	106	14.5%	730	100.0%

LEP * PROFFLEV Crosstabulation

			PROFFLEV			Total
			partially proficient	proficient	advanced proficient	
LEP	limited English proficient	Count	5	5[1]	2[1]	12
		Expected Count	1.5	5.5	5.0	12.0
		Std. Residual	2.8	-.2	-1.3	
	English speaker	Count	74	281[2]	257[2]	612
		Expected Count	77.5	280.5	254.0	612.0
		Std. Residual	-.4	.0	.2	
Total		Count	79	286	259	624
		Expected Count	79.0	286.0	259.0	624.0

1. LEP students passing NJASK/4 = 5 + 2= 7/12 = 58%
2. Non-LEP students passing NJASK/4 = 281 + 257 = 538/612 = 88%

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
SES * PROFFLEV	624	85.5%	106	14.5%	730	100.0%

SES * PROFFLEV Crosstabulation

			PROFFLEV			Total
			partially proficient	proficient	advanced proficient	
SES	economically disadvantaged	Count	6	17[1]	3[1]	26
		Expected Count	3.3	11.9	10.8	26.0
		Std. Residual	1.5	1.5	-2.4	
	not disadvantaged	Count	73	269[2]	256[2]	598
		Expected Count	75.7	274.1	248.2	598.0
		Std. Residual	-.3	-.3	.5	
Total		Count	79	286	259	624
		Expected Count	79.0	286.0	259.0	624.0

1. Economically disadvantaged students passing NJASK/4 = 17 + 3 = 20/26 = 77%
2. Non-economically disadvantaged students passing NJASK/4 = 269 + 256 = 525/598 = 80%

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
ETHNIC2 * PROFFLEV	624	85.5%	106	14.5%	730	100.0%

ETHNIC2 * PROFFLEV Crosstabulation

			PROFFLEV			Total
			partially proficient	proficient	advanced proficient	
ETHNIC2	white	Count	52	219[1]	180[1]	451
		Expected Count	57.1	206.7	187.2	451.0
		Std. Residual	-.7	.9	-.5	
	asian	Count	11	45[2]	69[2]	125
		Expected Count	15.8	57.3	51.9	125.0
		Std. Residual	-1.2	-1.6	2.4	
	black, hispanic native American and other	Count	16	22[3]	10[3]	48
		Expected Count	6.1	22.0	19.9	48.0
		Std. Residual	4.0	.0	-2.2	
Total		Count	79	286	259	624
		Expected Count	79.0	286.0	259.0	624.0

1. White students passing NJASK/4 = 219 + 180 = 399/451 = 89%
2. Asian students passing NJASK/4 = 45 + 69= 114/125 = 91%
3. Students from other ethnic groups passing NJASK/4 = 22 + 10 = 32/48 = 67%

Cross Tabulation for Table 9: *Relationship Between Ethnicity and Performance on the NJASK/4 in Mathematics (page 78)*

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
ETHNIC2 * PROFFLEV	624	85.5%	106	14.5%	730	100.0%

ETHNIC2 * PROFFLEV Crosstabulation

			PROFFLEV			Total
			partially proficient	proficient	advanced proficient	
ETHNIC2	white	Count	52	219	180	451
		Expected Count	57.1	206.7	187.2	451.0
		Std. Residual	-.7	.9	-.5	
	asian	Count	11	45	69	125
		Expected Count	15.8	57.3	51.9	125.0
		Std. Residual	-1.2	-1.6	2.4	
	black, hispanic native American and other	Count	16	22	10	48
		Expected Count	6.1	22.0	19.9	48.0
		Std. Residual	4.0	.0	-2.2	
Total		Count	79	286	259	624
		Expected Count	79.0	286.0	259.0	624.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	32.364(a)	4	.000
Likelihood Ratio	28.269	4	.000
Linear-by-Linear Association	2.874	1	.090
N of Valid Cases	624		

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 6.08.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Contingency Coefficient	.222	.000
N of Valid Cases		624	

a Not assuming the null hypothesis.
b Using the asymptotic standard error assuming the null hypothesis

Oneway ANOVA and Multiple Comparisons for Discussion of Between Group Effects (p. 79-81)

Oneway

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
ETHNICIT	Between Groups	7.471	2	3.736	5.649	.004
	Within Groups	410.681	621	.661		
	Total	418.152	623			
SPED	Between Groups	2.620	2	1.310	15.041	.000
	Within Groups	54.009	620	.087		
	Total	56.629	622			
LEP	Between Groups	.189	2	.094	5.055	.007
	Within Groups	11.581	621	.019		
	Total	11.769	623			
SES	Between Groups	.418	2	.209	5.293	.005
	Within Groups	24.499	621	.039		
	Total	24.917	623			

