2007

An analysis of 2005 NAEP 8th grade mathematics achievement items by content strand, problem type and language complexity

Yvette Marie Fagan

The University of San Francisco

AN ANALYSIS OF 2005 NAEP 8TH GRADE MATHEMATICS ACHIEVEMENT ITEMS BY CONTENT STRAND, PROBLEM TYPE AND LANGUAGE COMPLEXITY

A Dissertation Presented
to
The Faculty of the School of Education
Learning and Instruction Department

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

by

Yvette Marie Fagan
San Francisco
December, 2007

This dissertation, written under the direction of the candidate's dissertation committee and approved by the members of the committee, has been presented to and accepted by the Faculty of the School of Education in partial fulfillment of the requirements for the degree of Doctor of Education. The content and research methodologies presented in this work represent the work of the candidate alone.

Yvette Marie Fagan
Candidate

5-14-07
Date

Dissertation Committee:

Robert Burns
Robert Burns, Chair

5-14-07
Date

Susan Evans
Susan Evans

5.14.07
Date

Patricia Mitchell
Patricia Mitchell

5-14-07
Date

UNIVERSITY OF SAN FRANCISCO
School of Education

DOCTORAL PROGRAM
Certificate of Completion of Dissertation Requirements

Yvette Marie Fagan	Learning and Instruction
Name of Student	Program Area
USF Student ID Number: ______	Date: January 22, 2008
Address: 1706 McAllister (Street)	San Francisco, CA (City and State) 94115 (Zip Code)

This is to certify that the above-named student has successfully fulfilled all requirements for the completion of the dissertation for the degree of Doctor of Education.

Robert Burns
Committee Chairperson

January 22, 2008
Date

Dissertation Title: An analysis of 2005 NAEP 8th grade mathematics achievement items by content strand problem type, and language complexity

ABSTRACT

An Analysis of 2005 NAEP 8th Grade Mathematics Achievement Items By Content Strand, Problem Type And Language Complexity

The purpose of this study was to conduct a descriptive analysis the 2005 NAEP 8th-grade mathematics assessment. In order to determine if a relationship between mathematical language fluency and mathematics achievement exists, the Mathematics Assessment Language Framework was created to classify the 2005 8th-grade NAEP mathematics assessment test items according to problem type and language complexity. The magnitude of the achievement gap on each content strand was then related to the percentage of items classified by problem type and language complexity.

Three procedures were used to analyze the research data. First, a comparative data analysis disaggregated racial/ethnic group data and compared mean scores by five mathematic content strands to examine differences in achievement on the 2005 NAEP Math. A series of *t*-tests were performed to compare White student group mean performance to group mean performance of Black students, Hispanic students, and Asian students. Second, a content analysis of the items was completed first by problem type and by MALF categories.. Third, the magnitudes of the achievement gaps within each strand were related to the percentage of items classified according to problem types and language complexity and a rank order correlation was computed.

Results revealed that measurement was the most difficult of the five strands based on overall mean achievement scores. Data analysis was the least difficult of the five content strands based on mean achievement but showed the second highest gap for Black students and Hispanic students when compared to their White peers. Achievement gap differences in the content strand of numbers and operations could be attributed more to computational competency than language complexity based on the analysis of problem types x language categories. Achievement gap differences in the content strand of data analysis could be attributed more to language complexity than computational competency based on the analysis of problem types x language categories. The Spearman rank order correlation suggested that relationships exist between achievement gap rank and problem type; achievement gap rank and language complexity category; and problem type and language complexity category.

ACKNOWLEDGMENTS

Dr. Robert Burns

Who have insured that
this dissertation is a "Perfect Manhattan"

Lanna Andrews
Who became my Mother Protectorate

Susan, Yvonne, Paula, and Nikki
For cheering me on with loving support and infinite patience

Candace Chinn
The "Oracle of Wisdom" for teaching me how to be a Willow

Val, Leslie, and Tyrone
My partners and Doc-mates without whom this
journeys-end would have been all but impossible

The Teachers and Staff of Dr. Wm. Cobb

Whose quiet support and pride in my accomplishments have
been unstinting over the last too-many years

Mother and Daddy
As usual I did it the hard-headed way and made it work for me, I want you
To know this time all your daughters, grandchildren and great-grand children Sang

Laura and Jeff
For your consistency and understanding

Lavay and Chris and all of the Red Hot Skillet Lickers
I found my voice in your company

Edee
My mother in spirit for being a source of renewal
…the joys of your garden

Helen
My best friend through distance, time, and the many experiences

Leslie, Al, and Matt
The original crew

All the friends who remain unnamed especially in the
City of San Francisco
Allison, Lawrence and Jeff of NOPA
Mike and staff at the Bean Bag Café
Laurent and Erin at Chez Spencer

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER ONE

STATEMENT OF THE PROBLEM

In 1966, the landmark Coleman Report (Coleman, et al, 1966) used student test score performance in reading and mathematics for the first time as indicators of equality in the American educational system for children of different races, gender, and socioeconomic status. Using data from over 600,000 students from across the nation, the Coleman Report found that White students out-performed minority students, wealthy students outperformed poorer students, and students with highly-educated parents outperformed students with less-educated parents.

More than 40 years of research based on the factors identified by the Coleman Report have not resolved the issue of the achievement gap between White and minority students (Abedi & Lord, 2005; Atweh, Bleicher, & Cooper, 1998; Boaler, 2002; Fenema & Leder, 1990; Ladson-Billings, 1997; Lee, 2004; Lubienski, 2000a, 2000b, 2004, 2006; National Center for Educational Statistics (NCES), 2005-2007; National Science Foundation (NSF), 2004, 2007; No Child Left Behind (NCLB), 2002; Okpala, Okpala, & Smith (2001); Tate, 1997; U.S. Department of Commerce, 2001; U.S. Department of Education, 2006; Wenglinsky, 2004). In *Reaching the Top: A Report of the National Task Force on Minority Achievement* (The College Board, 1999), the differential in academic achievement between minority students and their White peers was thought to be predicated on factors similar to those found by Coleman: (1) economic circumstances, (2) parent educational background, (3) racial and ethnic prejudice and discrimination, (4) cultural attributes of the home, community, and school, and (5) the quality, amount and uses of school resources. While achievement gaps exist in reading, mathematics and science achievement, this study is restricted to mathematics achievement.

The social and economic impact of the mathematics achievement gaps makes them a severe national problem (Kober, 2001). Mathematics performance in schools acts as an important determinate of successful educational attainment, career choice, and economic rewards (NSF, 2004). Nations require a workforce that is mathematically literate to adapt to increasing technological innovations and global markets for goods and services (Erpenbach & Forte-Fast, 2004; National Council for the Center for Excellence [NCCES] 1984; NCES, 2005). Yet despite the need for mathematical literacy, education has failed to ensure that all students are able to function with high mathematical attainment on assessment measures. For example, minority students' mathematic achievement scores today are lower than non-minority students’ scores were 10 years ago, before a decade of focused national school reform efforts to raise minority mathematical attainment (Educational Research Service, 2001; National Council of Teachers of Mathematics [NCTM], 2000; NCES, 2005). Investigations are needed to understand why certain groups consistently lack the necessary skill sets to reach proficiency on national and state mathematic assessments. Therefore, this study focused on student mathematics performance using data from the 2005 National Assessment of Educational Progress (NAEP).

Since 1973, the federal government has required that a sample of the nation’s students be assessed in reading and mathematics to ascertain the achievement of all students in reaching proficient levels of educational attainment. Biannually, since 1996, *The Nation’s Report Card* has presented the achievement results in mathematics, disaggregated by gender, race/ethnicity, and socioeconomic status (Elementary and Secondary Education Act of 1965 [EASA]; National Assessment Governing Board

[NAGB], 2004 NCES, 2005). Despite some improvement in overall performance, gaps between various racial and ethnic groups continue to persist and have not narrowed significantly since 1990 (NAEP, 2005; NSF, 2004). For example, in a cross-grade comparative analysis on overall mathematics achievement over the decade from 1990 through 2000, the difference in the achievement gap showed that Black and Hispanic 8th-graders' mathematic scores remained close to the achievement of White 4th graders' scores. Another grade level comparison showed that in 1990, Black 12th graders scored similarly in mathematics to White 8th-graders, but by 2000, White 8th-graders outperformed Black 12th graders by an average of eight points.

Disparities in achievement gaps due to socioeconomic status (SES) were similar to race-related gaps. Whites outperformed Blacks and Hispanics on mathematics in each of the SES quartiles. For example, in the 12th grade, low SES White students scored within three points of Black students in the highest SES. Between low SES White students and low SES Black students there was a 22-point differential (Lubienski & Shelley, 2003). Recent data from the 2005 NAEP mathematics assessment indicated that the proportion of Black and Latino 8th-graders reaching the *proficient* level was 9% and 13%, respectively, compared to 39% of White students and 47% of Asian students. Achievement at the *proficient* level means the student has demonstrated competency with challenging grade-level content (NAGB, 2003).

Another set of studies using 1996-2001 NAEP mathematics assessment data examined the effects of different instructional practices employed with students during mathematical problem solving on achievement. The findings suggested that students who had the most experience with a more discursive, student-centered approach recommended by National Council of Teachers of Mathematics (NCTM) were from higher SES groups, disproportionally White, with achievement scores on the proficient level or above on NAEP mathematics assessments (Lubienski, 2000a, 2000b, 2004).

Achievement gaps identified by these and other studies are based on NAEP test items that are reviewed extensively for fairness and bias and examined statistically for differential item functioning (DIF), or evidence that an item might put a subgroup at a disadvantage. DIF analyses compare the item performance of sub-groups (e.g., Black or Hispanic examinees), and when an item is substantially harder for one group than for another group, the item is reviewed to uncover, if possible, causes for the differential performance. If it is determined that group differences can be explained by characteristics unrelated to the purpose of the test, the item can be deleted from the scoring. However, if it is determined that the differential performance is caused by relevant knowledge being assessed for mastery (e.g., algebraic equations on a mathematics exam), then it is likely that the item would be retained and scored (Zumbo, 1999).

Thus, the achievement gaps between minority and White students continue to persist even with items screened for potential bias, and these achievement gaps have important implications for America's ability to compete globally. Understanding the achievement gaps is a top priority among educators and researchers alike.

One area not well studied is the effect of mathematical language fluency on mathematical achievement (RAND Mathematics Study Panel, 2003). Language fluency in mathematics refers to the ability of a student to understand what is required in a mathematics test item and delineates the differences between language used on a daily basis and the language associated with problem solving. One aspect of mathematical fluency is the use of complex language that is required to represent abstract structures and relationships using words, mathematical notation, symbols, and logic, which in mathematics, is more careful and accurate than everyday speech (Nowak, Komarova, & Niyogi, 2002; Wakefield, 2000). Though the acquisition of mathematical literacy draws on many of the same skills as print literacy (Adams, 2003; Wakefield, 2000), Wakefield (2000) theorized that mathematics qualifies as a separate language based on the socio-linguistic structures of words, symbols, and expressions used to communicate ideas. He suggested that the understanding of mathematical concepts depends upon a student's fluency, proficiency, and comprehension of mathematics vocabulary.

This theory suggests that language complexity is one factor that may contribute to the achievement gap between White and minority students. To date, the NAEP achievement gaps have been based on subtests defined by the five strands of mathematics curriculum--number sense, measurement, geometry, data analysis, and algebra (NCTM, 1989). NAEP items consist of multiple-choice and open-ended questions of varied complexity that meet the standard parameters of current mathematics assessments (Levine & Reed, 2000). The content strands across item types may use language that confounds the measurement of students' mathematics abilities with their language fluency - the ability to understand complex language in problem solving. If language complexity plays a role in the achievement gap in mathematics, then analyzing mathematical items according to language-based categories may reveal a source of the achievement gap. Few studies have examined the language difficulty of NAEP math items to determine to what extent, if any, language complexity plays a role in the achievement gap. This underlies the need for the present study.

Purpose of the Study

The purpose of this study was to conduct a descriptive analysis the 2005 NAEP 8th-grade mathematics assessment. In order to determine if a relationship between mathematical language fluency and mathematics achievement exists, the Mathematics Assessment Language Framework was created to classify the 2005 8^{th-}grade NAEP mathematics assessment test items according to problem type and language complexity. The magnitude of the achievement gap on each content strand was then related to the percentage of items classified by problem type and language complexity. It was thought that analyzing NAEP mathematics items according to problem type and language-based categories may reveal sources of the achievement gaps.

The analysis was conducted in three steps. First, achievement gaps on each of the five content strands (number and operation, measurement, geometry, data analysis, and algebra) were computed by race. Second, the percentage of item problem types and categories of language complexity from the Mathematics Assessment-Language Framework were calculated within each strand. Third, the magnitudes of the achievement gaps within each strand were related to the percentage of items classified according problem types and language complexity.

Significance of the Study

This study was important for three reasons. First, even though the influence of language and language factors in mathematics has become more important in our pluralistic society, research on such factors has not kept up with their importance. Most research on the 2005 NAEP Mathematics Assessment, the largest national representative sample of students' mathematics performance ever constituted by the NCES, has focused on student performance, attitude, instructional practice, and race-related equity issues. Very little NAEP research has considered the impact of language factors on NAEP mathematics performance.

Second, this study used data from the restricted-use data set, enabling analysis of language complexity, problem type, and achievement gaps to be disaggregated by strand. Most NAEP studies do not have access to content strand data and must rely on the total mean score of the NAEP assessment for analysis purposes. Disaggregating by content strand should help to identify language related factors in mathematics achievement.

Third, examining the relationship between language complexity and mathematics achievement may reveal bias not uniformly identified through current differential item function (DIF) techniques. Drawing attention to the underlying language factors may help improve mathematical assessments' ability to capture the knowledge and skills that students need to know and better portray students' academic ability (Lubienski, 2000a, 200b, 2004, 2006; NAGB, 2003, 2004; NCES, 2005; NCTM, 2004, RAND, 2004).

Theoretical Framework

The theoretical foundation for this study is the language-based framework of Wakefield (2000). This framework defines the foundational social-linguistic constructs of mathematical language that identifies the interdependence of words, symbol, and expressions used to construct meaning and communicate ideas. Wakefield suggested that, similar to the processes children use to acquire fluency in their native language, mathematic language fluency has innate qualities that are predetermined prior to formal instruction. For example, prior to entering school, most preschoolers have mastered mathematical concepts such as the concepts of equality, greater than, lesser than, addition, and subtraction. Wakefield uses *Chomsky's(1975) Theory of Language Acquisition* and *Piaget's (1969) Theory of Cognitive Domain* to support his contention that acquiring mathematic language fluency shares many attributes of early language acquisition such as when children learn to talk. Wakefield equates the early conceptual mastery of mathematical concepts as math-acquisition devices (MAD) that are similar to the language-acquisition devices (LAD) suggested by Chomsky (1975). Chomsky theorized that a complex but finite set of rules governs all languages. Further, humans are born with an innate capacity to learn whatever language they hear and this learning of language is an integral part of neurological functioning.

Wakefield draws further parallels to mathematics as a separate language when citing Piaget's (1969) preoperational stage of cognitive development (ages 2-7) as the stage when rapid development of language ability and early acquisition of mathematical concepts of preschool children coincide. Wakefield posits that if speaking to and engaging young children in activities that build vocabulary and help them to make sense of the world increases cognitive development and language acquisition, then mathematics ability may be increased as well by engaging in conversations and using play involving mathematic operations such as counting, sorting, sharing, and valuing.

Wakefield's (2000) mathematics framework details ten attributes and/or characteristics of mathematical language based on unique socio-linguistic structures required to reason, communicate, and express ideas. These attributes and/or characteristics are: (1) abstractions (verbal or written symbols representing ideas or images) are used to communicate; (2) symbols and rules are uniform and consistent; (3) expressions are linear and serial; (4) understanding is based on opportunities to practice;

(5) memorization of symbols and rules are required to engage in discourse and practice; (6) a continuum from novice to expert requires translations and interpretations; (7) meaning is influenced by symbol order; (8) communication requires encoding and decoding; (9) increasing intuition, insightfulness and spontaneity accompany fluency; and (10) the possibility for expression is infinite.

Wakefield's attributes characterize the components of mathematics that distinguish it as a separate language. However, this framework does not provide objective measures which can be used to quantify the words, terminology, and vocabulary that are key factors in the communication process nor does the framework identify how these attributes can be used to evaluate mathematical performance by students. In order to develop objective measures for classifying items by language categories on the 2005 NAEP mathematics assessment, Wakefield's epistemological framework had to be adapted for use in the present study.

Successful problem solving depends on students' conceptual understanding, their relative strength in each series of steps to interpret meaning in terms of mathematical symbols (e.g., graphs, equations, relevant details), and whether the students can see and make use of the relationships among those steps. Because students' first judgments about how to approach mathematics items are based on whether graphic material is included in the item (Lai, Griffin, Mak, Wu, & Dulhunty, 2001), the present study first classified items into graphic versus non-graphic content. Graphic representations are used as a visual aid to interpret mathematical data such as symbols, pictures, graphs, grids, charts, maps, geometric shapes, and numerical graphics that include number lines, computation items with less than three word directions, frequency tables, and extended numerical patterns. Non-graphic representation refers to those items that have no visual or pictorial representation.

The graphic items were then were subdivided into three categories based on language complexity: (1) graphic vocabulary only (2) organize and plan, and (3). draw/manipulate to solve. *Graphic Vocabulary* refers to items that require understanding of specific mathematical terms to identify or confirm mathematical notation, geometric shapes, location on a map or grid, or to find discrete information on a graph or chart. *Organize and plan* refers to graphic representational items that also require knowledge and skills of syntax (word order), words (e.g., prepositional, proportional, multiple meaning), and directional signs to set up numerical expressions for computation. *Draw/manipulate to solve* refers to those items that require the use of additional resources not found within the stem of the question. These items may require the respondent to confirm an answer, extend, transform, locate, or plot patterns, or apply new information to solve an equation.

The non-graphic items were also subdivided into three categories based on language complexity: (1) non-graphic vocabulary,(2) convert-to-solve, (3) convert only. These items require an evaluation of relevant written information to solve. *Non-Vocabulary only* refers to items that require understanding of specific mathematical terms to identify or confirm mathematical notation, operations and formulas. *Convert and solve* refers to those items that require the understanding of the interrelationship of symbols and words to mathematical notation, and requires words or symbols to be changed into numerical notation prior to computation. *Convert only* refers to those items that require the knowledge of technical vocabulary and/or mathematical notation to locate or identify the same information using symbols, and words. This classification includes the automaticity of basic facts in addition, subtraction, and multiplication of whole numbers.

Background and Need

The No Child Left Behind Act of 2001 (NCLB) is the most recent in a series of legislative acts (e.g., Elementary and Secondary Education Act [ESEA], 1965; Goals 2000 - Educate America Act, Title III, Sec.302; Improving America's Schools Act of 1994) intended to improve the educational system and increase student performance. Policies designed in NCLB focused on the public accountability of student learning and the achievement gap among groups of students. The critical difference between NCLB and other federal educational acts is that states must show gains in total achievement and demonstrate progress in closing the achievement gap among subgroups defined by ethnicity, SES, English Language Learners and students with disabilities (Erpenbach, Forte-Fast, & Potts, 2003; NCLB, 2001; Resnick, 2005). This accountability measure is based on an "Annual Yearly Progress" index of the percentage of all students and the percentage of disaggregated subgroups scoring at proficient levels on statewide language arts and mathematics assessments. The NAEP assessments are used as comparative benchmarks to assess the rigor of the statewide assessments.

The effectiveness of NCLB to foster higher achievement for minority students has thus far proved inconclusive (Forte-Fast & Erpenbach, 2004). Despite changes in federal educational policy and state curriculum standards, the disparity in achievement remain especially in mathematics and science among White, Black, Latino, and Asian students. The achievement gap continues to be an intractable measure of the inequality of educational opportunities for all students to learn (Kober, 2001; NCES, 2005; PISA, 2005; TIMSS-R, 2003). As minority students are becoming an increasingly larger part of the school age population, there is an economic and social imperative to enhance the performance of minority groups so they may be integrated into an increasingly global workforce (NAEP, 2005; RAND Mathematics Study Panel, 2003; Resnick, 2005).

Assessing the Achievement Gap

Academic success in mathematics requires precision, consistency, attention to detail, conceptual agility, problem-solving flexibility, quick processing and recall, and cumulative learning that is integrated with verbal skills and reading proficiency to create strategic approaches to new tasks or learning (Levine & Reed, 1999; Resnick, 1987; 2005). Mathematics fluency depends on cumulative skills in computation, the ability to integrate words, symbols, and vocabulary to create meaning and communicate ideas.

According to Smith (2004), students acquiring mathematical fluency go through a series of stages that include: (a) initial and advanced *acquisition* where students learn to perform a skill with accuracy; (b) *proficiency* where students develop automaticity of a skill while maintaining accuracy; (c) *maintenance* where students have mastered the discrete skill at proficient levels over time and develop conceptual understanding; (d) *generalization* where students apply a skill(s) to different situations; and (e) *adaptation* where students apply their understanding to problem solving, reasoning, and real-life situations. While assessments vary according to grade level and state frameworks, many current mathematics achievement assessments focus on the skills found at the adaptation stage of student learning. According to Levine and Reed (2001), every mathematics assessment consists of the eight basic parameters for mathematic skills identified in Table

1, adapted for grade level complexity and content strand. The content of NAEP mathematics assessment questions is aligned to these parameters as identified in the test specifications for the NAEP framework (Federal Register, 2003; NAGB, 2004; Vinovskis, 1998).

The NAEP mathematics framework specifies mathematical content in five strands (number and operations, measurement, geometry, data analysis, and probability, and algebra) and as shown in Table 2, defines the proportion of questions assessed within each content strand for grades 4 and 8. The framework also specifies that students' ability to understand mathematics be demonstrated by using three problem types: multiple choice (50%), short-constructed response (25%) and extended-constructed response (25%). In addition, there are predetermined levels of the difficulty of test items across strand (NAGB, 2004; NCES, 2005).

Table 1

Basic Parameters of Mathematics Assessments (Levine & Reed, 2001).

Skill	Components
Basic number knowledge	Rote counting
	Object Counting
	Symbol recognition
	Symbol formation
	Symbol quantity association
Math Facts	Use of counting strategies
	Accurate recall
	Automatic fluency of basic facts
Math Notation	Knowledge of operational signs
	Understanding of quantitative and directional signs (e.g., <, <. =)
Written computation	Procedures and algorithm knowledge
	Sequencing of steps
	Fact recall
	Alignment
	Attention to signs
	Understanding of place value
	Self-checking
Mathematics vocabulary	Examples include: intersection, rounding, perimeter, mean
Concept Formation	Equation
	Proportional reasoning
	Prepositional reasoning
	Place value
	Spatial appreciation
	Imaging
Problem solving	Identification of relevant information
	Understanding of syntax and vocabulary
	Selection of operations
	Setting up problems
	Sequencing of steps
	Computational accuracy
	Self-monitoring
	Flexibility
Mental Calculation	With number problems
	With word problem

Table 2

Distribution of Test Questions Across the Five Content Strands for Grades 4 and 8

Content Strands	4^{th} Grade	8th-grade
Number and Operations	40%	20%
Measurement	20%	15%
Geometry	15%	20%
Data Analysis	10%	15%
Algebra	15%	30%

NAEP mathematics scores are typically reported in three ways. First, scaled scores are reported on a 500-point scale from 0 to 500. Second, the percentage of students scoring at five predetermined national percentiles are reported and labeled as *far below basic, below basic, basic, proficient,* and *advanced* achievement. These five proficiency levels and their cut-off scores are shown in Table 3. Finally, for some reporting purposes, three levels of student competency are reported and aligned to NCLB: *basic* (minimal grade level proficiency), *proficiency* (mastery of grade level standards), and *advanced* (exceeds grade level standards) (NAGB, 2004; NCES, 2005).

Because the current NAEP framework was first used in 1990, there are comparative data available for the past 15 years. In the overall national comparison between students in 1990 and 2005, average scores for all groups have increased for both 4^{th}- grade and 8^{th} -grade students: fourth grade students improved 25 points and the percentage of 4^{th}-grade students at the proficient level increased from 13% in 1990 to

36% in 2005; 8th-grade students improved 16 points overall and the percentage of 8th -grade students at the proficient level increased from 15% to 30%.

Table 3

Minimum Score Required for Performance at Each NAEP Achievement Level on a 500 point scale

Achievement Levels	Grade 4	Grade 8
Far Below Basic	Less than 170	Less than 240
Below Basic	170	240
Basic	214	262
Proficient	249	299
Advanced	282	333

Despite overall scores increasing over the past 15 years, the percentage of Black and Hispanic students at the proficient level is low relative to White and Asian students. Figures 2 and 3 present the 1990 and 2005 percentages of 4th -grade and 8th-grade students at three NAEP proficiency levels for Whites, Blacks, Hispanics, and Asians. The percentage of White and Asian students at the proficient level is considerably higher than the percentage of Black and Hispanic students (NCES, 2006).

Moreover, the actual achievement gaps have remained roughly the same from 1990 to 2005. In 1990, there was a 32 point gap in 4th-grade achievement between White and Black students; in 2005, the gap was 26 points. Between White and Hispanic students in 1990, the score gap was 20 points; in 2005, the score gap remained exactly the same. At the 8th-grade, the score gap between White and Black students in 1990 and

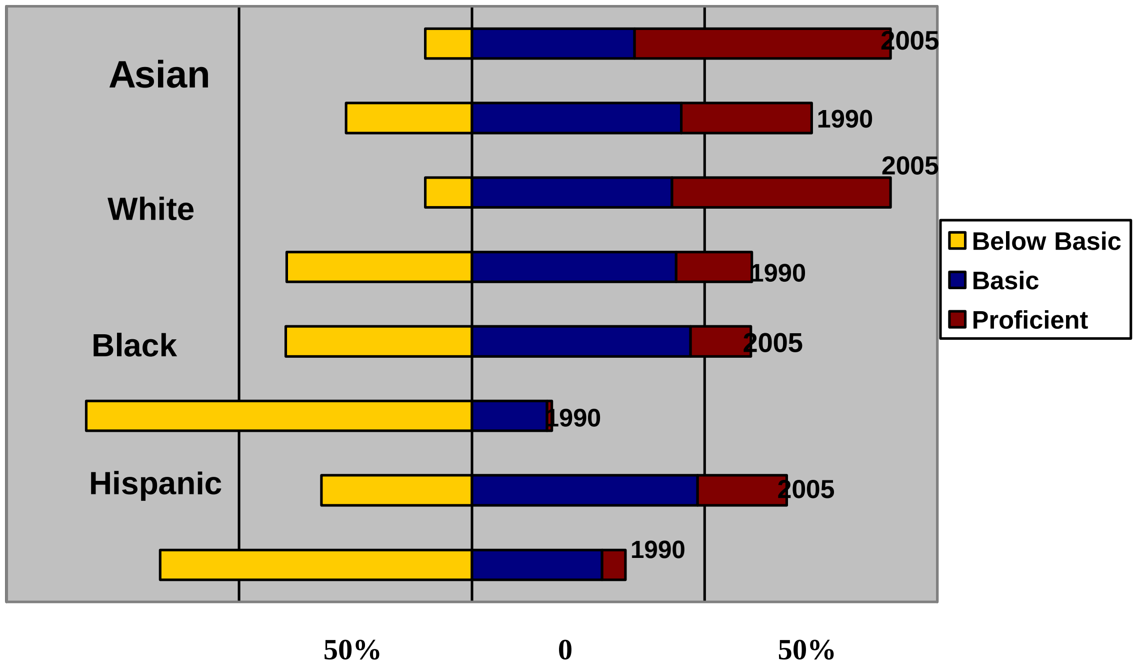


Figure 1. Average scale score comparison by achievement level results by race/ethnicity in grade 4 between 1990 and 2005 on the NAEP Mathematics (NCES, 2006).

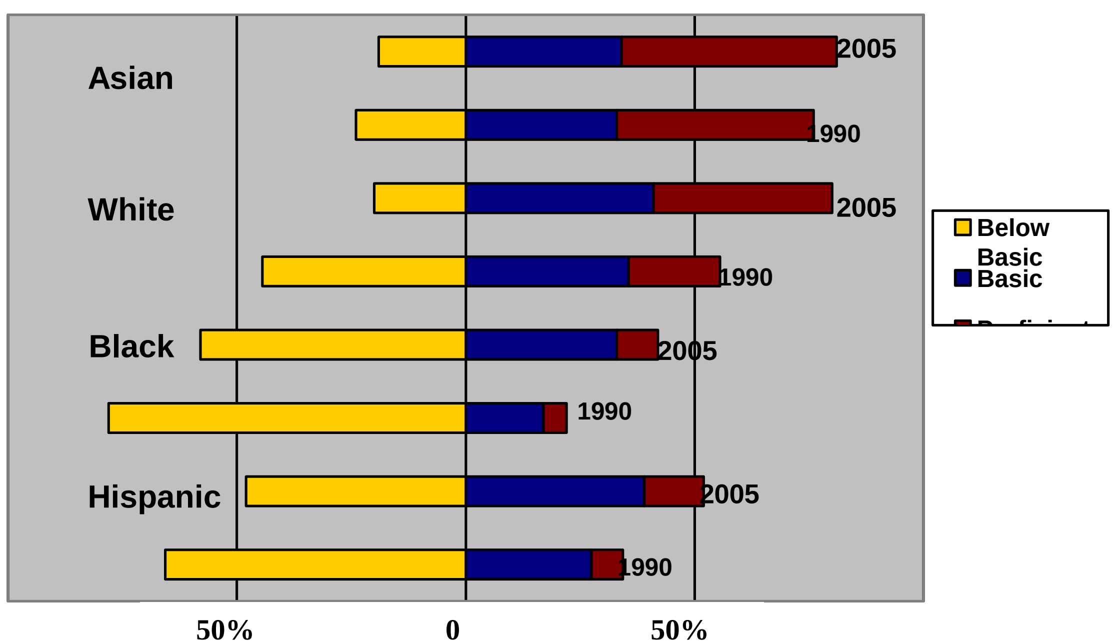


Figure 2. Average scale score comparison by achievement level results by race/ethnicity in grade 8 between 1990 and 2005on the NAEP Mathematics (NCES, 2006).

2005 remained basically unchanged, 33 points in 1990 and 34 points in 2005. From 1990 to 2005, the White-Hispanic score gap increased from 24 points to 27 points.

Mathematics Language and the Achievement Gap

Fundamental to achievement in mathematics is the premise that students have the skill set to perform tasks using and applying the four operations (addition, subtraction, multiplication, and division) across the content strands. Secondary to achievement in mathematics is the understanding of how written language interacts with the mathematical skills to perform problem-solving tasks across the content strands. Previous research has identified English language structures, vocabulary, and inferential language as areas that may inhibit mathematical achievement (Abedi & Lord, 2001).

The connection between language fluency and mathematical achievement has also been studied to identify isolated language structures which may affect academic achievement in the classroom (Curry, 1996; Fuchs & Fuchs, 2002). Fuchs and Fuchs (2002), for example, examined the functional performance of students with mathematics disabilities (MD), with and without reading disabilities (RD), on a range of mathematical problem-solving tasks involving arithmetic story problems, complex story problems, and real-world story problems. The results suggested that the performance of students decreased across the three problem-solving tasks as language complexity of the items increased for both groups of students. In a similar vein, Leong and Jerrod (2001) conducted an experimental study to examine word problems involving two different linguistic structures found in mathematics word problems. The results indicated that an interaction between problem type and ability level existed in the students' capacity to find information to use in the text of word problems. Additionally, the relationship between the ability to read and comprehend text with the ability to problem solve may be confounded by the students' understanding of complex language including non-literal directions and prepositional phrasing.

Other studies have been conducted to ascertain the role of vocabulary in mathematical problem solving. Lachance and Confrey (1995) conducted a quasi-experimental study using opened-ended problems to develop an understanding of decimals and connect new concepts to earlier multiplicative constructs of ratio and fractions through strategizing solutions individually and in various discussion groups. A paired t-test showed that students made significant gains (p <.01) between the pre- and posttests, from an average of 15.5 scale points on the pretest to 80.8 scale points on the posttest. The researchers contend that the development of ratio and proportional reasoning established early in the curriculum, along with opportunities to develop the mathematical language to explain student thinking, provided strategies to successfully connect and apply knowledge in a variety of assessment measures.

A longitudinal study by Huntsinger, Jose, Larson, Krieg, and Shaligram (2000) examined cross-cultural beliefs about mathematics performance between parents of White and Chinese primary students. Differences in belief systems were found regarding mathematic achievement: Chinese parents believed hard work resulted in high math achievement; White parents believed that innate ability was the primary reason for high mathematics achievement. Another difference was found in homework practice between the Chinese parents and the White parents. The Chinese parents spent more time on homework and had a significant influence on their children's vocabulary instruction. The Chinese parents all reported spending time at home to ensure their children understood the terminology; no White parent reported spending any time on homework support for vocabulary. The results suggested that the emphasis on mathematics vocabulary by their parents may be a primary reason the Chinese students were outperforming their White peers in mathematics by third grade.

Another group of studies investigated the influence of graphic representation on students' understanding of the mathematical task needed to solve word problems. Blinko (2004), for example, examined the effect of three different ways of presenting mathematics problems: manipulatives, graphic representation, and words only. The evidence suggested that the context of the layout and design of a question may influence whether or not a student will consider a problem approachable. If context matters in the outcome of attempted questions, and those questions without visual presentation are perceived as harder, the visual representation of word problems may impact a student's rate of omissions on a mathematics assessment.

Language frameworks created to investigate the effect of various problem-solving strategies in mathematics have focused on English-language proficiency and reading comprehension. Few studies exist that explore the relationship between the characteristics of verbal ability to performance on the mathematical tasks contained within standardized assessments. One such study by Abedi and Lord (2001) used items adapted from the 2000 NAEP assessment to investigate the importance of language proficiency. Modification of math items were based on six linguistic features: (a) familiarity of non-math vocabulary -infrequent words were changed; (b) voice of verb phrase – active from passive tense changed to active; (c) length of nominal – shortened; (d) conditional clauses – replaced with separate sentence and direct literal language; (e) relative clauses –

removed; question phrases – rephrased to simple directional questions; and (f) abstract and/or impersonal presentations – made concrete and personalized. The results suggested that differences in linguistic structure of math word problems affected performance of English-proficient students more than students who were not English proficient. The results also suggested that changes in the complexity of the language most benefited the lower achieving segments of the middle school population.

What role does mathematical language fluency play on the achievement gap differential on the 2005 NAEP mathematics assessment? Currently, the NAEP mathematics achievement gap is based on items that vary by content strand (NAGB, 2004). Within content strand the NAEP items may also vary by problem type and language complexity. If language plays a role in the achievement gap, an analysis of the language complexity of NAEP items may help reveal sources of the achievement gap that previously have been unidentified.

Research Questions

Consequently, three research questions were examined in this study:

Research Question 1: What are the achievement gap differences between racial and ethnic groups (White, Black, Hispanic, Asian) on the 8th-grade 2005 NAEP Mathematics Assessment by content strand (Number and Operation, Measurement, Geometry, Data Analysis, and Algebra)?

Research Question 2: How are the five strands characterized in terms of problem type (multiple response, constructed response, and extended response), and language complexity (graphic vocabulary, non-graphic vocabulary, operate and plan, convert-to-solve, draw and manipulate, and convert only)?

Research Question 3: What is the magnitude of the relationships between the achievement gaps and the percentage of items of different problem types and different language complexity categories?

Definition of Terms

Academic achievement -- generally, defined by grades in pre-college courses, class rank, science or literary prizes, National Assessments of Educational Performance (NAEP) scores, Advanced Placement (AP) course enrollments and test scores of an individual and or group of students on state assessments (Hombo, 2003). In this study, academic achievement refers to the scores on the NAEP mathematics assessment for 8th-grade students.

Ambiguous mathematical language -- refers to words or phrases found within a word problem that can have multiple interpretations. Ambiguity increases the range of possible interpretations of natural language and two primary forms are commonly found in mathematics: global and local. *Global ambiguity* means the whole sentence can have more than one interpretation. *Local ambiguity* means that part of a sentence can have more than one interpretation, but not the whole sentence (Inman, 2005).

Content strands of mathematics -- these are subject-matter content for the framework of NAEP mathematics assessments – number sense, measurement, data analysis and probability, geometry, and algebra. Each strand represents a specific subset of skills with an expected order of difficulty and has a direct relationship between overall assessment performance and proficiency on a specific skill (NAGB, 2005; NCTM, 2000).

Cut score -- the minimum score required for performance at each NAEP achievement level. NAEP cut scores are determined through a standard-setting process that convenes a cross-section of educators and interested citizens from across the nation. The group determines what students should know and be able to do relative to a body of content reflected in the framework. NAGB then adopts a set of cut scores on the scale that defines the lower boundaries of basic, proficient, and advanced levels of performance (NCES, 2006).

Differential item functioning (DIF) -- an item exhibits differential item functioning if the probability of doing well on the item depends on group membership, even after controlling for overall performance (NCES, 2006).

Item response theory (IRT) -- test analysis procedures that assume a mathematical model for the probability that an examinee will respond correctly to a specific test question, given the examinee's overall performance and characteristics of the questions on the test (NCES, 2006).

Problem types -- there are three types of item formats on the NAEP 2005 8th-grade mathematics assessment. The first item format is a standard multiple-choice with each item having five choices. The second item format is short-constructed response (SCR) and has two variations. In the first category of SCR questions, the student writes an answer in the space provided and it is scored dichotomously with full credit for a correct response and no credit for an incorrect response. The second category of SCR is polytomous items in which more than two responses are possible. Students answer multiple questions on data contained in one item or provide a rationale for a single response. The students may earn partial credit on this category of SCR questions. The third item format is the extended-construction response (ECR) questions. On these items students write out their responses to questions that ask for mathematical reasoning and justification for the students' problem solving. Items with ECR are scored using rubrics and students may be awarded on four different levels of credit for response (minimal, partial, satisfactory, or extended). In addition to coding all correct responses, NAEP codes items that students skipped as omitted (Arbaugh, Brown, Lynch, & McGraw, 2004; NAGB, 2004; NCES, 2006).

Language Complexity Categories – this refers to the difficulty of the words in the item used to convey the intentionality of the item (what is required to solve the problem). In this study, the Mathematics Assessment Language Framework (MALF) was used to classify 2005 NAEP 8th-grade Mathematics items into six language complexity categories. From most complex, the six categories are: graphics vocabulary, non-graphics vocabulary, operate-to-plan, convert-to-solve, draw/manipulate, and convert-only.

Mathematics literacy -- this refers to the amount and the nature of mathematic knowledge, how individuals obtain the knowledge, and the justification of the presence of knowledge through language and symbols (Resnick, 1989). In this study, *mathematics literacy* refers to the ability to use mathematic language to recognize and evaluate information whether written, video, or in conversation to make valued judgments in the contexts of daily life, employment, and personal decisions.

Mathematics language fluency -- Language fluency in mathematics refers to the ability of a student to understand what is required in a mathematics test item and delineates the differences between language used on a daily basis and the language associated with problem solving. This includes the development and communication of ideas and particularly within quantitative relationships using observation, reasoned analysis, and prediction.

National Assessment of Educational Progress (NAEP) -- The NAEP assessments are considered a monitor of student achievement. These assessments report on progress and identify achievement gaps by gender, racial groups and other demographic factors and record changes in achievement over time. NAEP assessments were retooled by the National Assessment Governing Board in 1992 and in 1996 to reflect these national trends toward a basic national standard of what children should know in the 4th, 8th, and 12th grades (NAGB, 2004; Vinovskis, 1998; Walberg, 2003). The importance of NAEP as the only assessment that provided information for different geographic regions, individual states, and demographic population groups and comparative information on how students performed on state assessments versus national assessments raised its prominence from a reporting agent to a measurement tool for state accountability (Hombo, 2003; NCLB *FAQ's*, 2005).

No Child Left Behind -- NCLB is a federal law that requires the accountability for each state's academic achievement standards. It demands that student assessments measure progress against common expectations for student academic achievement and that achievement scores are disaggregated by race/ethnicity and socioeconomic groups to insure that all groups are meeting academic standards. All groups are expected to show yearly progress and states are mandated to have 100% of all students proficient on standards-based assessments by 2014 in order to continue to receive federal funding for education. States must implement supplemental services to student in schools that are unable to meet yearly progress toward this goal. Under NCLB, the Annual Yearly Progress (AYP) must apply to all subgroups of students with the expectation that this level of school and district accountability will close the achievement gap for minority students (NCLB, 2001 HR PB. L. No. 107-110. 1425; NCLB *FAQs*, 2005).

Plausible values -- Proficiency estimates for an individual NAEP respondent, drawn at random from a conditional distribution of potential scale scores for all students in the sample who have similar characteristics and identical patterns of item responses. NAEP usually assigns five plausible values to each respondent. The plausible values are not test scores for individuals in the usual sense; they are offered only as intermediary computations for calculating summary statistics for groups of students (NCES, 2006)

Propositional referents -- This term refers to words that perform the function as directions of time (e.g., when, where), location (e.g., next to, sequential), size (e.g., greater than, lesser than) and/or relationships such as which, how, was, or prepositions (e.g., above, between, over) and prepositional phrases (e.g., in order to,) in problem solving.

Released item -- Refers to a test question that has been made available to the public. After each assessment, NCES releases nearly one-third of the questions. Released questions often serve as models for teachers who wish to develop their own classroom assessments (NCES, 2006).

CHAPTER TWO

REVIEW OF THE LITERATURE

This chapter is divided into three sections. The first section reviews NAEP background material, included the 2005 test specification framework for mathematics. The second section reviews epistemological frameworks that have been developed to examine the influence of language on problem-solving ability. The third section reviews research studies that examine the influences of language factors on mathematics achievement. These studies are grouped into six categories: (1) reading word problems, (2) mathematical discourse, (3) mathematics vocabulary, (4) problem-solving transfer, (5) student perceptions of test items, and (6) several studies using NAEP data to examine content strands

NAEP Background

The National Assessment of Educational Progress (NAEP), the only nationally representative and continuing assessment of what America's students know and can do in core academic subjects, have been conducted since 1969 (National Center for Educational Statistics [NCES], 2006). The National Assessment Governing Board (NAGB), appointed by the Secretary of Education, sets policy for NAEP and is responsible for developing the framework and test specifications that serve as the blueprint for the assessments. The NAEP assessments use oversampling procedures to obtain large enough samples of subgroups so that mathematics performance among the subgroups can be distinguished within each grade level. The NAEP assessments are designed to be cross-sectional and report what a group of students are able to do at one point in time; individuals or cohorts of students cannot be tracked for performance data over time (Elementary and Secondary Education Act [ESEA], 1965).

The NAEP assessments have a strong influence on national curriculum in a number of ways. NAEP scores are used to validate an individual state's measure of yearly progress toward achieving state academic standards as part of the No Child Left Behind accountability measure called "Adequately Yearly Progress (AYP)". Adequate Yearly Progress is the minimum level of improvement that states, school districts, and schools must achieve each year (NCLB Act 2001). Second, states must necessarily follow the subject matter content used in the NAEP assessment because these assessments reflects national educational practices of what students are learning at specific grade levels. Finally, the NAEP achievement scores are used to compare student performance by state and on international assessments.

NCLB targeted K-12 science and mathematics education under *Title II Part B: The Math and Science Partnership Program.* The intent of NCLB is to improve student achievement across the grade levels by focusing on the content knowledge and teachings skills of the classroom teachers (NCLB, 2001). An important component in focusing on yearly improvement is the development of standards for benchmarking the performance of all students. The improvement standards are organized into frameworks by subject matter content and include grade level learning objectives, teaching guidelines, and assessment benchmarks. The frameworks provide the context for what students should know and learn at every grade level from kindergarten to the 12th grade (NAGB, 2003; NCES, 2005; NCTM, 1989, 2000).

There are four types of frameworks for mathematics, and all four are affected by NAEP in some way. The first framework is published by The National Council of Teachers of Mathematics (NCTM, 1989; 1996, 2000). The NCTM mathematics framework provided the initial core framework of principles and standards in teaching mathematics in 1989, and continues to makes recommendations for national and state standards by grade level and topic strands. For example, in reaction to the increased emphasis on accountability testing brought about by the requirements of NCLB, NCTM (2006) recently released a national outline of mathematics curriculum, national assessments, and mastery benchmarks. This outline, called the *Curriculum Focal Points for Pre-kindergarten through Grade 8 Mathematics,* document the key mathematical concepts and skills at each grade level with an emphasis on number and operations and spatial reasoning aimed to reduce discrepancies of grade-level expectations and learning objectives between states (NCTM, 2006).

The second type of framework is the set of specifications for the 2005 NAEP Mathematics Assessment which are written by the National Assessment Governing Board (NAGB, 2003). The NAEP specifications include descriptions of the mathematical content of the test, the types of test questions, and recommendations for administration of the test. The core concepts used for benchmarking testing objectives in the NAEP assessment are based on the mathematic principles and standards of the NCTM framework (NAGB, 2003). The NAEP framework delineates by content strand what is to be assessed at each grade level, unlike the state frameworks which emphasize what content should be taught at each grade level by mathematic strand.

The third type of framework consisted of individual state frameworks. States vary in setting benchmarks and academic content for different grade levels. Students from one state may and often have different learning objectives being assessed at different complexity levels and at different grade levels than those from another state. Even though 48 states have aligned their content standards in mathematics to those of the NCTM by 2000 (Swanson & Stevenson, 2002), there continues to be a broad range of individual state grade level expectations. For example, a study of 4th-grade mathematics curriculum of 10 states found mastery benchmarks ranged from a low of 26 benchmarks in South Carolina to a high of 89 benchmarks in Florida (Reyes, Dingman, McNaught, Regis, & Togashi, 2006).

The fourth type of framework is the international assessment framework which allows countries to cross-reference student achievement to global educational standards. The two most influential are the *Trends in Mathematics and Science Study* (TIMSS, 2000) and the *Program for International Assessment* (PISA, 2004). The TIMSS measures students' progress in mathematics and science achievement on a regular 4-year cycle for grades 4 and 8. The PISA assesses one subject in depth every year, focusing on the subject matter literacy of 15-year-old students in reading mathematics and science. For 2003, mathematics literacy and problem solving was assessed. Forty-one nations were included in both of these assessments. The United States uses the NAEP achievement data to compare the learning and teaching of national practices to global educational standards (Hombo, 2003; NAGB, 2003; NCLB FAQs, 2005; PISA, 2004; TIMSS-R, 2000).

Epistemological Frameworks

Mathematics has a specialized vocabulary which must be learned. Mathematics fluency depends on the ability to integrate words, symbols, and vocabulary to create meaning and communicate ideas. This ever changing format between words and symbols to arrive at innumerable solutions in problem solving is the essence of mathematical fluency (Levine & Reed, 2001). The literature reviewed in this section uses epistemological frameworks to examine issues of language contained within mathematical problem solving on academic achievement.

Epistemological frameworks for mathematics concentrate on categorizing the language complexities needed to master mathematical content by examining how linguistics features (the requisites and developmental functions) facilitate the attainment of proficiency in mathematics problem solving by students. The epistemological frameworks help to illuminate how potential breakdowns in vocabulary and semantic content of mathematics may affect achievement. The frameworks provide structures that outline linguistic complexities involved in becoming mathematically fluent. These epistemological frameworks have been used as the context to explore discrepancies between different subgroups by examining language involved in problem solving. Table 4 presents the chronology of epistemological frameworks that have influenced the research studies reviewed in this chapter.