Post Hoc Tests

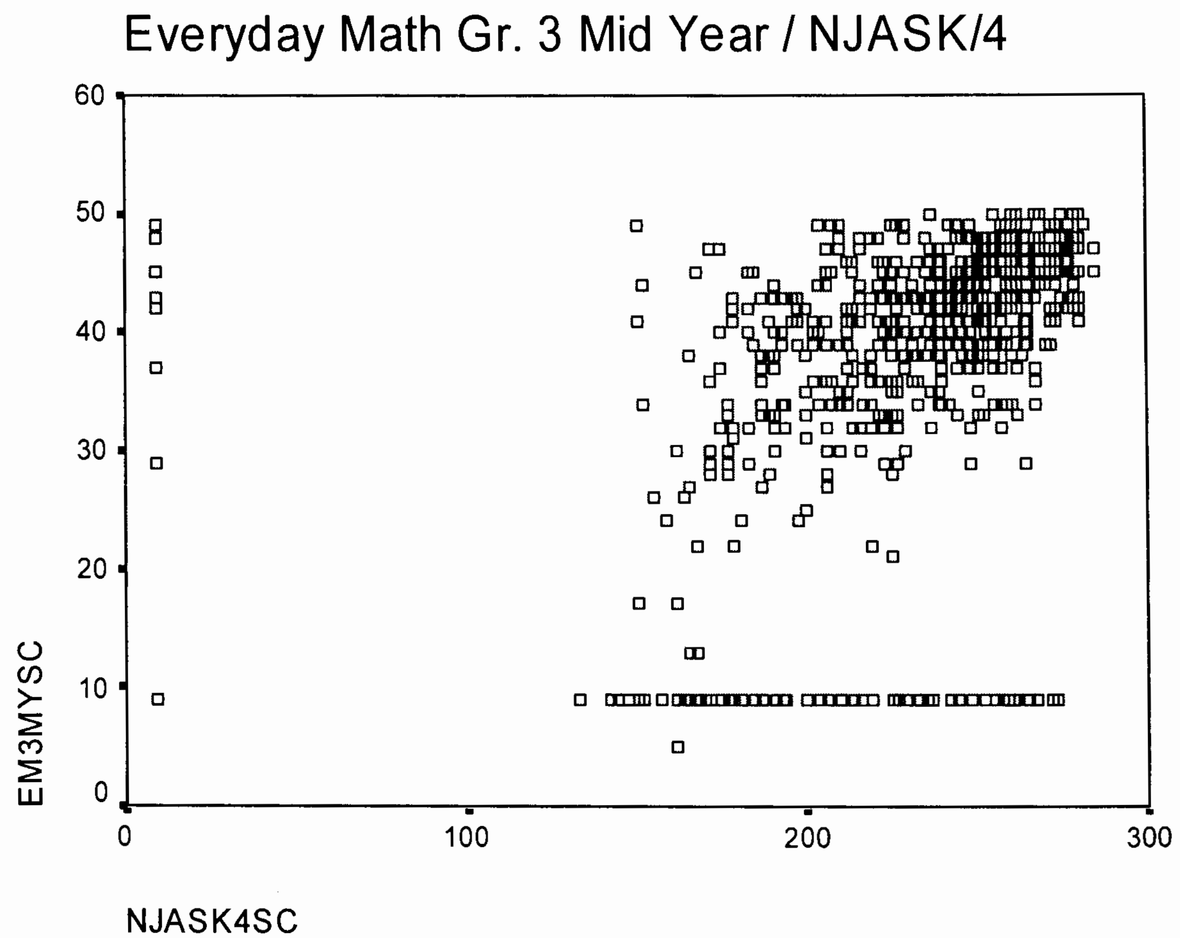
Multiple Comparisons

LSD

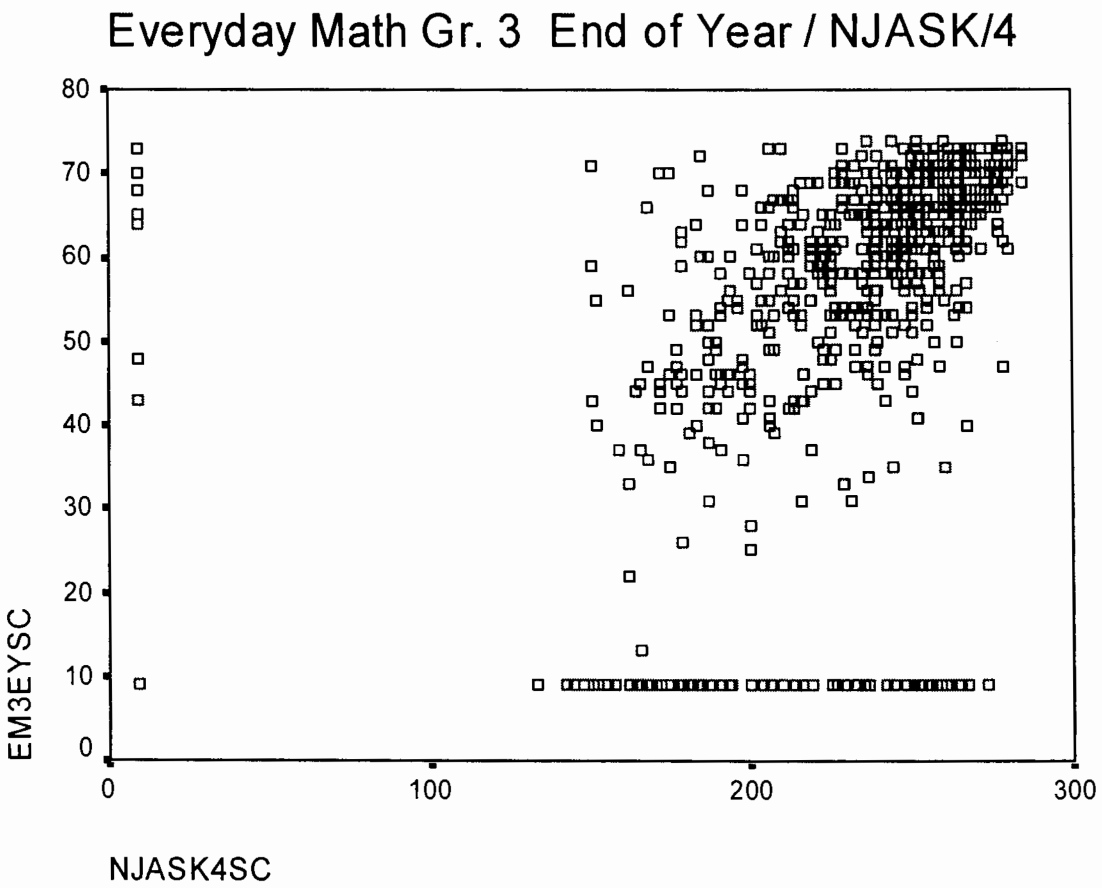
Dependent Variable	(I) PROFFLEV	(J) PROFFLEV	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
ETHNICIT	partially proficient	proficient	.33(*)	.103	.002	.12	.53
		advanced proficient	.33(*)	.105	.002	.13	.54
	proficient	partially proficient	-.33(*)	.103	.002	-.53	-.12
		advanced proficient	.00	.070	.969	-.13	.14
	advanced proficient	partially proficient	-.33(*)	.105	.002	-.54	-.13
		proficient	.00	.070	.969	-.14	.13
SPED	partially proficient	proficient	-.14(*)	.038	.000	-.22	-.07
		advanced proficient	-.21(*)	.038	.000	-.28	-.13
	proficient	partially proficient	.14(*)	.038	.000	.07	.22
		advanced proficient	-.06(*)	.025	.014	-.11	-.01
	advanced proficient	partially proficient	.21(*)	.038	.000	.13	.28
		proficient	.06(*)	.025	.014	.01	.11
LEP	partially proficient	proficient	-.05(*)	.017	.009	-.08	-.01
		advanced proficient	-.06(*)	.018	.002	-.09	-.02
	proficient	partially proficient	.05(*)	.017	.009	.01	.08
		advanced proficient	-.01	.012	.405	-.03	.01
	advanced proficient	partially proficient	.06(*)	.018	.002	.02	.09
		proficient	.01	.012	.405	-.01	.03
SES	partially proficient	proficient	-.02	.025	.513	-.07	.03
		advanced proficient	-.06(*)	.026	.012	-.11	-.01
	proficient	partially proficient	.02	.025	.513	-.03	.07
		advanced proficient	-.05(*)	.017	.005	-.08	-.01
	advanced proficient	partially proficient	.06(*)	.026	.012	.01	.11
		proficient	.05(*)	.017	.005	.01	.08

- The mean difference is significant at the .05 level.

Scattergram for Everyday Mathematics Grade 3 Mid Year Assessment Results as Compared to NJASK/4 Results (page71)



Scattergram for Everyday Mathematics Grade 3 End-of-Year Assessment Results as Compared to NJASK/4 Results (page 71)



Scattergram for Everyday Mathematics Grade 4 Mid Year Assessment Results as Compared to NJASK/4 Results (page 71)