In the early 1970s, Aiken (1971) reviewed the literature on the verbal factors in learning mathematics. At that time, positive correlations between verbal language and mathematics were considered by most researchers to be a function of general intelligence and ability. Aiken's review of the literature suggested that the vocabulary and syntax of word problems consistently interfered with students' ability to problem solve in mathematics. This issue, whether it is "innate" ability or an understanding of vocabulary, syntax, and readability of text that promotes successful problem solving and mathematics achievement, persists today.

Table 4
Chronology of Epistemological Frameworks used in Literature Review

Date	Author(s)	Framework
1942	Cronbach	Establishes the role of vocabulary in academic assessment measures.
1970-71	Aiken	Meta-analysis of research on the relationship between reading comprehension and mathematics problem-solving.
1978	Pactman & Riley	Structure for the teaching of mathematics vocabulary as part of daily instruction.
1981	Ciani	Instructional framework for providing specific reading comprehension strategies to improve mathematics problem solving.
1982	Mayer	Defines mathematic problem solving as an ability to classify problem-solution methods into schema, and transfer knowledge from known to novel problem-solving situation.
1985	Halliday	Outlines the role of academic discourse on mathematical understanding and the ability to use that language in problem-solving.
2000	Wakefield	Framework defines mathematics as its own language based on its unique structures, symbols and word order to convey meaning.
2001	Levine & Reed	Identifies the basic parameters of mathematics assessment into categories and discrete skills.

Cronbach (1942) developed a framework describing types of word knowledge needed for students to demonstrate understanding within academic subject matter. The purpose of Cronbach's framework was to define word knowledge and provide a context for evaluating vocabulary on diagnostic tests. According to Cronbach (1942), there were qualitative dimensions to vocabulary assessments like how many words a student may know, how refined a student's understanding was on a specific word, and what technical subject specific words were likely to cause difficulty in school if the student did not understand the meaning. Within the framework, five categories were used to determine the range and depth of student understanding of word knowledge: (1) generalization, (2) application, (3) breadth, (4) precision, and (5) availability.

More specific to the teaching of vocabulary in mathematics, Pachtman and Riley's (1978) framework was linked with the teaching of vocabulary connected with word problems. The purpose of the framework was to provide teachers with systematic instruction for teaching students to develop the relationship between mathematical vocabulary and mathematical concepts. Instruction was based on a structured overview for teaching the vocabulary related to word problems. This framework identified mathematics vocabulary necessary to solving word problems such as technical vocabulary, symbols, everyday words used in a mathematical context, general vocabulary, words with meaning specific to mathematics, and words representing mathematical concepts implied in the problem.

Ciani's (1981) framework outlined the reading comprehension skills inherent in becoming proficient in solving word problems in mathematics. This framework identified specific areas of reading instruction that teachers could use to facilitate the understanding of the correspondence of the mathematical symbols to mathematical words. Ciani presented a four-step reading process hierarchy, with comprehension skills at each level. Level 1 was word recognition skills, where terms are introduced in both language and math. The next level is literal meaning in which a student is able to decode symbols and attach definitions without conceptual understanding. The third level is interpretation of meaning where a student is able to recognize the symbol and describe its meaning in mathematical terms. The final level is application of meaning, where a student is able to solve word problems successfully based on a conceptual understanding of symbols and mathematical language.

Influential in explaining differences in problem-solving performances have been frameworks focusing on the identification of students' knowledge of problem schemas, level of conceptual understanding, and metacognitive skills. Mayer (1983), a primary researcher on schema acquisition, posited that learning can be viewed as change in schema organization in long-term memory based on experience. Schemas provide a framework that can be used to understand new information and to retrieve that information later when existing knowledge or skills are transferred to novel situations. The more complete one's schema for information, the better the encoding and retrieval of that information will be. In mathematic problem solving, Mayer (1992) suggested that poor schema development of problem-solution methods impairs academic performance due to the inability of students to apply skills in novel word problems.

The social-semiotic perspective defined by Halliday (1985) used an epistemological framework to examine the relationship of language to the social dynamics of academic achievement. The framework provided a background for understanding the linguistic expectancies required to learn, participate, and communicate within the learning environment, and to modify and adapt individual understanding through language. Halliday's (1985) theory characterized the social environment by place, level of formality and spontaneity, and type of discourse. This framework postulated a specific relationship between instructional practice, verbal discourse, and written text. Using this framework, an individual question or a recording of whole class instruction could be analyzed to determine the language needed to support the cognitive processes in learning.

Levine and Reed's (2001) framework identified the relevant skills universally measured by standardized mathematics assessments. All assessments, according to Levine and Reed, require quantitative reasoning, the manipulation of rules and symbols, and the ability to handle high density of ideas per number of words. Assessment words must be combined with abstract symbols; comprehension is dependent upon the student's ability to memorize symbols and discern the relationships between the symbols and words. According to Levine and Reed, mathematics is often taught and assessed using non-verbal logic, specialized vocabulary, and applying nuanced differences between ordinary language and word-problem language. The basic skills in the framework included number knowledge, mathematics facts, mathematic notation, math vocabulary and verbal concepts, concept formation, problems solving, estimation, and application.

Wakefield's framework (2000) defined the foundational social-linguistic constructs of mathematical language that identifies the interdependence of words, symbols, and expressions used to construct meaning and communicate ideas. Table 5 defines Wakefield's 10 attributes that designate mathematics as a specialized language. A detailed description of Wakefield's framework is outlined in Chapter One.

Wakefield's attributes characterize the components of mathematics that distinguish it as a separate language. This framework does not provide objective measures which can be used to quantify the words, terminology, and vocabulary that are key factors in the communication process nor does the framework identify how these attributes can be used to evaluate mathematical performance by students. Wakefield's epistemological framework was adapted for use in the present study to create a framework for classifying items by language categories on the 2005 NAEP mathematics assessment.

Table 5
Wakefield's Framework of the Attributes of Mathematical Language

Attributes and/or Characteristics
1. Abstractions are used to communicate.
2. Symbols and rules are uniform and consistent.
3. Expressions are linear and serial.
4. Memorization of symbols and rules are required.
5. Continuum of experiences requires translations and interpretation.
6. Meaning is influenced by symbol order.
7. Communication requires encoding and decoding.
8. Understanding increases with practice.
9. Increasing intuition, insightfulness and spontaneity accompany fluency.
10. Possibilities of expression is infinite.

The Influence of Language Factors on Mathematics Achievement

The epistemological frameworks reviewed in the previous section generated a number of studies that examined the influence of language on mathematics achievement. The research is organized into the following subsections: (1) reading word problems; (2)

mathematical discourse, (3) mathematics vocabulary, (4) problem-solving transfer, (5) student perceptions of test items, and (6) several studies that examine NAEP achievement by content strands.

Reading Word Problems

The studies in this section explore the construct of mathematical fluency as the relationship between mathematical language and mathematic computation skills based on Ciani's (1981) framework that identified specific areas of reading instruction that teachers could use to help facilitate the understanding of the correspondence of mathematical symbols to mathematical words, researchers have examined the structures of language in reading mathematics and word problems.

Leong and Jerrod (2001) examined word problems involving two different linguistic structures found in mathematics word problems. This experimental study investigated the effect these structures had on elementary students' ability to understand and solve mathematical word problems. The first structure can be described as the literal representation of the information needed to solve word problems. There is little inferential or extraneous information within this type of problem. The second structure dealt with the adequacy of linguistic information provided to solve a word problem. This structure involves word problems with inconsistent language which require higher levels of reading comprehension to understand the inferential subtexts and/or inverse relationship of words that differ from daily language.

The types of information contained in these word problems were subdivided into three groups: adequate information, inadequate information, and redundant information. A supplementary battery of tests in general ability, vocabulary, reading comprehension, mathematical concepts, and working memory was administered to determine the effects of nonmathematical ability on mathematical problem solving.

The study was conducted with 91 elementary students in grades 3–5 in two schools in western Canada. All students were administered the *Canadian Test of Basic Skills* (CTBS). Based on the Total mean scores on the CTBS, students were divided into two groups of “more able” and “less able” within each grade level. Students within each group were given two tasks. The first task examined the effects of consistent and inconsistent language on the students’ ability to successfully solve word problems. Twenty-four problems were used for this measure, equally divided between the two linguistic features. The second task examined the effects of language information on students’ ability to solve word problems. Thirty-six problems were used, divided into 12 problems containing just enough information (JE), 12 problems with insufficient information (NE), and 12 problems with redundant information (NN). All items were administered in small groups to the students at their schools. Students were credited for correct representation of the problems regardless of accuracy of calculations.

For consistent versus inconsistent word problems, a 3 (grade: 3rd, 4th or 5th) x 2 (ability: high or low) x 2 (problem type: consistent or inconsistent) ANCOVA using the CTBS as the covariate found significant main effects for grade and problem type. Word problems with inconsistent information were more difficult to solve than those with consistent information despite ability grouping. For adequate versus inadequate word problems, a 3 (grade: 3rd, 4th, or 5th) x 2 (ability: high or low) x 3 (problem type: NN, JE, NN) ANCOVA showed significant main effects for grade, ability, and problem type in understanding and solving mathematic word problems in elementary school. The results suggested an interaction between problem type and ability level where there were the expected grade level differences on the language information task with student in the lower grades having a more difficult time with the word problems than students in higher grades. Finally, a stepwise regression analysis indicated that scores on the adequate/inadequate linguistic information task was the most predictive of mathematical problem solving, accounting for 58% of the variation followed by chronological age (7%), consistent/inconsistent information (5.8%), and general ability (1.4%). This study suggested that specific language fluency skills have a strong impact on mathematical problem solving.

Hegarty, Mayer, and Monk's (1995) experimental study compared the reading comprehension of successful and unsuccessful problem solvers of two-step arithmetic word problems. It was hypothesized that one comprehension strategy, the direct translation approach, would be more consistently used by unsuccessful problem solvers, and the second strategy, the problem-model approach, would be used more consistently by successful problem solvers. Two experiments were conducted. The first experiment examined different patterns of eye-fixation on the premise that unsuccessful problem solvers would be more inclined to look at numbers and relational terms whereas successful problem solvers would be more likely to look at variables such as names when they reread a part of a problem. The participants were 38 undergraduates from a university psychology subject pool. Each participant was randomly assigned to a test version and tested individually. Test questions ($N = 48$) were displayed on a screen, and participants were videotaped discussing how they would solve each problem. There were no time limitations. Eye fixations were recorded with special digital equipment called a

Vaxstation. The analysis suggested that unsuccessful problem solvers relied more significantly on rereading number and relational words in a word problem than non-successful problem solvers $t(14)= 2.37$, $p<.05$. The analysis suggested unsuccessful problem solvers struggle to construct a representation of the problem. Contrary, to the initial hypothesis of the researchers that unsuccessful problem solvers fixates on numbers more than successful problems solvers, both groups (successful problem $t(7) = 7.18$, $p<.01$; unsuccessful problems $t(7) = 6.79$, $p <.01$) fixated on more numbers than any other variable such as names and relational words.

The second experiment compared how successful and unsuccessful problem solvers remember story problems they had solved. This experiment was based on the premise that successful problem solvers would more likely remember the situation described in the problem whereas unsuccessful problems solvers would more likely remember the key words such as less or more. The participants included 37 undergraduates from a university psychology subject pool. Students were given a problem-solving test and one minute to solve each problem. Following the problem-solving test, a recall test was given to determine if there was a difference between successful and unsuccessful problem solvers.

In a 2 (groups: high, low) x 2 (types of questions: literal, semantic) ANOVA, it was found that there were significant differences between the two groups. Confirming the prediction, the successful problem solvers were more able to recall the situation in which the arithmetic problem was set and less able to recall the literal numerical details; for the unsuccessful problem solvers, the inverse was true. These experiments provided evidence that successful and unsuccessful problem solvers tend to use qualitatively different strategies. It also suggested that while the direct-translation strategy may be effective for many word problems at the elementary level, using a situational problem-solving strategy may be a key component to comprehension when math language becomes more complex.

Kelly and Mousley (2001) investigated the impact of reading level on solving mathematical word problems with deaf college students. This quasi-experimental study examined the effect of measured reading levels on students' ability to solve mathematical word problems with two versions: numeric/graphic representations or words only. The performances of deaf college students (n = 33) were compared to a group of hearing college students (n = 11) at the same college. The deaf college students were administered the reading comprehension subtest of the *California Achievement Test* and were divided into ability groupings of low, medium, and high reading ranges. The comparison group of hearing college students reading scores on the ACT was used as the pretest score. The hearing college students' mean scores were in the mid-range of reading ability. All students were enrolled in 1st or 2nd year mathematics requisite course at a northeastern university.

All 44 students were given three sets of problems sequenced for increasing mathematics complexity involving computation of geometric dimensions. Half of the problems were shown with numeric/graphic representations. The other half were given a corresponding word-only version of each numeric/graphic problem except for differences in actual numbers being calculated. Word problems were designed and written with short sentences and literal descriptions, and contained no language structures that had been shown to impact reading comprehension for deaf students such as conditionals (if-when), comparatives (greater, least), inferential (because, could) or pronouns (it, something).

The reading passages in the 2nd and 3rd sets were longer due to the amount of information needed to describe more complex shapes. Readability for the three sets of problems ranged from a low grade level of 3.0 to a high grade level of 6.7.

The results suggested that the level of complexity of the language influenced the ability to solve mathematical problems in both deaf and college students regardless of whether information is presented in numeric/graphic representation or word-only versions. A 4 (deaf: low, medium, high; and hearing) x 3 (levels of complexity: low, medium, high) ANOVA showed a significant interaction between group performances and word problem complexity that shows the performance patterns of the four groups were different with respect to the three sets of word problems. The post hoc tests revealed that all three levels of deaf students performed significantly lower than hearing students on word-only problems with computation errors exceeding procedural. There was a higher incidence of omissions on word-only problems and graphic representations by deaf students with 20% to 48% of all problems omitted compared to no omissions by the hearing students.

Additional analysis compared the actual number of correct graphic representation problems with those for the corresponding word-only versions by ability group (low, middle, high). The mean conditional probabilities for solving corresponding word problems correctly when numeric/graphic problems were solved correctly suggested that the deaf students' ability to solve the word-only versions was weaker than the hearing students. The probability of deaf students answering the corresponding word-only version was lower when the level of reading became more complex.

Results suggested that although reading comprehension was a factor, the rate of omissions had the greatest impact on the overall performance of deaf students to solve mathematical word problems. Follow up interviews suggested that students declined to attempt the word-only problems because of prior negative experiences. The interview findings suggested that students' belief that problems with numeric/graphic representations were easier improved the rate of success.

Mathematical Discourse

Halliday's (1985) linguistic theory of learning used a socio-semiotic framework to examine the relationship of language to the social dynamics of academic achievement. His theory centers on student discourse and the language used to define student thinking about mathematic content. The following studies investigated the relationship of ability to discuss mathematics concepts to written performance assessment on measures of mathematics achievement.

Moss and Case (1999) investigated the effect of instructional practice on developing conceptual understanding of rational numbers as assessed through student discourse and problem solving in an quasi-experimental study of fourth grade students ($N = 29$). The researchers designed the *Rational Number Test* to assess conceptual understanding of fractions, decimals and percents, and the proportional relationships among them. The experimental group received twenty, 40-minute instructional sessions on average of one a week over five months. The control group received twenty five, 40-minute instructional sessions during the regular allotted math period. Both teachers used manipulatives, group work, and class discussions. The experimental group of students ($n = 16$) were given a specifically designed curriculum that focused on semantic knowledge and conceptual meaning of rational numbers; the control group (n = 13) received the traditional curriculum on rational numbers as established by district mandates.

The experimental curriculum introduced the concept of percent in a specific sequence. First, the students were introduced to exercises which developed vocabulary and mathematical terminology to define observations and exploration with manipulatives. Then, students developed strategies to solve problems involving calculating precise values. Finally, students were shown how to use algorithms in addition to alternate strategies devised by students. A similar format was followed for the introduction and teaching of decimals and fractions. The control group curriculum adhered to a program used by a mathematics textbook series adopted by the district. Fractions, decimals, and percents were taught with a direct instruction approach to teaching the rules and algorithms, with an opportunity for students to practice using manipulatives and game-like activities.

Test questions were presented individually in an interview format. The results from the pretest (k = 41 items) to the posttest (k = 45 items) interview were based on 12 questions on percentage, 13 questions on fractions, and 16 questions using decimals. The posttest was expanded to include four additional questions. Six subcategories were used to compare conceptual, semantic, and mathematical understanding of rational numbers: (1) nonstandard computation, (2) comparison and ordering, (3) misleading appearance (graphic representation), (4) word problems, (5) interchangeability of representations, and (6) standard computation. Students were interviewed three times: pre, midpoint, and post treatment. All questions were scored dichotomously.

A repeated measures ANOVA showed the experimental group statistically outperforming the control group in five categories: (1) nonstandard computation, (2) comparison and ordering, (3) misleading appearance (graphic representation), (4) word problems, and (5) interchangeability of representations. The control and experimental group performed equally in the category of standard computation. The results indicated that mathematical language fluency is an important factor in building abstract thinking. This study's mathematical vocabulary intervention, prior to the topic of instruction, suggests that developing the semantic knowledge of the learner in order to discuss the proportional representations accurately and to think in terms of rational number constructs may improve students' problem-solving ability.

Koponen, Mononen, Rasanen, and Ahonen (2006) examined basic numerical skills in children with specific language impairment (SLI) and how well linguistic factors explain the variance in these children's number skills. The performance of Finnish children with SLI (3rd grade, $n = 29$) was compared with that of children within the general population ranging from preschool to third grade (preschool, $n = 20$; 1st grade, $n = 47$; 2nd grade, $n = 40$; 3rd grade, $n = 33$). Numerical skills were characterized by a battery of tasks then divided into two categories. Tasks measuring verbal numerical skills included counting, fluency, and accuracy of single digit calculations; tasks measuring nonverbal numerical skills included comparison and estimation of numbers using numerals and play money. The students were assessed twice for 45 minutes, one using computer simulations (30 minutes) and the other using pencil and paper (15 minutes).

When the SLI group's mean score was compared to the control group's mean scores, the SLI group mean was similar to 1st grade students' verbal and nonverbal skills.

However, when the SLI group was subdivided into three groups based on their differential performance on the verbal and nonverbal tasks, the mean age and performance score of the SLI subgroups showed differences: 12 children showed difficulties in verbal and non-verbal number skills (V-/N-), 8 children showed difficulties in verbal only (V-/N+), 9 children showed no difficulties in verbal or nonverbal number skills (V+ /N+), and no child had difficulties solely in nonverbal number skills.

In verbal numerical skills, the SLI subgroup V+/N+ performed at the level of the third graders. The V-/N- and V-/N+ SLI subgroups scored comparably with 1st graders. In the nonverbal skills, the V+/N+ subgroup performed at the level of third graders and were currently studying third-grade texts. The V-/N+ subgroup performed a year below their educational age peers and were currently studying second grade texts. The V-/N- subgroup performed worse than first graders despite studying second grade texts.

This study concluded the numerical skills of children with SLI are very different and the differentials in performance cannot be explained fully by only referencing their language skills, nonverbal reasoning, or number-specific attributes. The researchers suggested that some language skills are associated with some number skills such as the development of calculation fluency. These skills share the underlying processing ability required to access the names and objects rapidly from long-term memory and that linguistic deficits may also negatively influence developing numerical skills.

Fuchs and Fuchs (2002) examined the functional performance of students with mathematics disabilities (MD), with and without reading disabilities (RD), on a range of mathematics problem-solving tasks involving arithmetic story problems, complex story problems, and real-world story problems. A hierarchy of mathematical problem-solving tasks was created based on the comparative features of the tasks including text-based features like words per question, sentences per question, words per sentence, verbs per question, and numbers per question. Problem solving was further divided into two categories with levels of difficulty assigned three kinds of word problems: arithmetic story problems, complex story problems, and real-world problems. The first category was math steps with four levels (essential data, nonessential detail, irrelevant numbers, and location of question in narrative); the second category was math skills with three levels (number facts, algorithms and applications).

Sixty-two fourth-grade students from three schools in a southeastern city were selected based on intelligence (90 or higher), and identification as having a mathematical disability based on goals on the Individualized Education Program. The students were given the *Test of Computational Fluency* in which students wrote answers to 25 second-grade addition and subtraction problems involving basic facts and algorithms. Students (n = 40) whose scores fell more than 1.5 standard deviations below district norms were included in the MD sample. These 40 students were given the oral reading segment of the *Comprehensive Reading Assessment Battery*. Students whose scores were more than one standard deviation below district norms were identified as having MD with comorbid RD (MD+RD; n = 22); the other students were identified with MD only (n = 18).

Three tests were administered to students in small groups by a special educator trained in the testing protocol over three sessions. Story problems were read aloud; students worked at their own pace and could request the tester to reread portions of the test. All story problems were assessed on two dimensions: operations (accurate math work), and problem solving strategies. The language used in the three types of story problems varied by complexity and level of abstraction. Arithmetic problems (*n* = 14) were one-step addition and subtraction problems involving sums or minuends of nine or less; students were given manipulatives to aid in solving the equations. Complex story problems (*n* = 10), based on the district's third-grade school curriculum, involved problems using lists and graphs. Real-world story problems (*n* = 10) were based on third-grade skills identified by third- and fourth-grade teachers as critical to student understanding. These problems were multi-paragraph narratives involving tabular and graphic information.

For a profile performance of students, a 2 (disability: MD versus MD + RD) x 3 (task by type: arithmetic, complex, real-world) x 2 (performance: operations, problem-solving) ANOVA was conducted. Of significance for this study, the main effect for performance of students with MD only was higher than students with MD + RD. In the main effect for tasks, the arithmetic story problems were easier than complex problems, but there was comparable difficulty between complex and real-world problems. In a three-way interaction of arithmetic story problems, the effects of the disability did not affect the performance level of operations or the level of problem solving. On complex and real-world story problems the effects of the disability status affected performance. Students with MD with or without RD scored comparably on operations; the MD scored higher than students with MD + RD on problem solving.

The results suggested that the performance of students decreased across the three problem-solving tasks as language complexity increased for both groups of students. Students with MD went from 75% accuracy for arithmetic story problems to 14% for complex story problems to 12% for real-world problems. For students with comorbid MD

+ RD, the percentage of accuracy was 55%, 8%, and 5%, respectively. The effects of the interaction of language with mathematical problem solving could possible contribute to lower performance. The results suggested that the competence level of students with MD in both decoding and comprehension in a relevant factor in explaining the differential in skill levels in the mathematical hierarchy devised by the researchers.

Huntsinger et al. (2000) investigated the influences of ethnicity, parents' beliefs, and parents' practice on mathematics achievement of primary school age Chinese-American and European-American children in a three-year longitudinal study. The participants began as kindergarteners. Forty, second generation Chinese-American and forty European-American children from upper-middle class, well-educated, two parent families (both parents had a minimum of a bachelor's degree) participated. The children were followed for four years. Initial student testing and parent interviews were conducted at the onset of the study. Data collection procedures were followed at three points, in the spring of each year during grades 1, 2 and 3. Students were tested annually using the *Sequential Assessment of Mathematics Inventories* (SAMI), and the *Peabody Picture Vocabulary Test*. Two in-home interviews were also conducted. One interview was held jointly with both parents; the second interview was with the child separately at the same time. After the interviews, parents were videotaped interacting with their children around a mathematics task. The children's teachers also completed a questionnaire at each time point. Interviews included both Likert scale and open-ended questions.

Repeated measure analysis was conducted for mathematics acquisition, vocabulary knowledge, and parental teaching methods. A 2 (ethnic group) x 2 (time of measurement) MANOVA on mathematics scores between Time 2 and Time 3 showed that the Chinese-American students obtained higher scores than their European-American peers on both measures and the differential gap increased in Time 3. In a two-way interaction of Time of Measurement x Ethnic Group, European-American children knew more vocabulary than Chinese-American children; however, the Chinese-American children's vocabulary growth was steeper ($M = 42.20$) than the slope for European-American children ($M = 29.77$), indicating that the Chinese-American children's vocabulary would be equal and surpass their European-American counterparts by fourth grade. Parents' mathematics teaching methods were measured using a 2 (ethnic group) x 3 (time measurement) MANCOVA. Chinese-American parents were consistently more systematic and formal in their mathematics instruction with their children, while the European-American parents became more informal as the child became older. A series of hierarchical multiple regressions were completed that examined the relationship between the early and late time points regarding parental beliefs and practices on mathematics performance. According to the researchers, both parental beliefs and practices at Time 1 predicted the variance in mathematic scores at Time 3 for all 80 children.

The results from this longitudinal study suggested that parental practices in teaching children mathematics in preschool and kindergarten had a significant impact on the children's grade 3 mathematics performance. More formal instruction was positively correlated to mathematics scores. Parental beliefs about their children's success and positive attitude toward mathematics at Time 1 corresponded to Time 3's mathematics scores. There were also cultural differences: Chinese-American parents believed that hard work preceded academic achievement while European-American parents believed that natural ability preceded academic achievement. Chinese parents were much more likely to give additional homework in mathematics to support classroom instruction.

The most significant cultural difference between the two groups was parental influence on children's vocabulary acquisition. In the initial vocabulary scores at Time 1, the Chinese-American children had much smaller receptive vocabulary scores when tested on the PIAT vocabulary test than the European-American children. By Time 3, the situation had changed. The Chinese-American children's gap was not significantly different than their peers. During Time 3 interviews, all of the Chinese-American parents discussed the importance of vocabulary goals and spending additional homework time helping their children acquire the correct mathematical language; none of the European-American children parents discussed vocabulary building. Researchers contended that the parental efforts in formal skill building, homework help, and vocabulary development in early childhood and kindergarten positively fostered mathematical competence and high achievement in the later years in school.

Mathematical Vocabulary

Pachtman and Riley's (1978) framework linked the teaching of vocabulary to word problems. The purpose of the framework was to provide teachers with instruction on how to teach students to recognize the relationship between mathematical terms and mathematical concepts. According to Cronbach (1942), there are qualitative dimensions to vocabulary assessments that measure how many words a student may know, how refined a student's understanding was on specific word, and what technical subject specific words were likely to cause difficulty in school if the student did not understand the meaning. Knowledge of the specific mathematic vocabulary and language used in word problems is critical in achieving proficient levels of mathematics (Aiken, 1972; Ciani, 1981; Jitendra, DiPipi, & Perron-Jones, 2002; NCTM, 1989, 2000). The following studies examined the relationship of mathematics vocabulary knowledge to student achievement.

Tatsuoka, Corter, and Tatsuoka (2004) completed a comparative analysis of mathematics achievement of 8th-grade students across 20 countries using data from the TIMSS-R (1999). The purpose of this study was to provide a framework for a diagnostic profile on how test-takers performed on the underlying knowledge and cognitive process skills required to answer problems on the TIMSS-R assessment. The researchers measured student mastery on 23 specific content knowledge and processing skills using a psychometric model called the Rule-Space Method (RSM). The RSM generates an "ideal" pattern of item response scores, and then measures the "ideal" against actual student response scores.

The researchers assembled a team of experts to create a mathematics framework to classify predetermined performance scores from the RSM into three attributes sets (knowledge, skill, and process) to explain student achievement scores. There were 163 test items recoded as either knowledge, skill, or process, the recoding process was completed with 99.5% agreement among the team of experts. A multiple regression analysis was then performed to predict item difficulties and derive estimated attribute mastery probabilities for students in 20 of the most industrial countries. The researchers suggested that by analyzing the various attributes for student mastery, a hypothesis could be formed about the teaching of curriculum, skills, and culture by country.

A hierarchal multiple regression was performed and an adjusted R^2 value of .87 showed that the coded attribute composition fairly represented the predicted level of difficulty across countries. Attribute mastery probabilities were standardized to compare performance on single attributes across countries grouped in three categories of knowledge, skill, and process. A second comparison across countries was completed by analyzing a specific country's performance to the mean item percent score. Three composite variables were developed to investigate comparative achievement across countries: process, spatial, and reading. This second score was used to examine discrepancies between different country's approaches in teaching specific skills and topics across the mathematical strands assessed by TIMSS.

General patterns of mastery of subject matter were correlated by country based on student performance. Results were presented for all 20 countries; here, just the results for the United States are presented. In overall mathematic achievement, the students in the U.S. ranked 17th. Specifically, U.S. students showed relative weakness compared to other industrialized countries in the areas of geometry (ranked 18 out of 20 countries) and fractions (ranked 13 of 20 countries), but showed strength in the areas in the areas of algebra and in computational skills (5 of 20 for both). In a second comparison of composite variables based on three mathematical skills (process, spatial, and reading), the student in the US ranked 13/20 in process skills, 16/20 in spatial and 11/20 in reading tasks.

An analysis of attributes and skills found that geometry correlated highly with attributes measuring higher-order thinking skills including use of logical and proportional reasoning, application of knowledge, and processing data. The results indicated that geometry may be a better topic than algebra to teach the skills of logical reasoning and other higher-order skills, as those countries who have instituted middle school geometry into the curriculum performed highly on the TIMMS-R. The researchers suggested that a shift of the emphasis from algebra to geometry in current US mathematics curricula could improve student performance.

Lachance and Confrey (1995) conducted a quasi-experimental study with an intact class of 5th-grade students ($N = 20$) to evaluate a new type of mathematical curriculum. This curriculum was designed to develop understanding of decimal notation through contextual problem solving by making connections between decimals and multiplicative constructs. The curriculum is based on the mathematical concept of splitting action (how to group sets to show relationships), a key concept to develop multiplication, division, and ratio. This curriculum to teach rational numbers was piloted over three years and was initially introduced to the fifth-grade students when they were third graders. This study focused on the introduction of decimal notation over a six-week period in the final year of the pilot program. Students were given three open-ended problems constructed to develop the understanding of decimals and to connect new concepts to earlier multiplicative constructs of ratio and fractions through strategizing solutions individually and in various discussion groups.

Students were assessed through a series of pre-and-post written assessments and interviews. Assessment items were grouped into four tasks: meaning of decimal notation, ordering tasks, converting fraction into decimals, and computation with decimals. The assessment items were taken from four previous research studies on decimal instruction to establish comparative results between the piloted curriculum and other studies involving the performance of students after decimal instruction.

A paired *t* test showed the students made highly significant gains ($p \leq .001$) between the pre-and-post tests averaging 15.5 (pre) to 80.8 (post) on a scaled score. Compared to students in previous studies, a greater percentage of students were considered expert by consistently performed decimal number task correctly. The researchers contended that the development of ratio and proportional reasoning established early in the curriculum along with opportunities to develop the mathematical language to explain student thinking provided strategies to connect and apply knowledge.

Abedi and Lord (2001) investigated the importance of language in student test performance through an experimental study of the performance of 8^{th}-grade students (N = 1,174) on two word problems tests from:(1) the 1992 NAEP main mathematics assessment and (2) a parallel form that modified linguistic structures in the test items. The primary focus of the study was identifying differences in student performance of English language learners and proficient speakers of English on the modified test items. Modification of the math items was based on six linguistic features: (1) familiarity of non-math vocabulary – infrequent words were changed; (2) voice of verb phrase – active from passive tense changed to active; (3) length of nominal – shortened; (4) conditional clauses – replaced with separate sentence and direct literal language; (5) relative clauses – removed; question phrases – rephrased to simple directional questions; and (6) abstract and/or impersonal presentations – made concrete and personalized.

Two field tests were conducted using the parallel form of the NAEP math assessment. The first tested the perceptions of students (N = 36) to the parallel form of the mathematics content by interviewing students to read and compare pairs of similar math problems – original and revised – and to select the one which would seem easier. Students chose the revised items 63% of the time. The second field test examined the impact of the linguistically-revised items on students' mathematics performance using 20 NAEP released and linguistically-simplified items. For the second field test, 1,174 8^{th}-grade students from 11 schools in greater Los Angeles were selected for the sample. Students were given one of the two versions of the test for comparative analysis; if the A version had a non-modified item; the B version would have the same item revised. The two books were rated as similar based on four criteria: type and number of linguistic complexities, graphic representations, number of mathematical strands, and difficulty level. Data were analyzed using the following variables: language proficiency (ELL/ Non-ELL), social economic status (high, low), gender, booklet (Form A or B), and type of math class (ELL math, low math, average math, high math, algebra, honors algebra).

Three different ANOVAs were run to investigate the impact of the linguistic modifications on the different subgroups. Groups performed as expected with students in higher level mathematics classes outperforming the students in lower levels mathematics classes. The first model, item type (original and revised) by ELL classification (English learner and proficient English speaker), showed a significant impact of the revised items on student performance in both groups with proficient English speakers outperforming English learners. The second model, item type (original and revised) by SES, indicated that the higher SES students performed better than the lower SES students on both revised and original items. The third model, item type (original and revised) by math class (low, average, ESL, high, algebra, honors algebra), indicated that students performed differently depending on level with high algebra and honors algebra performing higher than low and average students on both the revised and the original items. However, students in the average groups had higher scores on the revised items, and students in the high algebra and honors algebra higher scores on the original items

A second analysis compared the performance of students on linguistically simplified items and original items by devising a measure of percentage of average gain score improvement. The results suggested the overall gains for the total sample using modified testing items was 2.9%; ELL students (3.7%) benefited more than non-ELL (2.45%), and students from low SES (3.3%) benefited more than student in high SES (2.6%). Among the different levels of math class, students in low and average classes showed the greatest gains 6.7% and 6.6%, respectively, while students in the highest level, honors algebra, had a slight negative gain (-0.8%). Students in the ESL class made slight improvements on revised items (0.9%). This study suggests that there is relationship between the ability to read and comprehend the text with the ability to problem solve. Two important findings of this study were: (1) differences in linguistic structure of math word problems can affect student performance; and (2) changes in the complexity of the language can possibly benefit those most affected in lower achieving segments of the middle school population. These studies suggest that the lack of understanding of specific mathematical terms and/or facility with vocabulary used during mathematics instruction interferes with the student's ability to problem solve beyond literal one-step tasks.

Problem-solving Transfer

The next set of studies examined mathematical problem solving as a form of transfer which requires students to transfer previously mastered problem-solution rules to novel situations. Accordingly, students who have poor mathematical language may have not developed appropriate schemas or classification systems to group problems into types that require similar problem-solution strategies (Mayer, 1992).

In a quasi-experimental study, Fuchs, Fuchs, Prentice, Hamlett, Finelli, and Coury (2004) examined: (1) the effects of schema-based instruction (SBI) on promoting mathematical problem-solving; (2) the effectiveness of explicit schema-based instruction on the development of mathematical problem solving; and (3) the impact of guided sorting practice on schema development and problem-solving skills. This study sought to broaden SBI by defining and refining two types of schemas: problem-type schemas and superficial-features schemas. The problem-type schemas had four problem types (shopping lists, half, buying bags, and pictographs) with three levels (immediate transfer, near transfer, and far transfer). Four superficial-features schemas were included: format, key vocabulary, question, and scope. Specifically, this study investigated if a SBI variation of explicitly teaching transfer skills to recognize how superficial features altered problems (without changing problem type and solution method) would improve mathematic problem-solving performance of third-grade students.

Twenty-four teachers in six schools within one district volunteered to participate in the study. Student participants were the 366 children in these classrooms who were present for the pre- and post-testing. In a stratified random assignment by school, eight teachers were designated to be in one of three treatment conditions: contrast, SBI, and

SBI plus sorting. Based on pretreatment scores on a transfer problem-solving measure, students were designated low, average, and high performing. Student distribution by performance levels and demographic factors (race, gender, SES, ELL) were comparable across the three treatment groups.

All three conditions shared three instructional features: district curriculum with a weekly pacing guide, a proscribed basal text, and a three-week unit on general mathematic problem-solving strategies. The four problem types used in the study were selected from the basal texts including: shopping list (multi-step addition/subtraction), halving and sharing groups of items (introductory division), buying groups of items (introductory multiplication), and pictographs (graphic representation). In addition, all students received three weeks of instruction (two lessons per week) designed by the researchers and taught by the research assistants in a whole-class format on basic problem-solving strategies including making sure answers made sense, lining up numbers from text to perform mathematical operations, checking computation, labeling work with words, monetary signs, and mathematical symbols.

All SBI and SBI plus sorting received 26 SBI lessons over 13 weeks grouped in four three-week units by problem type and one week of review of the previous lessons. All units had four lessons of problem-solutions methods by problem type and two lessons developing superficial-feature schemas that make a problem appear novel without altering problem type or solution method. All lessons were taught by the research assistants and used examples from the basal with explicit instruction, dyadic practice, independent practice, and homework. In the SBI plus sorting group, students were given instruction in categorizing problems by problem type and superficial features.

Student schema development to recognize problem types by superficial features was measured in three problem-solving situations: immediate transfer, near transfer, and far transfer. The immediate transfer measure was similar to problems used in the problem-solution method instruction with novel cover stories. The scoring rubric awarded points for correct computation and correct labeling of each step of the problem. The near transfer included one novelty feature per problem type. The scoring rubric awarded points for correct computation and correct labeling of each step of the problem including monetary signs and symbols. Far transfer introduced multiple sources of novelty by: (1) combining all four problem types, (2) varying all four superficial-features, (3) adding irrelevant texts and numbers, and (4) assessing six additional skills from the district curriculum. The measure was formatted to look like a commercial test to decrease association from the experimental treatment.

Data were analyzed using a 2-way ANOVA of ability by treatment with change in scores from pretest to post test as the dependent variable. To evaluate the data for significant effects, a pairwise comparison was used and effect sizes (ESs) for practical significance on the problem-solving scores were computed. A regression analysis was performed to explore the relationship of schema development to problem-solving development.

The results showed that high ability students performed better than average and low ability students regardless of treatment group on pretest/posttest gain scores. Across the three ability levels, students who received SBI and SBI plus sorting outperformed students in the contrast group in schema development by problem type mean scores: SBI, 12.39; SBI plus sorting, 12.72; contrast 8.67; and by superficial-feature schema: SBI, 4.06; SBI plus sorting, 4.67; contrast, 2.94. In problem-solving transfer measures, students in the contrast group showed the least gains when compared to students in the SBI and SBI plus sorting groups. On immediate transfer problems, the SBI and SBI plus sorting outperformed the contrast group by approximately 24 points, on near transfer problems by 15 points, and on far transfer by 12 points. Differences in the SBI group and the SBI plus sorting were statistically insignificant for immediate transfer, near transfer, and far transfer measures with actual gain score differences less than one point between the two groups in all three measures. No statistically significant correlation was found between schema development and problem-solving learning. A stepwise regression using pretreatment problem solving and schema development to predict post treatment problem solving found that schema development accounted for approximately 39% of the variance versus 2.9 for initial problem-solving competence on immediate- and near-transfer problems. On far transfer problems, schema development accounted for 11% of the variance versus 18% for initial pretreatment problem solving.

The study suggests that schema training can positively affect problem-solving scores across the range of ability levels of third-grade students even on measures which do not resemble tasks used during instruction. The research suggests SBI provided concurrently with general mathematic strategies can induce schema development in young children across all performance ability levels. Guided sorting practice provided no additional advantage to students who received the sorting practice instruction along with SBI except with students with very low achievement profiles (students with disabilities) who showed a small effect size (.37) in the far transfer achievement measure.

Student Perceptions of Test Items

Students' perceptions of test items and their perceived ability to successfully complete the test item have been acknowledged as a factor in mathematics achievement (Fenema, 1989; Ladson-Billings, 1997; Tate, 1997). The following group of studies suggests that the visual representations found in word problems (e.g., charts, diagrams, and illustrations) may play a crucial role in students even attempting a test item and having an opportunity to complete and item.

Blinko (2004) examined the perceived discrepancies between student responses to assessment questions when there were representational features embedded such: pictures, graphs or diagrams, versus abstract items that were words only using variety of mathematical problem contexts. The researcher suggested that the context of the problem affected the students' performance. Mathematical contexts were defined as one of three different types: (1) realistic based on what the child have met or experienced, (2) abstract based on knowledge of mathematical terms and concepts, and (3) graphic based on models, diagrams and/ number lines.

This study examined the influence graphic representation may have on assessment item performance by asking student to sort items by level of difficulty. Students were given items with similar mathematic content knowledge and skills to sort based on the following four categories: (1) realistic – words only; (2) realistic – graphic representation; (3) abstract – words only; and (4) abstract –graphic representation. Fourteen fifth-grade students from four schools were selected by their teacher to be of average ability and socially adept at discussing their thinking articulately. Students were given a series of questions to sort, based on first impressions of primarily looking at the graphics; they were then given an opportunity to re-sort after reading the questions. Questions were sorted into three piles based on the student's perceived accessibility into the following groups: questions most likely to attempt, those that were too hard to attempt, and unsure. Students were interviewed individually to discuss their response to sorting prior to and after reading the questions.

In the initial sort, students were presented with two versions of a question involving the same mathematical skills and knowledge. Without specific direct instructions to read the problems, students were told to select problems that they thought they could solve from a stack on cards. The majority of students (93%) chose the version of each question with the visual presentation as being more accessible and easier to do. In the second sort, students were given similar problems and told to read each question closely and then sort the problems into two stacks of "could solve" or "could not solve" easily. The students sorted the questions, after reading each of the problems. In the second sort the majority of students (79%) felt confident about attempting the question with or without illustration. The evidence suggests that the context of the question (realistic versus abstract) may be less important than graphic representation in influencing students' decision-making to attempt a test item.

Because an analysis of the TIMSS-R (2001) performance of 8th-graders suggested that Singapore's 8th-graders scored higher than U.S. 8th-graders on all items in the strands of number sense and algebra, Beckmann (2004) investigated the use of elementary mathematics texts as a factor in Singapore's students having the highest academic achievement. Elementary mathematics texts and workbooks are designed by the Singaporean Curriculum Planning and Development Division, Ministry of Education, and are the only adopted texts for the nation. Compared to the major elementary school mathematics texts in the U.S., there is a heavy use of pictures and diagrams to accompany problems. The use of pictorial aids may help students make sense of problems as part of solution strategies. Unlike most texts found in the U.S., Singapore's curricula materials place less emphasis on lengthy explanations, procedural aids like cartoon characters, and factoids unconnected to basic content.

The same types of pictures and diagrams were used repeatedly across problem types and grade levels to facilitate problem-solving skills and develop an understanding of mathematical concepts over time. Strip diagrams are one primary pictorial representation in Singaporean texts that helped students calculate some of the addition, subtraction, multiplication, division, fraction, and decimal story problems from grades 3-8. These same types of strip-diagrams were used across the strands enabling students to be involved in the study of complex algebra problems in fourth and fifth grade as well as more traditional problems involving Number and Operations. Beckmann contended that the effectiveness of the Singaporean instruction was based on the problem-solving methods of pictorial representation and this allowed for those students to be exposed to more challenging and linguistically complex story problems.

Examination of NAEP of Content Strands

The culture of mathematics instruction, curriculum, assessment, and pedagogy are based on the experiences of the White middle class (Ladson-Billings, 1997; Pennington, 2000; Smith, 2004). Those students who do not share in White middle class social norms may be at a disadvantage in developing high levels of mathematical skills because of bias overlooked within the structure of the tests. Differential item functioning (DIF) is a statistical procedure that looks for cultural and gender bias within NAEP to ensure that individual items assess equally across groups. The following studies examined the five content strand of mathematics (Number and Operation, measurement, geometry, data analysis, and algebra) for differences in student performance and difficulty level.

Beginning around 1990, reform movement in mathematics intended to rectify past inequities by offering all students a mathematics education centered on problem solving and critical thinking (NCTM, 1989, 1995, 2000). This focus may actually have exacerbated the gap in achievement between racial/ethnic groups (Boaler, 2002; Lubienski, 2000, 2002; Wenglinsky, 2004). As assessments and instructional practices became more aligned with reform movement standards and expectations, disparities were created regarding how reform practices should be instituted at the school and classroom level according to race and socioeconomic status (Cohen & Hill, 2000; Lee, 2004; Lubienski, 2004).

Lubienski and Shelly (2003) examined trends related to race, SES, and mathematics achievement using NAEP data. This descriptive study investigated how the achievement gaps in mathematics had changed over time among White, Black, and Hispanic students. In addition, the study sought to determine if there were race-related differences in the implementation of NCTM mathematics reform-oriented instructional practices and to identify instructional factors that may correlate to race-related achievement gaps. In order to do this, Lubienski and Shelly used data from the 1990, 1992, 1996, and 2000 NAEP main mathematics assessments regarding 4th, 8th and 12th graders. The 1990 samples had 8,072 students divided equally among the three grade levels. Samples for 1992 and 1996 totaled approximately 21,000 students; the 2000 student sample size was over 42,000. The 2000 NAEP assessment was the last year in which 12th graders were assessed as part of the national representative sample until 2005 (NAEP, 2005).

Achievement and survey data were taken from the restricted-use main NAEP mathematics CD-ROM and the NAEP's web-based data tool. The NAEP web-based tool feature for cross-tabulation was used to calculate mean and standard errors for student achievement and analyze the differences among racial groups. The variables in the analysis included mathematics achievement, student demographics, student course-taking practices, attitudes toward mathematics, teachers' instructional practices, and teacher's educational backgrounds. Between 1990 and 2000, mean scores for mathematics achievement and instruction-related practices were compared for race-related achievement gaps across the five strands.

A SES composite variable was created using seven factors: types of reading materials, computer and internet access in the home, extent to which studies are discussed at home, school lunch, Title 1 eligibility, and education of the mother and father. In fourth grade, parent education levels were not reported so the SES composite variable consisted of only five factors. SES quartiles based on the weighted sample of students were examined for race/ethnicity differences. The higher SES quartiles had a greater proportion of White students and the lower quartiles had a greater proportion of Black students across all five strands.

The largest gap differential among the content strands was between Hispanic and White students in the 4th and 12th grades in measurement. For 8th-grade students, statistics and probability showed the greatest difference between White and Hispanic students. Measurement showed the largest disparity in performance for all three grade levels between Black and White students. The 1996 and 2000 NAEP mathematics assessment showed race-related differences in two statements on the student survey: "There is only one correct way to solve a math problem" and "Learning mathematics is mostly memorizing facts." Black and Hispanic students were more than twice as likely to agree with these statements regardless of SES or grade level. Access to instructional practices as recommended by NCTM were similar for all students as reported in a teacher survey with two exceptions: White students were more likely to have access to calculators for daily use and tests than their non-White peers, and Black and Hispanic students were twice as likely to be assessed monthly by multiple-choice tests.

Lubienski and Shelly (2003) also reported differences in teacher quality, instructional access, and instructional practices. The analysis of teacher educational background revealed that there were no significant differences between access to credentialed teachers by demographics or SES; however, those who had teachers certified in mathematics tended to score an average of 14 points higher than students who did not have a certified mathematics teacher. An examination of patterns of student course taking for 8th and 12th graders showed that course taking was more associated with SES than race. Students in the lower SES groups took less mathematics than students in higher SES groups.

Instructional practices are factors that can be affected by educators and policy makers (Lubienski & Shelly, 2003; NCLB, 2001). Lubienski and Shelly contended that an examination of the disparities between groups' access to mathematical instruction provided an opportunity to identify factors of instructional practices that underlie the achievement gap.

Most prevalent was the emerging pattern that inequality still existed regarding access to grade-level curriculum. The study found that there were race-related disparities with student access to content based on NCTM (1995) mathematics standards mostly focused on the use and access to high problem-solving instruction. The researchers suggested that this lack of access to higher level math content may negatively impact achievement between Whites and their minority peers. In order to meet the *proficient* standard at each grade level on NAEP assessments, students must have the ability to answer complex and open-ended questions. Findings based on student beliefs and experiences taken from the teacher and student surveys suggested that lower-SES, Black and Hispanic students were being taught and assessed with an emphasis on basic skills.

Schulz et al. (2005) used a multi-stage content analysis to categorize the 2003 NAEP mathematics assessment items according to levels of difficulty within each of the content strands. This study examined if a relationship existed between performance test items by strand and achievement levels (basic, proficient, advanced) to an order of difficulty based on a hierarchal criterion of skill mastery on the NAEP mathematics assessment. Four assumptions were met prior to the reclassification of items: (1) multiple strands were defined within the test and that each strand addresses specific skills; (2) each strand had an expected order of difficulty; (3) scores within a strand represented a relationship between the overall test performance and proficiency on a specific skill; and (4) items are assigned based on content not by item statistics or level of difficulty.

In the first phase of the study, items were reclassified as either dichotomous with one correct answer (multiple choice and short constructed responses), or polytomous, scored with rubrics with three or more levels (extended response). Items were selected from three sources: (1) secured items on the NAEP 2000 mathematics assessment (*n* = 159), (2) released items on the NAEP 2000 mathematics assessment (*n* = 136), and (3) released items from previous NAEP mathematics assessments between 1990 and 1996 (*n* = 136). Released items were used to pilot the classification system and to train raters. Using Item Response Theory (IRT) models, items from the NAEP assessment were recalibrated to create domain characteristics based on the content strands. These models showed a relationship between the mastery level of each content strand (65% correct) and the achievement levels of the students (basic, proficient, and advance). Because of the relationship, strands could be ordered from hardest to easiest as follows: measurement, data analysis, geometry, algebra, and number sense, respectively.

In the second phase, the researchers designed new content domains using the difficulty order as a framework to organize and rate NAEP mathematic items into two criterion-referenced categories, introduction and mastery. Ratings were based on grade-level curriculum frameworks. Using a 7 point scale ranging “from below grade 5” to “above grade 9,” each test item was assigned three ratings. The first rating assigned the grade level when the skill should have been introduced to the average student. The second rating assigned a grade level when the same skill should have been mastered. The items were then ordered by difficulty based on the average instructional time between introduction and mastery. Items that were introduced prior to fifth grade and expected to be mastered by fifth grade were ranked the lowest. Items that were introduced in 8th-grade and projected to be mastered in later grades were ranked the highest.

In the last phase of the study, three item classifications were performed. The classification procedures defined ordered difficulty within each of the five NAEP content strands by mutually exclusive subsets of sequential skills. Using instructional timing (mean introduced or mean mastered ratings), three teams of curriculum specialists and researchers identified 26 subclassifications within the mathematics content strands thought to match the skills needed to perform at the highest levels of achievement on the NAEP 8th-grade assessments. Next, five teachers sorted NAEP 2000 secure items (n = 148) corresponding to the domain definitions of each of the sub-classifications within a given strand. An unclassified category was available if the teachers felt the item did not fall into one of the sub-classifications. There was unanimous agreement on 67% percent of the secure items and a majority agreement (3 out of 5 teachers) on 96% of the secure items. The agreement rate between preliminary classifications and teacher classifications was 91%. Eleven items were designated as unclassified. Final item classifications were made by the consensus of the domain content team with 95% inter-rater reliability to the teacher classifications (Schulz et al., 2005).

Results showed that within strands, sub-classifications were ranked in the same order by mean IRT item difficulty (*b*-value), mean introduced, and mean mastered ratings. Item classification was consistent by teachers. The depiction of item level statistics to the relationship of between achievement and mastery of a skill was unreliable. Some items appeared to differentiate between proficient and basic achievement levels; other items appeared to differentiate between proficient and advanced levels. Some items indicated partial mastery of sub-classification by lower level students and some of the skills needed to perform items were not included in the subclassification.

Procedures in the Schulz et al. (2005) study followed the general recommendation of Popham (1994) for validating instructional relevance. These include: (1) established domain definitions for mastery levels across strand, (2) proved consistency of domain definitions as qualified experts were able classify items based on definitions, (3) confirmed reliability by teacher classifications of items, (4) validated criterions levels to performance levels on the achievement test, and (5) verified usefulness and creditability of domains with teachers using a Likert rating scale. The researchers suggested that there were important practical implications to defining the expected order of difficulty within a learning strand on assessments for both educators and the general public.

Criterion-referenced mastery levels can possibly better explain what students can or cannot do at any given achievement level and show student growth in achievement through the mastery of difficulty-ordered sub-skills found in strands of subject-matter content (Shultz et al., 2005). This study suggested that there are factors within the NAEP mathematics assessments at the item level that may affect achievement of students across the content strands. These item level differences have strong educational implications with regard to instructional practices and accessibility to mathematics content by students. Mathematical language fluency may also have implications at the item level on the NAEP mathematics that has not been identified as a factor in achievement performance levels of students.

Summary

The literature review has examined language factors in math achievement. Most of this research examined the influence of problem-solving language on mathematics performance in the classroom and did not focus on the influence of mathematics language on assessment performance. Mathematically-specialized vocabulary must be learned. Words are combined with symbols, and comprehension depends on the student's ability to remember the relationships between words and symbols and use this knowledge to show mastery of mathematic content. The studies in this literature review do not investigate how potential breakdowns in vocabulary and the semantic content of mathematics may affect achievement directly nor do they examine discrepancies that may exist between different subgroups when reviewing the linguistic complexities involved in becoming mathematically fluent. This dissertation focused on the mathematical language needed to acquire mathematical literacy.

CHAPTER THREE

METHODOLOGY

This chapter contains five sections. The first section provides a brief overview of the study's design. The second section discusses the content analyses of the mathematics content strands, problem types and the Mathematics Assessment Language Framework for the 2005 NAEP 8th grade Mathematics Assessment. The second section describes the development of a Mathematics Language Assessment Framework (MLAF) classification scheme used in this study to classify 2005 NAEP mathematic items into language categories. The third section describes the sample procedures for students and the concomitant procedures for data analysis. The fourth section describes study procedures. The final section presents the statistical procedures for each of the research questions. Hereafter, for the purpose of readability, the 2005 NAEP 8th-grade Mathematics Assessment will be referred as the 2005 NAEP Math.

Design of the Study

This study was a secondary and content analysis of items on the 2005 NAEP Math. Typically, secondary analyses use existing data for reanalysis by asking new research questions or applying new statistical techniques not available at the time of the primary analysis (Heaton, 1998). In the current study, two statistical analyses were performed. The first analysis identified the extent of achievement gaps by strand disaggregated by race (White, Black, Hispanic, and Asian). The second analysis was a content analysis of items by problem type as assigned by the NAEP framework, and by language complexity on the Mathematics Language Assessment Framework created for this study. This study used the 2005 NAEP Math because it is the largest representative sample of mathematics assessment with a wide range of implications regarding educational policy and teaching and learning. The design of this study incorporated the unique feature of using national data collected by the National Center for Education Statistics. Both public-released data and secure-license data were used in accordance to the regulations defined by the National Assessment Governing Board (2005).

Content Analysis

For the content analysis, three areas were analyzed: (1) NAEP Math content strands, (2) problem types, (3) language complexity. The three areas are described below.

Content Strands. The 2005 NAEP Mathematics Assessment is based on five content strands that are universally used in the U.S. for K-12 instruction: (1) number and operation, (2) measurement, (3) geometry, (4) data analysis and probability, and (5) algebra. Each strand is further subdivided by objectives and mastery benchmarks intended for each grade-level assessed. All strands are assessed at every grade level; however, the distribution of items varies depending upon grade. Table 6 lists the mathematics content strands, learning objectives and an example of an 8th-grade mastery benchmark for the NAEP 2005 Math.

Table 6

Item Specifications by Strand, Learning Objective, and Exemplar Benchmark for 8th-grade 2005 NAEP Mathematics Assessment (NAGB, 2004)

Strand	Learning Objectives	Exemplar Benchmark
Number and Operations	Number sense	Apply representations of rational numbers (fractions, decimals, and percents)
	Estimation	Make estimates appropriate to a given situation analyzing the accuracy of results
	Number operations	Solve application problems involving rational number and operations
	Ratios and Proportional Reasoning	Use fractions to represents and express ratios and proportions
	Properties of Number and Operations	Recognize or use prime and composite numbers to solve problems
Measurement	Measuring physical attributes	Use appropriate measurement instrument to determine a given length, area, volume, angle, weight, or mass
	Systems of measurement	Estimate the measurement and the conversion factor of an object
Geometry	Dimensions of shape	Represent or describe a three-dimensional situation in a two dimensional drawing
	Transformation of shapes and preservations of properties	Identify and use the relationships of conservation of angle and proportionality of side length and perimeter
	Relationships between geometric figures	Use the Pythagorean theorem to solve problems
	Position and direction	Describe the intersection of two or more geometric figures in the plane
	Mathematical direction	Make and test a geometric conjecture about regular polygons

Table 6, continued

Item Specifications by Strand, Learning Objective, and Exemplar Benchmark for 8th-grade 2005 NAEP Mathematics Assessment (NAGB, 2004)

Strand	Learning Objectives	Exemplar Benchmark
Data Analysis and Probability	Data representation	Given a graph or a set of data, determine information is represented effectively
	Characteristics of data set	Calculate, use, or interpret central tendency
	Experiments and samples	Evaluate the design of an experiment
	Probability	Interpret probabilities within a context
Algebra	Patterns, relations, and functions	Identify functions as linear or nonlinear or contrast properties of functions from tables, graphs or equations
	Algebraic representations	Solve problems involving coordinate pairs
	Variables, expressions, and operations	Perform basic operations, using appropriate tools, on linear algebraic expressions
	Equations and inequalities	Solve problems using linear equations and inequalities with rational coefficients

Problem Types. Items were categorized into one of three problem types with in each of the strands: multiple choice, constructed response and extended response. These problem types are used to define the level of mathematical ability of students on the 2005 8th-grade NAEP Math across the five strands. The first problem type is standard multiple-choice with each item having five choices. The second problem type is short-constructed response (SCR) and had two variations. For the first category of SCR questions, the student writes an answer in the space provided. This item format is scored dichotomously with full credit for a correct response and no credit for an incorrect response. The second category of SCR is polytomous items in which more than two responses are possible. Students answer multiple questions on data contained in one item on the assessment or provide a rationale for a single response. The students may earn partial credit on this category of SCR questions. The third problem type is the extended-construction response (ECR) questions. On these items students write out their responses to questions that ask for mathematical reasoning and justification. Items with ECR are scored using rubrics and students are assigned one of four levels of credit (minimal, partial, satisfactory or correct). In addition to scoring all correct responses, NAEP also coded skipped items as omitted. Table 7 shows the items overall specifications of the 2005 NAEP Math. The breakdown of the three types of item formats were 50% multiple choice and 50% either SCR or ECR questions (Arbaugh et al., 2004; NAGB, 2004; NCES, 2005).

Language Complexity. To code NAEP items according to language demands, a Mathematics Language Assessment Framework (MLAF) was developed by the author using the mathematical language theory of Wakefield (2000) and the language epistemology of mathematical problem solving (Aiken, 1971, 1972; Ciani, 1981; Cronbach, 1942; Halliday, 1985; Levine & Reed, 2001; Mayer, 1982; Pachtman & Riley, 1978). The MLAF established new language category boundaries. Q-sort methodology (Stephenson, 1935) was used to define language category boundaries and the unique characteristics of each category. The items used to develop the MLAF were from the 2003 and 2005 NAEP release mathematics items; the items that were actually reclassified for this study were 2005 NAEP 8th-grade mathematics items from the secure CD-ROM.

Table 7

Distribution of Items on the 2005 NAEP Mathematics Assessment by Content Strand and Problem Type Established by NAGB (2004)

Content Strand and Problem Type	
Content Strand	
Number and Operation	20%
Measurement	15%
Geometry	20%
Data Analysis	15%
Algebra	30%
Problem Type	
Multiple Choice	50%
Constructed-Response	25%
Extended-Response	25%

Q-Sort Procedures. Q-sort methodology has typically been used to categorize and rank complex or partially overlapping qualitative statements into quantifiable units. Traditional Q-sort procedures use belief statements by participants in the study. The belief statements comprise the Q-set and are based on the theoretical model relevant to a particular study and are used in a way similar to a questionnaire. Typically, participants are asked to divide statements into three piles consisting of the statements “most like their beliefs”, “those most unlike their beliefs”, and those statements for which they have “neutral feelings”. A basic principle of the Q-sort methodology is that items are evaluated relative to each other. This is usually accomplished by providing the items on cards which the subject lays out and sorts into horizontally ordered category piles on a desk. Then those statements are rated and ranked against a standard measure to operationalize and scale attitude or belief statements. As data are confirmed by repeated Q-sorts, validity of the statements toward the subject of inquiry is established (McGowan & Brown, 1988; Stephenson, 1935).

A major activity in conducting a Q-sort is to establish the exclusiveness and inclusiveness of the categories. The criterion of exclusiveness, called the Q-standard, is met when the characteristics defining any single category do not overlap the characteristics of another category. For example in category A, the characteristics A1, A2,… An are not included as characteristics B1, B2,…., Bn in category B. The criterion of inclusiveness, called a Q-sort, stipulates that the category system has specific definitions for data that applies within that system. As part of the validation process, the Q-sort methodology involves the contrasting of the different Q-sorts compared to the Q-standard (Boyd, 1996).

As applied in this study, the Q-sort technique was adapted to develop and validate the MLAF. First, the 2003 NAEP 8th-grade mathematic released items were sorted into groups to examine if mutual exclusivity could be found among the items. This initial sort was the basis for the mathematics categorization scheme. Second, the 2005 NAEP 8^{th}-grade mathematic release items were used to validate the criteria of the Q-sort for the exclusiveness and inclusiveness of the mathematics' framework categorization. Third, the validation of the *exclusiveness* (Q-standard) of the MLAF was completed by a curricular team composed of the researcher and two curriculum specialists. Fourth, the validation of the *inclusiveness* (Q-sorts) of the MALF was completed by a team of mathematics teachers. Two teams (curricular and teacher) made up the validation panel for the MLAF used in this study. All members of the panel were fully credential with extensive multi-cultural urban teaching experiences. The curriculum team's experience in education ranges from 15 to 35 years, and the teacher team's experience ranges between 7 and 28 years.

To set the Q-standard, the curricular team (speech pathologist, special education K-12 content specialist, and the researcher) studied release items from the 2003 NAEP Math for 8th-grade to establish the characteristics of language differences between items across content strands and item formats. At this stage, the curricular team's primary function was to establish the categories and characteristic boundaries defined in the MLAF. The curricular team reviewed all release 2003 NAEP math items and focused on the question stems only as the essential component of mathematics assessment. Items on the assessment are designed based on IRT theory (NAGB, 2004). Item parameters include difficulty, discrimination, and pseudo-guessing. Items may be questions that have incorrect and correct responses, or statements that allow respondents to indicate level of agreement and are subject to multiple interpretations. The ability to understand and act on ambiguous information in the question stem requires both the understanding of the complexities of mathematics language and computational skills.

The ambiguity found in assessment word problems increases the range of possible interpretations between everyday language and the two types ambiguity found most commonly in mathematics: global and local. *Global ambiguity* means the whole sentence can have more than one interpretation. *Local ambiguity* means that part of a sentence can have more than one interpretation, but not the whole sentence (Cotton, 2000; Inman, 2005). This analysis focused on the local ambiguity found in the question stem of items on NAEP Math to determine the categories for the MALF and the placement of items within each category.

Graphic representations were used as the first identifying marker for differences between assessment items. Questions were then subdivided into the two categories (graphic or non-graphic) in order to examine each group separately. For each category, the team determined the intentionality of the question stem (what each student would need to do to solve the question). See Final Coding Categories in Methodology for more details on the MALF categories. The team determined that the question stems fell into two major groups (graphic and non-graphic representation) and a total of six subcategories (graphic vocabulary, non-graphic vocabulary, operate-to-plan, convert-to-solve, draw/manipulate, convert only).

The criterion of *inclusiveness* for the MLAF was measured by the middle school mathematics teachers using the Q-sort method with the 2003 release items. Five middle school teachers of mathematics independently classified release items into established categories. Each teacher met individually with the researcher at various school locations. The researcher reviewed the definitions and exemplars for the three subcategories in each of the two main categories with the teachers. The researcher used six release items to train the teachers by modeling how to use the definition of each subcategory with the examples to sort the items by question stem, according to the designated classifications. Teachers were then given an opportunity to ask clarifying questions. Teachers had headers on the table in order to sort and place the questions by category. All five teachers sorted questions into the two categories, graphics and non-graphics matching the sorting of the curricular team. All five teachers opted to sort graphic representation items first. A separate category was available for items that may not fit into any of the classifications. The teachers sorted the items into piles based on category definitions and examples. All items fit into one of the two classifications. Once the teachers completed the sort, they were asked by the researcher to review the work and confirm choices. Each teacher changed approximately 2-3 items from one subcategory to another within the graphic or non graphic category. No teacher changed an item from graphic to non-graphic or vice versa.

The placement of the item into a category was recorded by the researcher. After the completions of the final sort by the teachers, interrater agreement was calculated using percent agreement among raters. Overall, inter-rater agreement was 90%. Three quarters of the items had 100% inter-rater agreement, 25 % of the items had 66% agreement. After the preliminary classification of items occurred, the curricular team met to compare teacher classifications to the original Q-sort. Similar Q-sort procedures were used to classify the actual items from the 2005 NAEP Math.

A final field test of the MLAF was completed using 2005 release items. Two members of the teacher team and three members of the curricular team classified the 60 release items from the 2005 NAEP Math. This composition team followed the exact procedure described above that determined the inclusiveness of the Q-sort classification system. Each group classified items separately and then compared results. The researcher clarified questions and participated in discussion on reaching consensus for each of the release items. Using percent agreement, the first classification yielded a 94% inter-team reliability. The two groups discussed each item until 100% agreement. This 100%

agreement indicated that it was appropriate for the researcher to reclassify the secure items. The curricular team rank ordered the six MALF language subcategories by the degree of language complexity from most (1) to least complex (6) as follows: (1) graphic vocabulary, (2) non-graphic vocabulary, (3) operate and plan, (4), convert-to-solve, (5) draw/manipulate, (6) convert only. These rankings were supported by the research (Abedi and Lord, 2001, Blinko 2004; Fuchs et al, 2002; Tatsuoka, Carter, Tatsuoka 2004).

Final Coding Categories. The procedures described above resulted into six categories. The MALF categories met the Q-sort criteria for exclusiveness and inclusiveness as determined by the curricular panel assembled for this study. The six categories are described below.

Three distinct subcategories emerged for graphic representation, and three subcategories for non-graphic representation. The three graphic categories were: (1) draw/manipulate and solve; (2) organize and plan; and (3) relate vocabulary to recognition of formulas, graphs, or numerical expressions. The non-graphic categories were: (1) convert only; (2) convert-to-solve; and (3) relate vocabulary to definitions for formulas and or mathematical notation. The curricular team jointly (by consensus) synthesized the final set of items by language category classifications. Examples were selected for each of the classification categories of the released items.

Graphic representations are used as a visual aid to interpret mathematical data such as symbols, pictures, graphs, grids, charts, maps, geometric shapes, and numerical graphics that include number lines, computation items with less than three word directions, frequency tables, and extended numerical patterns. These items allow the students to examine and use the representation in order to support their understanding of the written part of the question on the assessment. Non-graphic representation referred to those items that have no visual or pictorial representation and items that rely on the understanding of the interrelationship of symbols-to-words and words-to-mathematical notation to solve. There are two types of problems presented in this group: words with numbers include items that have mathematical notation using the four operations (+, -, *, /), and words-only items that do not contain any numbers within the question stem. All of the items in non-graphic representation depend upon the understanding of written mathematical terminology to evaluate relevant information required to solve. The following categories are Non-graphic representation: non-graphic vocabulary, convert-to solve, and convert only. Figure 3 shows examples for each of the classification categories on the MALF taken from the 2003 NAEP release items.

This section provides details of each of the six categories on the MALF in the order of the most linguistically complex to least:

Graphic-Vocabulary only. This graphic category refers to items that require understanding of specific mathematical terms to identify or confirm mathematical notation, geometric shapes, location on a map or grid, or find discrete information on a graph or chart. This also includes mathematical formulas either as definitions or confirmation commonly used in geometry and measurement. The formulas can be depicted as words to picture or picture to words.

Non-graphic vocabulary only. The next category, non-graphic vocabulary, refers to items that require understanding of specific mathematical terminology in order to solve problems without pictorial representation. This category includes confirming mathematical definitions and formulas. In non-graphic vocabulary, these confirmations found as formulas can be depicted as key words in scenarios or examples.

Organize and plan. The third category includes graphic representation and refers to representational items that require knowledge and skills of syntax (word order), words (prepositional, proportional, multiple meaning), and directional signs to set up numerical expressions for computation. These are typically two-step word problems and require the student to use or validate the graphic representation included in the assessment.

Convert-to-solve. The fourth category, containing non-graphic items, usually includes two-step problems that require words or symbols to be changed into numerical notation prior to computation without graphic illustration. These problems, for example, ask students to translate between proportional relationships of decimal, percents and fractions; solve for unknowns, or apply a formula to solve hypothetical situation.

Draw or manipulate. This fifth category is graphic and refers to those items that require the use of additional resources not found within the stem of the question. These items may require the respondent to confirm an answer, extend, locate, or plot patterns, or apply new information to solve an equation. “Draw” in this study means to create a two dimensional shape based on directions in the item. “Manipulatives” are tools used during the NAEP mathematics assessment as part of the testing protocol to aid students in setting up or solving an item including paper rulers, protractors, markers, and grid paper to draw or extend graphic representations.

Convert-only. The sixth category is non-graphic and refers to those items that require the knowledge of technical vocabulary and/or mathematical notation to locate or identify the same information using symbols and words. This classification includes the automaticity of basic facts in addition, subtraction, and multiplication of whole numbers. There are no pictorial representations in these items i.e. (graphs, tables, charts or figures).

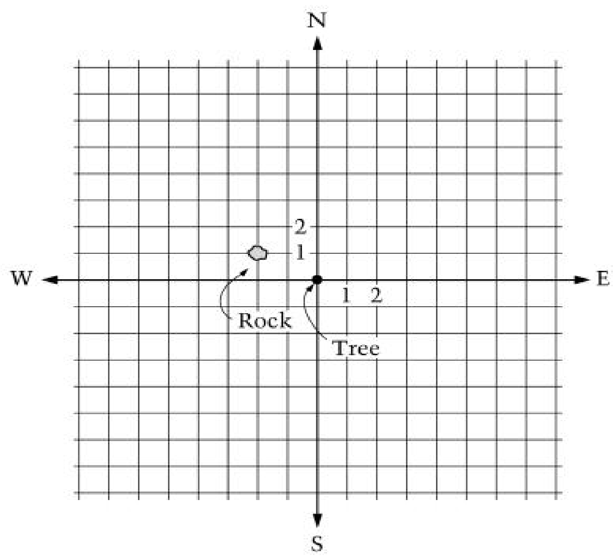
Exemplars for each of the six categories were selected from the 2003 NAEP 8th grade Mathematics Assessment. Each item was downloaded from the 2003 released questions from NAEP Questions Tool (http://nces.ed.gov/nationsreportcard/itmrls/) as part of the initial Q-sort. Theses examples were selected by the curricular team as best meeting the criteria for each category of the MALF. These items were used with the math teachers along with writing descriptors of each category. The category, Operate-and-Plan, had two examples: the first example contains both words and a figure, and the second example have symbol notation only with no words or directions. The examples for the Q-sort to establish the criteria of inclusiveness of the language categories on the MALF among the curricular panel are shown in Figure 3.

(1) Graphic Vocabulary
Description: Find location on a grid

On the map below, the rock is located 2 miles west and 1 mile north of the tree. A treasure chest (not shown) is located 8 miles east and 4 miles north of the rock. Mark the location of the treasure chest on the map with an X.

. What is the position of the treasure chest with respect to the tree?

Answer: _____ miles east and _____ miles north of the tree



(2) Non Graphic Vocabulary
Description: Identify property of a graph line

In a coordinate plane, the points (2,4) and (3,-1) are on a line. Which of the following must be true?

A) The line crosses the *x*-axis.
B) The line passes through (0,0).
C) The line stays above the *x*-axis at all times.
D) The line rises from the lower left to the upper right.
E) The line is parallel to the *y*-axis

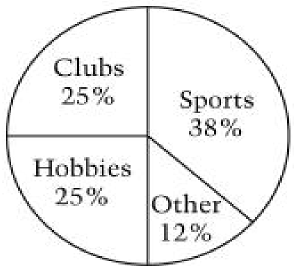
Figure 3

Exemplars for the Mathematics Assessments Language Framework Categories with Graphic and Non-Graphic Representation

(3) Organize and Plan (with figure)

Description: Use pie chart and percents to solve an equation

STUDENT PARTICIPATION IN ACTIVITIES AT ADAMS MIDDLE SCHOOL



1 There are 1,200 students enrolled in Adams Middle School. According to the graph above, how many of these students participate in sports?

A) 380
B) 456
C) 760
D) 820
E) 1,162

Organize and Plan (symbols only)

Description: Compute using order of operation

$3 + 15 \div 3 - 4 \times 2 =$

A) -9
B) -2
C) 0
D) 4
E) 5

(4) Convert-to-solve

Description: Counter-example of even –odd numbers

Consider the statement "If n is an even number, then n is two times an odd number." For which of the following values of n is the statement FALSE?

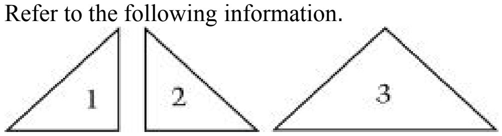
A) 2
B) 6
C) 8
D) 10
E) 14

Figure 3 (continued)

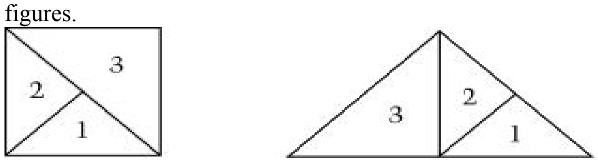
Exemplars for the Mathematics Assessments Language Framework Categories with Graphic and Non-Graphic Representation

<u>(5) Draw and manipulate</u>
Description: Form a parallelogram using shapes

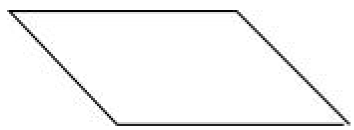
Refer to the following information.



Triangles 1, 2, and 3 shown above can be rearranged with no overlap to form either of the following figures.



. Draw lines on the figure below to show how triangles 1, 2, and 3 can be rearranged without overlap to form this parallelogram.



<u>(6) Convert only</u>
Description: Identify number of feet in 15 miles

1 mile = 5,280 feet

How many feet are in 15 miles?

A) 352
B) 35,200
C) 79,200
D) 84,480
E) 89,760

Figure 3 (continued)
Exemplars for the Mathematics Assessments Language Framework Categories with Graphic and Non-Graphic Representation

Sample and Sampling Procedures

The 2005 NAEP Math used a multistage stratification procedure to select sample schools and students for each state. Schools were stratified based on variables such as region of the country, urbanization, percentage of minority enrollment, and median house income based on state data reported to the Department of Education and Census Bureau. The primary sampling units (PSUs) were based on four geographic locations: Northeast, South, Central, and West. Schools were then randomly chosen within each state or PSU unless that school had special characteristics that made it unique within the state demographics (e.g., an urban school within a majority rural state).

On average, 100 public schools in each state were chosen with a total population of 2,500 students per grade, per subject assessed. States with large populations (i.e., California - 400 schools selected, 9,800 students) or with very small populations (i.e., North Dakota - 200 schools selected, 2,400 students) had more schools selected. Each selected school was intended to represent about 1% of the students in public schools nationally. The student participants were selected using a nesting procedure based on national, state, district, and school demographics. About 30 students were randomly chosen within each school as the representative sample for the assessment. A similar procedure was used to select participants for the non-public school sample (NCES, 2006).

Students in more than 6,500 schools participated in the 2005 mathematics assessment administered from January to early April 2005, yielding approximately 172,000 students at grade 4 and 162,000 students at grade 8. Each state's sample was designed to be proportionally representative of the different demographic characteristics of students in the state. Before the data were analyzed, responses from specific subgroups of students were assigned sampling weights to guarantee that the representation actually matched the percentage of the school population in the nation at the grade-level assessed. Sample weights were adjusted for oversampling or undersampling of a particular group of students based on national census data. Weighting was also used to adjust for missing data, within school variability and student non-responses (NCES, 2006). For each release item, mean scores were reported on a secure-licensed CD-ROM by content strand, problem type and item-level disaggregated by race/ethnicity, SES, and gender (NCES, 2006).

Unique to NAEP was the stratified multistage probability sampling design that included over-sampling of certain subpopulations at higher rates. For example, schools with high Black or Hispanic populations were sampled at twice the rate to obtain larger sample of respondents from these subpopulation in order to have a total sample that was representative of the nation based on Census Data (poststratification). Because of the differences in probability of selection rates due to oversampling and to allow for compensation for nonresponses, each student was assigned two sets sample weights: (1) reporting sample weights and (2) modular sample weights. Reporting samples are weights series of individual attributes that are assigned to each individual student and reflect the proportional representation of the characteristics and variety of individuals in the population. Modular weights are attributes that are assigned to individual students that allow for desegregations of groups of students from the total sample or for comparing results. A third procedure called “trimming” was conducted to equate reported and modular weights to equal the total population of the sample.

In NAEP, the complex sampling procedures affect the estimate of variance. The estimation of variability of the sample was affected by the weighting procedures used to compensate for oversampling. The estimation of variability or sample error is the condition of probability that a normal distribution of the population would fall within confidence interval of a range of values, defined by a lower and upper bound. Because not every student answered every question and some students within certain blocks of items either answered all questions correctly or below chance level, standard marginal maximum likelihood (MML) estimates used for missing values was not an inappropriate statistical procedure to arrive at mean score achievement for the NAEP's population proficiency distribution (Johnson, 1992; NAEP Primer, 1994; NCES, 2005). To compensate the fourth and final procedure (jackknife), NAEP calculates standard error measurements for the total population. The jackknife procedure uses 62 replicate weights to account for differences among students, and PSU's characteristics. (Allison, 2001; Arbaugh et al., 2003; Carlson et al., 2000; Lubienski, 2001; Lubienski & Lubienski, 2005; Lubienski & Shelley, 2003; NCES, 2006; Sowder, Wearne, Martin, & Strutchens (2004).

Missing data were handled through a predictive statistical procedure requiring multiple imputation that created five estimates of the overall score, called plausible values, to generate complete data sets with a value estimate for each question for every student taking the test (Allison, 2001; Carlson et al., 2000; Honaker & King, 2006; Horton & Kleinman, 2007). Plausible values were used to compute overall mean achievement for students in the 8th-grade, disaggregated mean achievement by race and ethnicity, and mean achievement by subsets of content strand and problem type (Carlson et al., 2001; NCES, 2005). The plausible values were combined to create a composite value for each content strand based on the distribution percentages established in the NAEP Mathematics Framework, and provided achievement data based on the performance range of the total of the actual students who answered the item on the assessment within a 500 point scale (NAGB, 2004; NCES, 2004).

Study Procedures

The study began when the researcher attended a NAEP Database Training Seminar in 2003 to become familiar with NAEP data structures and to obtain hands-on experience with NAEP-specific software. The researcher's qualifications to attend the training and conduct this study are itemized in Appendix A. The training seminar prepared the attendees to use the *NAEP Data Tool Kit* which was needed to access NAEP restricted-use data on the CD-ROM.

The *NAEP Data Tool Kit* contains two data analysis tools. One tool is NAEPEX, a data extraction program for choosing and customizing variables for extracting data into SPSS. The NAEPEX program allows for the identification and selection of data files that contain all responses, scores, weights, demographic variables, and derived variables from the sample. The second program, *AM* software, is a specialized tool used to estimate regression models through marginal maximum likelihood (MML) and to automatically provide appropriate standard errors.

The University of San Francisco obtained a secured-license agreement with National Center for Education Statistics (control number-030904785) in order to have access to restricted-use micro-level data contained on CD-ROM for the purpose of secondary analysis. In accordance with the licensing agreement, adequate security measures were put in place so that the data were secured from unauthorized disclosure, use, or modification. This entailed building a computer with special security features to run the data and the implementation of a protocol for use and storage of the CD. The details of the protocol are detailed in Appendix B (Computer Security Act of 1987, Public Law 100-235; E-Government Act of 2002, Title V, subtitle A; NAGB, 2005; NCES, 2006).

The release items commonly available to researchers and educators are only a small portion of the total items and do not represent complete coverage of the content, cognitive skills, and range of difficulty in the NAEP assessment for mathematics. Access to the total mathematic items from the 2005 NAEP assessment was to be provided for this study through two sources: (1) released items that were questions from the assessment and available to the public through the NAEP Questions Tool (http://nces.ed.gov/nations reportcard/itmrls/); and (2) a secure-licensed CD-ROM that had all of the items on the assessment.

Because the secure-licensed CD was not released in time for the completion of the current study, the researcher and one member of the curricular team went to the National Center of Education Statistics (NCES) in Washington, DC, to reclassify all of the 2005 NAEP Math items (secure and release) into language categories. With an NCES observer present at all times, the researcher and one member of the curricular team coded each item 2005 NAEP Math from the secure CD-ROM, in accordance with the MLAF designed for this study. The researcher and curricular team member independently rated each item and then discussed the rating of items until 100% inter-rater agreement was reached.

At the American Institute of Research (AIR) in Washington DC, the researcher used the NAEP software to download the 2005 NAEP Mathematics 8th-grade items, background variables, and assessment data into an SPSS spreadsheet. Each item was coded by content area (algebra, measurement, geometry, data analysis and algebra), and problem type (multiple choice, short-constructed response, extended response). The AM software extracted the data from SPSS spreadsheet following the protocol established by NCES licensing agreement. AM software was then used to recode items language categories and perform all statistical procedures.

Data Analysis Procedures

Three statistical procedures were used to analyze to determine the impact of mathematic language on achievement and examine if mathematical assessment language affected achievement gap differentials between Whites and minority students (Black, Hispanic, and Asian). First, a computation of achievement gaps on the NAEP 2005 8th-grade Mathematics Assessment was conducted by content strand. Second, a content analysis of items by content strand, problem type, and language complexity based on MALF language categories was conducted. Third, in order to relate the magnitude of problem type to language complexity, frequency tables were created, cross tabulations were analyzed and a Rank Correlation was performed. Analyses were discussed and organized according to research questions.

Research Questions

Research Question 1: What are the achievement gap differences between racial and ethnic groups (White, Black, Hispanic, Asian) on the 8th-grade 2005 NAEP

Mathematics Assessment by content strand (Number and Operation, Measurement, Geometry, Data Analysis, and Algebra)?

In order to compute achievement mean scores and analyze NAEP data for Research Question #1, the following factors were considered prior to data analysis. Test specifications that included the distribution of Research Question #1 used the average mean scores of the overall achievement for the 2005 NAEP Mathematics Assessment, and the five plausible values by content strand (Number and Operations; measurement, geometry, data analysis) to disaggregate achievement by race/ethnicity (White, Black, Hispanic, and Asian).

Research Question 2: How are the five strands characterized in terms of problem types and language complexity on the 2005 NAEP 8th Grade Mathematics Assessment?

The primary purpose of this study was to develop and use a mathematics framework to investigate the linguistic complexities of items on the NEAP mathematics assessment as a possible explanation of linguistic factors beyond English language proficiency that may contribute to mathematics achievement. The development of the Mathematics Language Frameworks (MALF) was created to examine language complexity using Q-sort methodology. Two content analyses were performed on the 2005 NAEP 8th grade Math items: first by problem type (multiple choice, short constructed response, and extend response); second by the MALF language complexity categories (graphic vocabulary, non-graphic vocabulary, operate and plan, convert-to-solve, draw/manipulate, convert-only).

Research Question 3: What is the magnitude of the relationships between the achievement gaps and the percentage of items of different problem types and different language complexity categories?

To analyze the data for Research Question #3, frequency tables were established for each category, which were reviewed for actual item distribution by category and problem type for the 2005 NAEP Math. Using SPSS, two crosstabulations were conducted to identify whether a relationship existed between achievement gap ranking and language complexity on each content strand. The first crosstabulation related achievement gap ranking to the number of items for each problem type and number of items for each language complexity code. The second crosstabulation related achievement gap ranking to a subset of language complexity codings representing just vocabulary. Spearman rank-order correlation coefficients were calculated.

Summary

This study examined mean score differences on the 2005 NAEP Math by content strand to find out if there are: (1) differences on the 2005 NAEP Math for achievement performance by race/ethnicity by content strand found in previous research (Arbaugh et al., 2004; Lubienski, 2001; Lubienski & Lubienski, 2005; Lubienski & Shelley, 2003; Schulz et al., 2005); (2) differences between problem type and language complexity; and (3) the relationship of problem type to language complexity identified in Mathematics Language-Assessment Framework (MLAF) across the content strands.

CHAPTER FOUR

RESULTS

The purpose of this study was to conduct an analysis of the 2005 National Assessment of Educational Progress (NAEP) 8th-grade mathematics assessment items' language complexity to examine if there is a relationship between language complexity and student achievement. Three statistical analyses were performed.

The first analysis identified the extent of achievement gaps by strand disaggregated by race (White, Black, Hispanic, and Asian). *T* tests were performed by strand to evaluate the mean differences between groups. The second analysis was a content analysis of items by problem type as assigned by the NAEP framework, and by language complexity on the Mathematics Language Assessment Framework created for this study. A descriptive analysis by strand was completed to characterize each strand in terms of problem type and language complexity. The third analysis was a rank-order correlation that examined the magnitude of the relationships between the achievement gaps and the percentage of items of different problem types and different language complexity categories. Correlation coefficients were calculated for achievement gap rank and problem types; achievement gap rank and language complexity categories; and problem types and language complexity categories. Results are discussed and organized according to research question.

Research Question 1: What are the achievement gap differences between racial and ethnic groups (White, Black, Hispanic, Asian) on the 8th-grade 2005 NAEP mathematics assessment by content strand (number and operation, measurement, geometry, data analysis, and algebra)?

Using AM Software and the NAEP plausible value scores, mean scores were generated for the overall scores and for each of the five content strands by race. *T* tests were completed to compare White achievement mean scores to three minority student group scores (Black, Hispanic, and Asian). Tables 8-13 present the mean scores, t-test scores, standard errors, and *p*-values for each comparison. Because of the large sample size, degrees of freedom were calculated as infinity (Brownlee, 1960; NCES, 2007). A summary of the performance benchmarks for each of the content strands is presented before analysis.

Overall scores. Results for the composite plausible values analysis indicated that the overall average mean score for the 2005 NAEP Mathematics Assessment for 8th-grade was 278. The White student group mean score was 289. The Black student group mean score of 253 was 36 points lower than the White student group mean score, the Hispanic student group mean score of 264 was 25 points lower, and the Asian student group mean score of 294 was 4 points higher than the White student group mean. Three *t*-tests were run to compare the significance of achievement mean differences between White and Black students, White and Hispanic students, and White and Asian students. The *t*-tests indicated that there were significance differences for race on composite mean scores between White students and Black, (t=13.30, $p < .01$), and between White students and Hispanic students, (t=21.08, $p < .01$), with White students receiving higher mean scores than Black or Hispanic students. There was no statistical significance between Asian students and White students. Table 8 gives the result of the composite achievement mean scores, achievement gap for the total measure (228) and for the white group score compared to minority groups, *t*-tests, standard errors, and p-values by groups.

Table 8

Weighted N's, Mean Achievement Scores, Achievement Gaps, and T-tests Disaggregated by Race for Overall Achievement

Race	Students (N)	Mean Score	Achievement Gaps Total	Achievement Gaps Whites	Std Error	*T*-value	*p*-value
White	114,149	289	11				
Black	43,075	253	-25	36	1.67	21.08**	.01
Hispanic	31,954	264	-14	25	1.87	13.30**	.01
Asian	8,690	294	16	5	3.44	-1.47	.16

*p < .05 ** p < .01

Number and Operations. Mastery of number sense is a major expectation of the 2005 NAEP Mathematics. This content area focuses on students' abilities to represent numbers, order numbers, compute with numbers, make estimates appropriate to given situations, use ratios and proportional reasoning, and apply number properties and operations to solve real-world and mathematical problems. By 8th-grade, students should be able to use rational numbers, represented either as decimal fractions (including percents) or as common fractions. They are expected to use them to solve problems involving proportionality and rates. In middle school, number begins to coalesce with geometry via the idea of the number line. It is recommended that the number line be connected with ideas of approximation and the use of scientific notation. It is also recommended that 8th-graders should be able to use occurring irrational numbers, such as square roots and pi (NAGB Framework, 2004).

Results of the achievement gap for Number and Operations analysis indicated that the strand average mean score was 278. The strand mean was the same as the overall test mean. When disaggregated by race/ethnicity, the White student group means score of 289 was 11 points higher than the average mean score for the strand. The Black student group mean of 254 was 24 points lower than the strand average and 35 points lower then the White student group mean. The Hispanic student group mean of 264 was 14 points lower than the group mean and 24 points lower than the White group mean. The Asian student group mean (292) was three points higher than White group.

Three *t*-tests were performed to compare the significance of achievement mean differences between White and Black students, White and Hispanic students, and White and Asian students. The *t*-tests indicated that there was a significance effect for race on Number and Operation content strand mean scores between the White students and Black, (t=18.98, p < .01 0); and between White students and Hispanic students, (t=10.73, p < .01), with White students receiving higher mean scores than Black or Hispanic students. There was no statistical significance found between Asian and White mean scores. Table 9 gives the result of the composite achievement mean scores, achievement gap for the total measure (277) and for the white group score compared to the minority groups, *t*-tests, standard errors, and p-values by groups.

Table 9

Weighted N's, Mean Achievement Scores, Achievement Gaps, and T-tests Disaggregated by Race for Achievement in Number and Operations

Race	Students (N)	Mean Score	Achievement Gaps		Std Error	*T*-value	*p*-value
			Total	Whites			
White	114,149	289	+12				
Black	43,075	254	-23	-34	1.80	18.98**	.01
Hispanic	31,954	264	-13	24	2.25	10.73**	.01
Asian	8,690	292	+15	3	4.08	8.02	0.4

*p < .05 ** p < .01

Measurement. This content area focuses on students' understanding of measurement attributes such as capacity, weight/mass, time, and temperature, the geometric attributes of length, area, and volume, attributes such as capacity, weight/mass, time, and temperature, as well as the geometric attributes of length, area, and volume. More emphasis is placed on area and angle in grade 8, nonstandard, customary, and metric units are assessed. Measurement at Grade 8 includes the use of both square and cubic units for measuring area, surface area, and volume; degrees for measuring angles; and constructed units such as miles per hour. Converting from one unit in a system to another (such as from minutes to hours) is an important aspect of measurement included in problem situations (NAGB Framework, 2004).

Results of the achievement gap for Measurement analysis indicated that the strand average mean score for the strand of Measurement was 273. When disaggregated by race/ethnicity, the White student group mean score of 288 was 15 points higher than the mean average score strand. The Black student group mean of 241 was 37 points lower the content strand mean and 47 points lower then the White student group mean. The Hispanic student group mean of 255 was 23 points lower then content strand mean and 33 point lower than the White group mean. The Asian student group mean of 293 was 15 points higher than the strand mean and five points higher than the White group mean.

Three *t*-tests were run to compare the significance of achievement mean differences between White students and Black students, White students and Hispanic students, and White and Asian students. The *t*-tests indicated that there was a significant effect for race on Measurement mean scores between the White students and Black students, (t= 16.78, p < .01); and between White students and Hispanic students, (t=9.88,

$p < .01$), with White students receiving higher mean scores than Black or Hispanic students. There was no statistical significance found between Asian and White mean scores. Table 10 gives the result of the composite achievement mean scores, achievement gap for the total measure (277) and for the white group score compared to the minority groups, *t*-tests, standard errors, and p-values by groups.

Table 10

Weighted N's, Mean Achievement Scores, Achievement Gaps, and T-tests Disaggregated by Race for Achievement in Content Strand Measurement (277)

Race	Students (N)	Mean Score	Achievement Gaps		Std Error	*T*-value	*p*-value
			Total	Whites			
White	114,149	288	+11				
Black	43,075	241	-32	-47	2.78	16.78**	.01
Hispanic	31,954	256	-18	-32	3.28	9.88**	.01
Asian	8,690	294	+21	+6	5.47	1.03	0.30

*p < .05 ** p < .01

Geometry. In middle school, students are expected to be familiar with simple figures and their attributes, both in the plane (lines, circles, triangles, rectangles, and squares) and in space (cubes, spheres, and cylinders) and use understanding of these shapes with cross-sections of solids. Students in 8th-grade are expected to demonstrate the beginnings of an analytical understanding of properties of plane figures, especially parallelism, perpendicularity, and angle relations in polygons. Right angles and the Pythagorean Theorem are introduced, and geometry becomes mixed with measurement. Students are expected to be familiar with the basic types of symmetry transformations of plane figures, including flips (reflection across lines), turns (rotations around points), and slides (translations), with each type of transformation being distinguished from other types by their qualitative effects. The basis for analytic geometry is laid by study of the number line (NAGB Framework, 2004).

Results of the achievement gap for Geometry analysis indicated that the strand average mean score was 275. When disaggregated by race/ethnicity, the White student group mean score of 285 was 10 points higher than the average mean score for the strand. The Black student group mean of 252 was 23 points lower then the strand mean, and 33 points lower than the White student group mean. Hispanic student group mean of 264 was 11 points lower than the test mean, and 21 points lower then the White group mean. The Asian student group mean of 292 was 7 points higher than White group.

Three *t*-tests were run to compare the significance of achievement mean differences between White students and Black students, White students and Hispanic students, and White students and Asian students. The *t*-tests indicated that there was a significant effect for race on Geometry content strand mean scores between the White students and Black, (t=17.84, $p < .01$); and between White students and Hispanic students, (t=11.01, $p < .01$), with White students receiving higher mean scores than Black or Hispanic students. There was no statistical significance between the mean scores of Asian students and White students. Table 11 gives the result of the composite achievement mean scores, achievement gap for the total measure (275) and for the white group score compared to the minority groups, *t*-tests, standard errors, and p-values by groups.

Table 11

Weighted N's, Mean Achievement Scores, Achievement Gaps, and T-tests Disaggregated by Race for Achievement in Geometry (275)

Race	Students (N)	Mean Score	Achievement Gaps		Std Error	*T*-value	*p*-value
			Total	Whites			
White	114,149	285	+10				
Black	43,075	252	-23	-33	1.84	17.84**	.01
Hispanic	31,954	264	-11	-21	1.90	11.01**	.01
Asian	8,690	292	+17	+7	3.51	2.59	0.4

*p < .05 ** p < .01

Data analysis. Data analysis covers the entire process of collecting, organizing, summarizing, and interpreting data. By the end of grade 8, students should be expected to apply their understanding of number and quantity to pose questions that can be answered by collecting appropriate data. They should be expected to organize data in a table or a plot and summarize the essential features of center, spread, and shape both verbally and written. Using summary statistics, students are expected to analyze statistical claims through designed surveys and experiments that involve randomization for making simple statistical inferences. Students are to begin to use more formal terminology related to probability and data analysis (NAGB Framework, 2004).

Results of the achievement gap for Data Analysis strand indicated that the average mean score was 281. The strand mean was three points higher than the overall test mean of 278. When disaggregated by race/ethnicity, the White students group mean score of 295 was 14 points higher than the average mean score. The Black student group mean of 255 was 40 points below the White student group mean. The Hispanic student group mean of 266 was 29 points lower than the White group mean. The Asian student group mean of 293 was two points lower than White group mean.

Three *t*-tests were run to compare the significance of achievement mean differences between White students and Black students, White students and Hispanic students, and White students and Asian students. The *t*-tests indicated that there was a significant effect for race on data and analysis content strand mean scores between the White students and Black, (t=17.07, p < .01); and between White students and Hispanic students, (t=11.18, p < .01), with White students receiving higher mean scores than Black or Hispanic students. There was no statistical significance found between the mean scores of Asian students and White. Table12 gives the result of the composite achievement mean scores, achievement gap for the total measure (281) and for the white group score compared to the minority groups, *t*-tests, standard errors, and p-values by groups.

Table 12

Weighted N's, Mean Achievement Scores, Achievement Gaps, and T-tests Disaggregated by Race for Achievement in Data Analysis (281)

Race	Students (N)	Mean Score	Achievement Gaps		Std Error	*T*-value	*p*-value
			Total	Whites			
White	114,149	295	+14				
Black	43,075	255	-26	-40	2.31	17.07**	.01
Hispanic	31,954	266	-15	-28	2.55	11.18**	.01
Asian	8,690	293	+12	-2	5.06	.24	0.4

*p < .05 ** p < .01

Algebra. By 8th-grade, understanding of functions and variables become more important and students are expected to have comprehensive background for the mathematics concept of function and set theory. Representation of functions as patterns, via tables, verbal descriptions, symbolic descriptions, and graphs are expected to be used as part of linear functions. Students are expected to connect the ideas of proportionality and rate including graphing by hand or with calculators and use linear equations to find solutions (NAGB Framework, 2004).

Results of the achievement gap for Algebra analysis indicated the average mean score 280. When the strand was disaggregated by race/ethnicity, the White student group mean score of 290 was 12 points higher than the average mean strand. The Black student group mean of 258 was 22 points lower than the strand mean, 32 points lower the White student group mean. The Hispanic student group mean of 268 was 12 points than the strand mean, and 22 point lower than the White group mean. The Asian group mean, 290, was two points higher than the White group mean.

Three *t*-tests were performed to compare the significance of achievement mean differences between White students and Black students, White students and Hispanic students, and White students and Asian students. The *t*-tests indicated that there was a significant effect for race on Algebra mean scores between the White students and Black, (t=17.90, $p < .01$); and between White students and Hispanic students, (t=9.88, $p < .01$), with White students receiving higher mean scores than Black or Hispanic students. There was no statistical significance was found between the mean scores of Asian students and White students. Table 13 gives the result of the composite achievement mean scores, achievement gap for the total measure (280) and for the white group score compared to the minority groups, *t*-tests, standard errors, and p-values by groups.

Table 13

Weighted N's, Mean Achievement Scores, Achievement Gaps, and T-tests Disaggregated by Race for Achievement in Algebra (280)

Race	Students (N)	Mean Score	Achievement Gaps Total	Achievement Gaps Whites	Std Error	*T*-value	*p*-value
White	114,149	290	+11				
Black	43,075	258	-22	-32	1.77	-17.90**	.01
Hispanic	31,954	267	-13	-23	2.30	-9.88**	.01
Asian	8,690	292	+12	+2	2.43	2.01*	0.4

*p < .05 ** p < .01

Summary. The achievement gap results on the 2005 8th grade NAEP Mathematics trends were consistent across the five content strands. Asian students outperformed all other groups on all strands except Data Analysis; White students outperformed Black and Hispanic students; and Hispanic students outperformed Black students. For all five of the content strands the achievement gap differences between White students and Black students, and between White students and Hispanic students were statistically significant. None of the gaps between White student and Asian students were statistically significant. Measurement was ranked the most difficult of the five strands based on overall mean achievement scores. This strand showed the largest gap for both Black and Hispanic students when compared with their White peers. Data Analysis was ranked the least difficult of the five content strands based on the overall mean achievement scores but it showed the second highest gap for both Black students and Hispanic students when compared to their White peers. Geometry ranked second in difficulty, but this strand had the smallest gaps between White students and their non-Asian peers and the largest gap between White students and Asian students. Table 14 presents a summary of the achievement gap data by content strand with the overall means, race/ethnic group means, and achievement gaps by strand mean, the achievement mean scores, and achievement gaps disaggregated by race, and difficulty ranking by strand mean (1=largest gap, 5 = smallest gap).

Table 14

Content Strand Achievement Gaps

Content Strand	Overall Mean	Means and Ranks						
		Whites	Blacks		Hispanics		Asian	
		Mean	Mean	Rank	Mean	Rank	Mean	Rank
Measurement	273	288	241 (-47)	1	256 (-32)	1	294 (+6)	2
Geometry	274	285	252 (-33)	5	264 (-21)	5	292 (+7)	1
Number and Operation	278	289	254 (-34)	4	264 (-23)	3	292 (+3)	3
Algebra	280	290	258 (-38)	3	267 (-23)	3	292 (+2)	4
Data Analysis	281	295	255 (-40)	2	266 (-29)	2	293 (-3)	3

Clearly, wide differences exist between among ethnic groups in 8th grade NAEP mathematics achievement. The question is whether these differences in achievement can be explained as a result of language complexity. On which strands are the achievement gaps the largest, and do these have more complex language and more difficult problem type? For example, is it possible that the Data Analysis achievement gap could be due to difficulty with understanding the language used in the items? This strand had the highest mean score, but the second highest achievement gap. The second research question examines these issues of language complexity using the MALF framework constructed for this dissertation.

Research Question 2: How are the five strands characterized in terms of problem types and language complexity?

The primary purpose of this study was to develop and use a mathematics framework to code assessment items according to linguistic factors that may contribute to mathematics achievement on the NEAP mathematics assessment. The first step was to identify the distribution of 2005 NAEP mathematics items by the problem types found within each of the content strands. The second step was to examine the distribution of these items when reclassified in accordance to the language categories. The third and final step was to examine the distribution of language categories by problem type across strands.

Problem Type. An analysis of the items on the 2005 NAEP Math was conducted within content strand and across problem type to determine if there were relationships between problem type and achievement gap among the content strands. Three problem types, multiple choice, constructed response and extended response, are used to define the level of mathematical ability of students on the 2005 NAEP Math across the five strands. The first item format is standard multiple-choice with each item having five choices. The second item format is short constructed response (SCR) and had two variations. The third item format is the extended construction response (ECR) questions. Items were placed in one of the three problem types by strand based on the NAEP coding contained in AM software. For each of the five content strands the numbers of items by problem type and percentages were calculated. For all five strands, multiple choice items represented the largest problem type, although the actual number of items varied by strand. Short constructed response represents the second largest percentage of items followed by extended constructed response, with the exception of data analysis. Table 15 summarizes the number of items by problem type for the five content strands.

Table 15

Number and Percentage of Items by Problem Type for the Five Content Strands

Content Strand	Problem Type			
	Multiple Choice	Constructed Response	Extended Response*	Total
Measurement	20 (71%)	6 (21%)	2 (7%)	28
Geometry	23 (61%)	11 (29%)	4 (11%)	38
Number and Operations	36 (69%)	9 (25%)	3 (6%)	48
Algebra	29 (66%)	10 (23%)	5 (11%)	44
Data Analysis	20 (69%)	1 (3%)	8 (28%)	29

Mathematics Language Assessment Framework Categories. The initial coding for the Mathematics Language Assessment Framework (MALF) organized 2005 NAEP Math items first by graphic representations or non-graphic representations, then into one of six language categories based on the questions stem across the five content strands. The six categories were ranked order by language complexity from highest to lowest as follows: (1) graphic vocabulary, (2) non-graphic vocabulary, (3) operate and plan, (4) convert-to-solve, (5) draw and manipulate, (6) convert only. After the initial sort identified and categorized each item according the language framework, items were then reorganized by strand. A comparative analysis of problem type x language category was conducted for items contained in each strand. A final analysis compared the content strands by mean scores, achievement gaps, and the percentage of total vocabulary items based on the MLAF categories.

According to the overall item distribution on the NAEP Math using the MLAF, 112 items (60%) were classified as graphic representation and 75 items (40%) were classified as non-graphic representation. Two of the sub-categories in graphic representation, graphic vocabulary (24%) and operate-and-plan (20%), equaled 44% of the total assessment. Non-graphic vocabulary represented 48% of the total items on 2005 NAEP Math, and 19% of the total reclassified items according Mathematics Assessment Language Framework. Table 16 summarizes the types of language complexity found in each strand by number and percentage of items for the five content strands.

Table 16

Number and Percentage of Items by Language Complexity Categories for the Five Content Strands

	Language Complexity						
Strand	GV	NV	OP	CS	DM	CO	Total
Measurement	5 (18%)	8 (29%)	3 (11%)	2 (7%)	9 (32%)	1 (3%)	28
Geometry	11 (29%)	6 (16%)	5 (13%)	2 (5%)	14 (37%)	0	38
Numbers and Operations	9 (19%)	6 (13%)	10 (21%)	14 (29%)	3 (6%)	6 (13%)	48
Algebra	9 (20%)	11 (25%)	13 (30%)	3 (7%)	2 (5%)	6 (14%)	44
Data	10 (34%)	5 (17%)	7 (24%)	4 (17%)	2 (7%)	1 (5%)	29

Note: GV = graphic vocabulary, NG = non graphic vocabulary, OP = operate plan, CS = convert to solve, DM = draw and manipulate, and CO = convert only.

Problem Type x Language Category. An analysis was performed to compare items on the NAEP Math problem type and language category by content strand. Table 17 provides a summary of the results comparing strand mean scores, achievement gap ranking, percentage of problem types, and percentage of vocabulary items. Tables 1-5 in Appendix C presents the breakdown of items on the NAEP Math by each content strand, problem type and language category. Language categories have been ranked from most to least language complex as follows: graphic vocabulary (GV), non graphic vocabulary (NG), operate plan (OP), convert to solve (CS), draw and manipulate (DM), and convert only (CO).

Measurement. This strand had the lowest overall mean score and the highest achievement gaps between Whites and Blacks, and Whites and Hispanics. It had the second highest gap between Asian students and Whites students. Measurement had the largest representation of content multiple choice items than any strand 71% to the total strand (20 /28items). The two vocabulary categories represented 13 items, 46% of the total strand. Graphic vocabulary had five items (18%), and non-graphic vocabulary had 8 items (28%). This suggests that language complexity may have effected the mean achievement of this strand. Appendix C - Table 1 shows measurement items by language categories and problem type.

Geometry. This strand had lowest achievement gaps between Whites and Blacks, and Whites and Hispanics. It had the highest gap between Asian students and Whites students. The majority of problems for this strand was content multiple choice, representing 61% of the strand (23/ 38 items). The two vocabulary categories represented 17 items, 45% of the total strand. Graphic vocabulary had eleven items (29%), non-graphic vocabulary had 6 items (16%).This indicates that language complexity may have effected the mean achievement of this strand. Appendix C -Table 2 shows geometry items by language categories and problem type.

Number and Operations. This strand followed the trend of the achievement gaps, as White students outperformed Blacks students and Hispanics students, and Asian students out performed White students. Number and Operation had a substantial percentage of multiple choice problems representing 69% or 33/48 of total items within the strand. The two vocabulary categories were the least represented in the five content strands with a total of 15 items (31%). Graphic vocabulary had nine items (19%); non-graphic vocabulary had 6 items (13%).The results suggest that language complexity may have less of an effect on the mean achievement in Number and Operations than in other content strands. Appendix C – Table 3 shows number and operations items by language categories and problem type.

Algebra. This strand had the third highest achievement gaps between Whites and Blacks, and Whites and Hispanics. It had the smallest gap between Asian students and Whites students. Algebra had the least amount of multiple choice questions relative to the total strand (36%) or 16 out of 44 items. The two vocabulary categories represented 20 items, 45% of the total strand. Graphic vocabulary had 9 items (20%), non-graphic vocabulary had 11 items (25%).These results indicates that language complexity may have affected the mean achievement of this strand. Appendix C –Table 4 shows algebra items by language categories and problem type.

Data Analysis. This strand had highest achievement overall and the second highest gaps between Whites and Blacks, and Whites and Hispanics. It is the only strand in which Asian students did not outperform Whites students. Data analysis items types were 69% (20/29) multiple choice and the largest representation of extended response relative to the strand of 28%. Because of the large extended response questions, partial credits may have had the largest effect on the strand mean achievement. For example one item had 6% of responses fully correct, when partial credits were calculated for percent correct the item's p-value increased to 49%. For the purposes of this study however, the researcher focused on total correct response. This strand had the greatest items categorized as vocabulary with 52% of the total strand classified as graphic or non-graphic vocabulary (15/29). Graphic vocabulary had ten items (34%); non-graphic vocabulary had 6 items (16%).These findings suggest that language complexity may have affected the mean achievement of this strand more than any of the other strands. Appendix C – Table 5 shows data analysis items by language categories and problem type. Table 17 presents a summary of 2005 NAEP 8th Grade Mathematics by achievement gap rankings with the percentages of problem types and language complexity categories by the five content strands.

Table 17

Achievement Gap Rankings, Percentage of Items in Each Problem Type, and Percentage of Items in Each Language Categories for the Five Content Strands

Strand	Gap Rank	Problem Type (percentages)			Language Complexity (percentages)					
		MC	CR	ER	GV	NV	OP	CS	DM	CO
Meas	1	71	21	7	18	29	11	7	32	3
Geom	5	61	29	11	29	16	13	5	37	0
Numb	4	69	25	6	19	13	21	29	6	13
Alg	3	66	23	11	20	25	30	7	5	14
D-A	2	69	3	28	34	17	24	17	7	5

Vocabulary is the most complex of the linguistic categories (Cotton 2000, Iman, 2005). Similarly, the combined vocabulary categories (graphic and non-graphic) were also considered the most complex on the MALF according the validity panel. To identify the impact that language complexity has on the overall assessment, vocabulary items were isolated. The results show that the overall assessment had a total of 80 items or 43% classified in two vocabulary categories: graphic or non-graphic. Results for vocabulary language complexity by strands ranged from a high of 52% of the total in Data and Analysis to a low of 33% in Number and Operations. Measurement and Geometry both had 45% of the total items classified as vocabulary. Algebra had 46% of the items. Table 18 presents a summary of mean achievement, achievement gap ranking, and total percentage of vocabulary items (graphic and non graphic) based on the MALF classifications.

Table 18

Content Strand Mean Scores, Achievement Gap Ranking, and Percentage of Items in the Vocabulary Category

Content Strand	Mean Score	Achievement Gap Ranking	Percentage Vocabulary
Measurement	273	1	45
Geometry	275	5	45
Number and Operation	278	4	33
Algebra	280	3	46
Data analysis	281	2	52

Summary. The analysis for this research question determined that problem types were consistent across all strand except data analysis. Language complexity factors were represented across all strands, with vocabulary showing the highest representation._What effect, if any, do problem type and language complexity, more specifically vocabulary, have on the achievement gaps by content strand? A Rank Correlation was calculated to analyze relationships of achievement gap ranks, problem type and language complexity.

Research Question 3: What is the magnitude of the relationships between the achievement gaps and the percentage of items of different problem types and different language complexity categories?

Spearman rank-order coefficients were calculated between the achievement gap ranking and the percentages of items in each problem type and each language complexity category (see Table 17). Obviously having only 5 content strands is too small to draw conclusions about relationships. However, the correlations maybe suggestive of possible relationships between achievement gaps, problem types, and language categories on the 2005 8th grade NAEP Math.

The findings for the relationships between achievement gap ranks by strand and item type or category suggested there was a strong negative correlation between achievement gap rank and multiple choice items (.81); a moderate positive correlation between achievement gap rank and short constructed response items (.59); and a small positive correlation between achievement gap rank and extended constructed response items (.29). In addition, there was a strong correlation between achievement gap ranks and non-graphic vocabulary items (.70), which is notable because non-graphic vocabulary is considered one of the most linguistically complex of the MALF categories.

The findings for the relationships between language category and problem type suggested a strong positive correlation between graphic vocabulary and extended constructed response item (.97); and strong negative relationship between graphic vocabulary and short constructed response items (.89). There were small statistically insignificant correlations that existed between the other language categories (non-graphic vocabulary, operate-and-plan, convert-to-solve, draw/manipulate and convert only) and the three problem types.

Summary. This current study introduces the new variable of language complexity which was used to begin to investigate language fluency by examining the 2005 NAEP Math using the MALF framework developed to study language complexity on assessments. A descriptive analysis of the 2005 NAEP mathematics for problem type on achievement content strand was completed. First, items from the NAEP Math were examined by content strand for achievement, and achievement gap differences disaggregated by race. Second, items were reclassified into six categories by language complexity using the MALF. A descriptive analysis was completed for language complexity by strand. Items were found to be organized into two primary language areas: vocabulary (43%) that encompassed the graphic and non-graphic vocabulary categories on the MALF, and multi-step problems (50%) encompassed operate-and-plan, convert-to-solve, and draw/manipulate categories; with a small subset of single-step items of convert-only items (7%). Finally, a correlation coefficient was calculated to analyze the relationships among achievement gaps, problem type, and language complexity categories.

The analysis of Research Question 1 substantiated the differences in achievement by racial groups (White, Black, Hispanic, and Asian) on the 2005 8th-grade NAEP Mathematics Assessment. The analysis of Research Question 2 suggested that language complexity as categorized by Mathematics Assessment Language Framework can be isolated as a separate component of mathematics. The analysis of Research Question 3 suggested that there may be relationships among achievement gap ranks, problem types, and language complexity categories. These results will be discussed in Chapter Five.

CHAPTER FIVE

SUMMARY, LIMITATIONS, DISCUSSIONS, AND IMPLICATIONS

The purpose of this study was to conduct a descriptive analysis the 2005 NAEP 8th-grade mathematics assessment. In order to determine if a relationship between mathematical language fluency and mathematics achievement exists, the Mathematics Assessment Language Framework was created to classify the 2005 8th-grade NAEP mathematics assessment test items according to problem type and language complexity. The magnitude of the achievement gap on each content strand was then related to the percentage of items classified by problem type and language complexity.

Even though the influence of language and language factors in mathematics has become more important in our pluralistic society, research on such factors has not kept up with their importance. Since Coleman (1966) first disaggregated mathematics achievement scores by race, gender, and SES, achievement gaps have been observed between minority students (Black and Hispanic) and Whites, demonstrating that many minority students are not developing the mathematical skills necessary to compete effectively in the workforce. Because mathematics has a specialized vocabulary which must be learned and depends on the ability to integrate words, symbols, and vocabulary to create and communicate ideas, it is possible that linguistic features of mathematics may play a role in mathematics achievement.

The effect of mathematical language fluency on mathematical achievement is not well studied (RAND Mathematics Study Panel, 2003). Language fluency in mathematics refers to the ability of a student to understand what is required in a mathematics test item and delineates the differences between language used on a daily basis and the language associated with problem solving. One aspect of mathematical fluency is the use of the complex language that is required to represent abstract structures and relationships using words, mathematical notation, symbols, and logic, which in mathematics is more careful and accurate than everyday speech (Nowak, Komarova, & Niyogi, 2002; Wakefield, 2000). Investigations are needed to determine if language factors are one reason why certain groups consistently lack the necessary skill sets to reach proficiency on mathematic assessments measured by disaggregated state and national achievement performance.

A number of epistemological frameworks have been developed to examine and characterizes language features on mathematic items(Aiken, 1970, 1971; Ciani, 1981; Cronbach, 1942; Halliday, 1985; Levine & Reed, 2001, Mayer, 1982; Pachtman & Riley, 1978; Wakefield, 2000). This study focused on linguistic complexities or features as these are seen as the prerequisites for the attainment of mathematical fluency.

Primary to this study was Wakefield's (2000) epistemological framework that characterizes mathematics as a separate language. Wakefield's framework defines the foundational social-linguistic constructs of mathematical language that identify the interdependence of words, symbol, and expressions used to construct meaning and communicate ideas. Though the acquisition of mathematical literacy draws on many of the same skills as print literacy, there are important characteristics of mathematics that qualify it as a language and, thus, require different strategies to engage students (Adams, 2004; Wakefield, 2000). Wakefield (2000) suggested that the understanding of mathematical concepts depends upon the students' fluency, proficiency, and comprehension of mathematics vocabulary.

Research has identified that mathematical language fluency may affect student performance on word problems, but little research has been done to examine language complexities contained on assessment items used to judge student achievement (Abedi & Lord, 2001; Heagerty, Mayer, & Monk, 1995; Pape, 2004; Quilici, & Mayer, 1996). Previous research related to NAEP examined if aspects of the content strands (number sense, measurement, data analysis, geometry, and algebra) or individual test items affected achievement performance (Lubienski & Shelly, 2003; Schulz et al., 2005).

Lubienski and Shelly (2003) examined the NAEP 2000 mathematics data by content strand and suggested that there was a relationship between the content area domain and the achievement gap between White students and their minority peers (Black and Hispanic). The Schulz et al. (2005) reclassification of items from the NAEP 2003 mathematics assessment suggested that the assessment contains subsets of questions that represent criterion-referenced mastery anchor points and are related to achievement performance levels. The current study proposed that there are embedded language factors within the assessment that impact mathematics achievement. In order to examine if the relationship between mathematical language complexity and mathematics achievement exists, items were categorized into the Mathematics Assessment Language Framework (MALF) in order to reclassify the NAEP 2005 test items. Then analyses of the assessment were conducted to compare the five content strands for differences in: (1) mean scores, (2) achievement gaps, (3) problem types and (4) language complexity.

Finally, a correlation coefficient was calculated to identify the relationships among achievement gap ranks, problem types and language complexity categories.

The Mathematic Assessment-Language Framework (MALF) created for this study was designed around the multistage stratification procedures of the NAEP assessment in order to establish new language category boundaries to examine achievement data from the 2005 NAEP Math for 8th-grade. Achievement was reported on the NAEP Mathematics Assessment by subpopulations through the use of two student weighting formulas and a post-stratification jackknifing procedure to provide likelihood estimates of mean achievement (plausible values) for subpopulations (Allison, 2001; Carlson et al., 2000; Honaker & King, 2006; Horton & Kleinman, 2007; Johnson, 1992; NAEP Primer, 1994).

The Q-sort (Stephenson, 1935) was used to: (1) define language categories boundaries and the unique characteristics of each category; and (2) establish objectivity of the framework through the classification of 8th-grade NAEP release items from the 2003 and 2005 NAEP mathematics website by language complexity categories. Secure test items from the 2005 NAEP mathematics were classified according to the MALF, and new categories were created according to this classification: (1) graphic representation – draw/manipulate, organize and plan, vocabulary; and (2) non-graphic representation – convert only, convert-to-solve, and vocabulary).

Three procedures were used to analyze the research data. First, a comparative data analysis disaggregated racial/ethnic group data and compared mean scores by five mathematic content strands (number and operations; measurement; geometry; data analysis, and algebra) to examine differences in achievement on the 2005 NAEP Math. A series of *t*-tests were performed to compare White student group mean performance to group mean performance of Black students, Hispanic students, Asian students, and to assess whether the means of each two group comparison were *statistically* different from each other (White/Black; White/Hispanic/; and White/Asian). Second, a content analysis of the items was completed first by problem type (multiple choice, constructed response, and extended constructed response) and by MALF categories: (1) graphic representation – draw/manipulate; operate and plan, and vocabulary, and (2) non-graphic vocabulary – convert only, convert-to-solve, and vocabulary within each strand. Third, the magnitudes of the achievement gaps within each strand were related to the percentage of items classified according to problem types and language complexity and a rank order correlation was computed.

These procedures were conducted to answer the following research questions:

Research Questions

(1) What are the achievement gaps by race for each of the five strands (number and operations, measurement, geometry, data analysis, and algebra) on the 8th-grade 2005 NAEP Mathematics Assessment?

(2) How are the five strands characterized in terms of problem types and language complexity?

(3) What is the magnitude of the relationships between the achievement gaps and the percentage of items of different problem types and different language complexity categories?

Major Findings

There were eight major findings from the results of the data analysis:

(1) There were differences in mean achievement between the five content strands. Data analysis (Mean = 281) had the highest mean achievement followed by algebra (Mean = 280), number and operations (Mean = 277), geometry (Mean = 275), and measurement (Mean = 273). The difference in mean achievement between the highest strand (data analysis) and lowest strand (measurement) was significant at the p <.01 level.

(2) Within achievement gap differences, when disaggregated by race/ethnicity, had the same overall pattern among White students, Black students and Hispanic students. White students outperformed Black students and Hispanic students on all five content strands at the p .<01 level. Asian students outperformed White students on mean achievement on every strand except Data Analysis. None of the differences between Asian students and White students were statistically significant.

(3) Within strands, mean achievement gaps varied in size. For example, measurement had the lowest overall mean achievement and the largest gap between White students and Black students (46.8) and White students and Hispanic students (23.5). Data analysis (281) had the highest mean achievement and the second largest achievement gap between White students and Black students (39.6) and White students and Hispanic students (28.3).

(4) Mathematics Language Assessment Framework categories were represented across five strands (number and operations, measurement, geometry, data analysis, and

algebra) and three problem types (multiple choice, constructed response and extended response) on the 2005 NAEP Mathematics Assessment

(5) Linguistic complexity as a component of mathematical fluency can be classified and categorized on the item level using the MALF for the 2005 NAEP Mathematics Assessment.

(6) Achievement gap differences in the content strand of numbers and operations could be attributed more to computational competency than language complexity based on the analysis of problem types x language categories.

(7) Achievement gap differences in the content strand of data analysis could be attributed more to language complexity than computational competency based on the analysis of problem types x language categories.

(8) Strong relationships exist between achievement gap rank and problem type; achievement gap rank and language complexity category; and problem type and language complexity category.

Limitations

There were four limitations to this study.

First, overall limitations in using NAEP Assessments are that they are cross-sectional, not longitudinal. The assessments represent a single point in time which cannot reflect achievement over time for individuals or groups of students. The design of the data collection procedures can not be generalized to other assessment measures that report large-scale achievement results. Because all states create their own individual achievement benchmarks for 8th-grade mathematics, findings on achievement performance on the 2005 NAEP can not be generalized to state performance measures by content strand.

Second, the need for specific statistical weighting adjustments of the sampling data requires the use of specific software and limits the type of procedures that can be run. Specific to this study, due to fact that the secure-license CD for the NAEP 2005 Mathematics data was not available for dissemination, the researcher was allowed only two days in Washington DC to code items and collect data and perform the statistical procedures used in this study. Further support was needed to run the reclassified data using the plausible values to insure the validity of the statistical procedures that were conducted through telephone conversations.

Third, it is difficult to create variables that are not pre-existing when using national data sets. The use of item statistics for categorization on the Mathematic Assessment-Language Framework does not align language categories across strands and problem types. The weighting formulas used to determine plausible values affected the cross-comparisons of items. Item means by language category to generate statistical comparisons could not be generated within the limited time the researcher had access to the data.

Fourth, the limited number of mathematical content strands (five) affected the statistical significance of the correlation coefficients generated among achievement gap rank, problem type, and language complexity category. These correlations are suggestive only.

Discussion

For many years educators and policy-makers have been concerned about the continued achievement gap between White students and their minority peers (Black and Hispanic students). Much of the research around achievement gap issues has been devoted to socio-economics (Common Core Data, 2003; Kober, 2001; Lubienski, 2001, 2002), student motivation (Greenberg, Skidmore, & Rhodes, 2004; Turner 2004; Singh, Granville, & Dika 2002; Stodlolsky, 1985), cultural differences (Boaler, 2002; Ladson-Billings,1997; Lee, 2004; Lubienski, & Shelley 2003; Okpala, Okpala, & Smith, 2001; Tate, 1997), and the integration of English–language learners in core classes (Abedi, 2000; Abedi, Lord, & Hofsetter, 2004; Gutierrez, 2002; Hofstetter, 2003).

Current educational policy has focused on content standards and accountability measures that emphasize teacher-quality and student achievement benchmarks based on federal guidelines (Cohen, & Hill, 2002; Coley, 2003; Erpenbach, Forte-Fast, & Potts, 2003; NCLB, 2001; Resnick, 1987; Volger, 2002). Disparities in mathematics among subpopulations of students evident since kindergarten continue to exist and are exacerbated as students move through the grades (Colman Report, 1966; NCES, 2006; NSF, 2007).

Few studies have looked at the content contained within mathematics to examine if there may be differences that may influence achievement measures (Abedi & Lord, 2003; Bailey, & Butler, 2003; Lubienski, 2003). This study examined the 2005 NAEP 8th-grade Mathematics Assessment to analyze the achievement gap by race and ethnicity in two ways: (1) to consider if differences among the content strands (numbers & operations, measurement, geometry, data analysis, and algebra) varied by subpopulations;

and (2) to investigate if mathematic language complexity impacted student performance by developing and testing a mathematics assessment-language framework.

Findings show that significant differences in achievement occurred between the content strands but the overall consistency of achievement among subpopulations remained the same: Whites students outperformed their Black and Hispanic peers, Asian students slightly outperformed White students.

Overall mean score differences between race/ethnic groups is the primary basis used for achievement gap explanation in mathematics on NAEP Mathematics Assessment (Lubienski & Shelley, 2003; Hombo, 2003; NAGB, 2004; NCES, 2005). Previous research on NAEP Mathematics Assessments (1996 – 2003) found significant differences between groups by race and SES when the mathematics achievement was examined by content strand (Lubienski & Shelley, 2003; Schultz, Lee, Mullin, 2005; Tatsuoka, Tatsuoka, Carter, Tatsuoka, 2004). When Lubienski & Shelley (2003) did a comparative study of NAEP Mathematics Assessment for 8th-graders on content strands, between 1992 – 2000, they found that measurement showed the greatest disparity of achievement for Black students and White students, and data analysis showed the greatest achievement gap between White students and Hispanic students. Similarly, this study's analysis of the 2005 NAEP Mathematics by content strand found that measurement had the greatest achievement gap between Whites and their minority peers (46.75 points for Black students and 32 points for Hispanic students).

Schultz, Lee, and Mullen (2005) suggested that there were difficulty factors within items on 2003 NAEP Mathematics that affect achievement. Achievement performance on the 2005 8th grade NAEP varied. Strand differences in achievement and achievement gaps related to problem type were inconclusive. The present study suggested that while content strands were relatively similar to each other, items in numbers and operations were found to be the easiest in terms of language complexity, and items in measurement were found to be the most difficult. Strand and gap achievement differences by linguistic complexity could not be discerned with the statistical procedures used in the current study. However, differences between strands by language complexity categories do exist, for example, items that were classified as vocabulary on the Mathematics Assessment Language Framework (MALF) tended to test for skills in measurement and geometry.

Particularly disturbing were the substantial performance differences found between racial/ethnic groups in mathematics, and the fact that those gaps generally remained stable from 1990 to 2005 (Fagan, 2006; Lubienski, & Shelley, 2003; NCES, 2005; NSF, 2007). Previous research and findings from this study suggest that minority students, Black students in particular, are getting shortchanged in receiving educational opportunities to learn higher level mathematical concepts needed to perform spatial relationships found in measurement and geometry (gap between Whites students and Black students 46.75 and 32.77 respectively). The research suggests minority students with the highest achievement gaps are also lacking skills in grade level computation and problem-solving. The current study corroborated these findings. On the 2005 NAEP mathematics assessment the achievement gap for the strand of numbers and operations between White students and Black student was substantial at 34 points.

In the content analysis of the items by language categories, it was found that vocabulary was an important aspect of the NAEP 2005 assessment. Based on this finding, knowledge of mathematical vocabulary may improve performance across the strands. Content strand items were fairly distributed by problem type (multiple choice, constructed response, extended response) and by language categories on the Mathematics Assessment-Language Framework. All NAEP items were able to be sorted into two major categories, graphic (having pictorial representation) or non-graphic. All subcategories, draw /manipulate, operate and plan, graphic vocabulary, convert-only, convert and solve and non-graphic vocabulary, had item representation from each of the content strands and problem types. This suggests that the distribution of the Mathematics Assessment-Language Framework language categories was aligned to 2005 NAEP Mathematics specification content and problem type and may be able to expose differences on achievement gap by language complexity.

Both short constructed response items (SCR) and extended constructed response items (ECR) have greater individual differences in achievement gap than the overall strand for all five content areas. Guessing is minimized, for ECR and SCR items, as students need confirm or prove their mathematical thinking using the correct vocabulary, symbols and notation to get full credit. Students lacking an understanding of the language complexity of mathematic assessments would be at an inherent disadvantage causing greater gaps in overall performance. Because SCR and ECR items are weighted at 50% of the NAEP Mathematics Assessments, knowledge of mathematics language directly influences performance.

Primary to this study was the analysis of mathematical language on achievement performance and the determination of the type of language that would have the greatest effect on mathematical performance. The analysis focused on the question stems because stems set-up the intention of the action needed to operationalize and solve each problem. The student, after reading the problem, must decide what to select as the best answer (multiple choice), or to write explanations based on their understanding of the mathematics required to order to solve the problem correctly (SCR and ECR).

Analyses of the relationships among achievement gap ranking, problem type and language complexity categories revealed both positive and negative correlations. There was a strong negative correlation(-.81) between the percentage of multiple choice items and achievement gap rank – the more multiple choice items the smaller the achievement gap; and a moderate positive correlation (.59) existed between the percentage of constructed response items and achievement gap rank – the more constructed response items the higher the achievement gap. These findings suggest that the percentage of an item type within a content strand could be a factor in achievement differences by race.

In the analyses of relationships among achievement gap rank and language categories, a moderate negative association (-.70) between percentage of non-graphic vocabulary items and achievement gap rank existed, suggesting that non-graphic vocabulary items effected achievement differences in the content strands with the lowest gaps. There were no other significant associations found between language complexity and achievement gap ranks. This suggests that, while overall language complexity categories may not be a critical factor in determining achievement gap differences, non-graphic vocabulary as part of language complexity may be a factor in mathematical performance of race/ethnicity subpopulations in individual content strands.

In the analysis of the relationships among problem types and language complexity categories, a strong positive correlation (.97) existed between extended constructed response items and graphics vocabulary, this suggest that pictorial representation using specific vocabulary was a predominate feature of ECR items. There was strong negative correlation (-.84) between short constructed response items and graphic vocabulary. These two polar correlations suggest that there are differences in language complexity between ECR and SCR items. This finding supports the research that mathematical statements need to be analyzed into their structural components. By definition, the question stem on assessment items give the instructional cues or stimuli that convey to students the content elements to be selected and directions for what they are to do and how they are to do it (Cotton, 2000, Iman, 2004).

Because the Mathematics Assessment-Language Framework focused on intentionality of the questions rather than on a specific word language-factor analysis, it is possible that it could not capture enough of the subtle distinctions of the language effects that would lead to statistical significance. A content analysis of key words or phrases found within each of the items on the NAEP Math using the MALF categories could contribute to understanding the effects of ambiguity and intentionally of questions stems on achievement for mathematics assessments.

According to MacGregor (1990) major stumbling blocks for student proficiency in mathematics were the fact that students lacked the experience and skill in reading analytic text, and unfamiliarity with the standard uses of prepositions. It may be that minority students do not understand the linguistic cues in the question stems that provide students with the basic relational concepts that provide directions on selecting the discrete mathematic skills needed to problem solve. For example, the preposition *of* in a phrase in the question stem should indicate to the reader that the use of the multiplication operation is probably needed solve the problem.

Wakefield suggested that acquiring mathematic fluency follows similar social-constructs children use to become fluent in their primary language (2000). Recent studies suggested that if students come to school lacking in primary understanding of relational mathematical language concepts, it will be difficult for them to show gains in achievement starting in kindergarten or maintain pace with grade level instruction when much more sophisticated academic language is required to move beyond basic skills (Bailey & Butler 2003; Boehm 2000; NSF, 2007; and Zhou & Boehm 2004).

Much of the language of mathematics is found within the set of basic relational concepts needed for understanding and following directions as identified by Boehm (2000) and should be mastered by second grade, including concepts as such most, every, all, each; and comparing and ordering attributes using adverbs and prepositions such as first, right, left, inside, between. For example, in a language analysis of a fifth grade mathematics texts, it was found that the 20 the most commonly used words accounted for 34% of the total words used in the mathematics selections overall. Words that pertained to specific mathematics content (e.g. mode, median) accounted for 8% of the total and often occurred just once in the mathematics selections analyzed (Bailey & Butler, 2003).

Current trends of achievement show that young children start school in kindergarten with small disparities in achievement of basic math skills between White students and minority peers -Black student and Hispanic students (NSF, 2007). By the end of third grade, however, the difference between groups widen, as Whites students gain more mathematics knowledge and skills at faster rates than minority students. White students continue to obtain more advance skills leading to increasing achievement gap in mathematics performance whereby 12th grade, Black students and Hispanic students are outperformed by 8th grade White students (NSF, 2007). The examination of achievement gap by strand and the content analysis of items on the 2005 NAEP 8th- grade Math by problem type and language category suggest that the lack of basic relational language may persist through later grades.

Conclusions

1. The Mathematics Language Assessment Framework was able to discern that mathematics language complexity categories was represented across strand and problem type.
2. Language complexity is a factor for mathematic fluency on the 2005 NAEP Mathematics Assessment.
3. The results are inconclusive about the effect of language complexity on achievement gaps performance on 2005 NAEP Math; however, language complexity affects achievement for all racial/ethnic groups (White, Black, Hispanic, and Asian).

Implications for Research

The current study was a preliminary analysis intended to provide the structure needed to conduct statistical procedures based on the blueprint of the NAEP assessment. One way to continue this research on language fluency and mathematical assessments would be to use statistical procedures to separate the MALF categories by performance in order to judge the impact of language complexity on achievement.

Researchers could reanalyze the 2005 NAEP Mathematics Assessment using the Mathematics Assessment-Language Framework with more sophisticated statistical procedures such as ANCOVA. The ANCOVA would allow for the covariance of socio-economic status, parental education, and student motivation variables to be partialed from student achievement.

Future research should examine earlier NAEP data for trends or patterns across content strands by language category. This could inform and improve the validity and reliability of the framework by uncovering specific weaknesses in achievement in mathematics by strand and/or subpopulation. Currently, there is consistent lower achievement in measurement and geometry when compared to the content strands of algebra, data analysis, and numbers and operations; and with items involving spatial reasoning (NAEP 2000, 2003, 2005; TIMMS, 2003; PISA, 2003).

A comparison study could be conducted using the Mathematics Assessments-Language Framework with other assessments such as the state achievement tests, Iowa Basic Skills, and other standardized tests. A comparative content analysis could examine differences in linguistic complexity and problem types at the item level. Rankings could then be determined by mathematic language category.

Similar to the research by Butler, et al (2004), a language content analysis of mathematics assessments could be implemented in order to measure the impact of grammatical structures on student performances. The content analysis could include such words as prepositions, transitives, and ordinals that could cue student response. Butler concluded that these grammatical forms of English should be directly taught as part of mathematics instruction. In addition, a comparison of conceptual development research and achievement research regarding mathematics language issues could produce a more sophisticated language fluency framework. This may unearth the root of the math language fluency factor on the achievement gap. For example, studies from conceptual development research have suggested that basic relational concepts such as temporal ordering and prepositions are key factors for determining mathematical proficiency by the end of kindergarten (Boehm, 2002).

A quasi-experimental study can be conducted with 8th-grade mathematics students regarding language intentionality in question-stems of word problems. Comparing students with strategic training to those students without the training may demonstrate statistical significance of language intentionality. The strategic training could include grammar lexical rules, and opportunities to transfer skills used in comprehending reading literature based on specific text structures to comprehending mathematics assessment based on specific language used in question-stems.

Implications for Practice

Fluency in specific mathematical terminology is an important aspect of mathematics education for all students. Further emphasizing the understanding of relational basic concepts in early grades may increase the foundational knowledge needed for language fluency to develop mathematical reasoning and skills for many minority and at-risk students. Teaching for mathematical language fluency is similar to teaching for computation automaticity and is an essential component of mathematics comprehension (North Central Educational Laboratory, 2004). Additionally, reinforcing relational concepts in conjunction with other problem solving strategies may improve student performance on recall and lower level mathematical computation.

Critical to the instruction of mathematics are the meanings of words found in everyday language that have multiple interpretations. Direct instruction of these words is critical to helping students decipher the subtexts of intentionality that exist in the question-stem and reduce ambiguity. It can not be assumed that students have the analytic reading skills or syntactic awareness to read and respond appropriately to assessment items in mathematics. Learning how to read mathematical texts should be part of the mathematics curriculum. Teaching students specific strategies for understanding how the question stems function and providing cues for problem solving may improve assessment achievement in mathematics across the strands for all students.

Teacher preparation programs need to refocus the curricula to strengthen teacher knowledge of mathematics content and mathematics language fluency. Perhaps teacher candidates need to develop their own mathematical language fluency skills. They could have more observations of teacher-practitioners who are successful in promoting minority students' mathematic achievements.

Summary

The No Child Left Behind Act of 2001 has held accountability to high educational standards, mandating every child in the US be proficient on state and NAEP assessments by 2014. The purpose of this stringent mandate is twofold: first, to eradicate the achievement gap in reading and mathematics between white students and minority students; and second, to improve the standing of US students in international achievement tests used to prove the readiness for an economically competitive workforce (NCLB, 2001, Resnick, 2006). The degree to which the requirements of NCLB are pressuring schools and teachers to narrow curriculums to the subject and content areas that appear on standardized tests may, in fact, be problematic to obtaining the necessary achievement gains for minority and under-achieving students. There is very little flexibility for the classroom teacher to place into practice supplemental instructional activities such as math language strategies (Keyes, 2007; National Education Association, 2007). Keyes (2007) describes this narrowing of curriculum as a class caste system where students in underperforming schools have very focused instruction on recall and test preparations, while students in high performing schools are taught subjects not being measured for accountability indexes used for NCLB such as science, social studies, and creative projects.

Despite the stricter accountability measures and higher standards of NCLB, the continuing trend in the achievement gap between White students and their minority peers on the NAEP 8^{th} Mathematics Assessments (NCES, 2001, 2003, 2005) remains. Student achievement levels for the majority Black and Hispanic students are at Basic or Below Basic. Given that all students must meet the goal of 100 percent proficient scores on and mathematic assessment by the 2013-2014 school year, minority groups' lack of mathematic attainment is cause for concern. Practical and tangible ways must be found to increase achievement. Incorporating mathematics language strategies may be one way to improve achievement for all students.

While language proved to be an important factor, further study is needed to identify the salient features of mathematics assessments that prevent the Black and Hispanic students from completing the final transfer task required for sequestered problem-solving in standardized testing situations (NCREL, 2004). Findings in this research study suggest that there is a relationship between mathematical performance and language fluency for all groups. The current study explored the achievement gap by mathematic strands and examined items by problem type and language categories in order to examine if there is a relationship between student performance and mathematical language fluency.

References

Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometrics issues. *Educational Assessment, 8*(3), 231-257.

Abedi, J., & Lord, C. (2002). The language factor in mathematics tests. *Applied Measurement in Education,* 14(3), 219-234.

Abedi, J., Lord, C., & Hofsetter, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Educational Research, 74*(1), 1-28.

Adams, T. (2003). Reading mathematics: More than words can say. *The Reading Teacher, 56*(8), 786-794.

Aiken, L. R. (1971). Verbal factors and mathematics learning: A review of research. *Journal for Research in Mathematics Education, 2(4)*, 304-313.

Aiken L. R. (1972). Language factors in mathematics learning. *Review of Educational Research, 42,* 359-85.

Allen, N. L., Carlson, J. E., & Zelenak, C. A. (2000). *The NAEP 1996 Technical Report,* NCES, National Center for Education Statistics, U. S. Department of Education. Washington, DC: U. S. Government Printing Office.

Allison, P. (2001). *Missing Data #136: Quantitative applications in the social science series 07-136.* Thousand Oaks, CA: Sage Publications.

Arbaugh, F., Brown, C., Lynch, K., & McGraw, R. (2004). Students' ability to construct responses (1992-2000): Findings from short and extended constructed-response items. In P. Kosstermand & F. K. Lester, Jr. (Eds.), *Results and interpretation of the 1900-2000 mathematics assessments of the National Assessment of Educational Progress* (pp. 305-306). Reston, VA: NCTM.

Atweh, B., Bleicher, R., & Cooper, T. (1998). The construction of the social context of mathematics classrooms: A sociolinguistic analysis. *Journal for Research in Mathematics Education, 29*, 63-82.

Bailey, A. L., & Butler, F. A. (2003). *An evidentiary framework for operationalizing academic language for broad application to K-12 education: A design document.* (CSE Tech. Rep. 611). Los Angles: University of California, National Center Research on Evaluation, Standards, and Student Testing.

Beckemann, S. (2004). Solving algebra and other story problems with simple diagrams: A method demonstrated in Grade 4-6 texts used in Singapore. *The Mathematics Educator, 14(1),* 42-46.

Blinko, J. (2004). Mathematics in context. *Mathematics in Teaching*, (8), 3-9.

Bock, D., Brennan, R., & Muraki, E. (2002). The information in multiple ratings. *Applied Psychological Measurement, 26(*4), 364-375.

Boaler, J. (2002). Learning from teaching: Exploring the relationship between reform curriculum and equity. *Journal of Research in Mathematics Education*, *33*, 230-258.

Brownlee, K.A. (1960). *Statistical theory and methodology in science and engineering.* New York: John Wiley & Sons (p. 236). Retrieved: March 31, 2007, from http://www.itl.nist.gov/div898/handbook/mpc/section7/mpc7.htm#Brownlee.

Bybee, R., & Stage, E. (2005). No country left behind*. Issues in Science and Technology, 1*, 69-75.

Chomsky, N. (1975). *Reflections of language*. New York: Pantheon Books.

Ciani, A. (1981). Mastering word and symbol language in mathematics. *School Science and Mathematics, (81),* 371-377.

Cohen, D., & Hill, H. (2002). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record, 102*(2) 294-343.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfield, F., & York, R. (1966). *Equality of educational opportunity*. Washington, DC: Government Printing Office.

Coley, R. (2003). *Growth in school revisited: Achievement gains from the fourth to the 8th-grade . (Policy Information Report).* Princeton, NJ: Educational Testing Service.

Common Core Data (2003) *Characteristics of the 100 largest public elementary and secondary school districts in the United States: 2000-01*. National Center for Education Statistics. Washington DC: Government Printing Office.

Cotton, K. (2000). Classroom questioning. *School Improvement Research Series.* Retrieved: April 9, 2007, from http://www.nwrel.org/scpd/sirs/3/cu5.html.

Cronbach, L. (1942). An analysis of techniques for diagnostic vocabulary testing. *Journal of Educational Research, 36*, 206-217.

Curry, J (1996). The role of reading and writing in mathematics. In D. Lapp, J. Flood, & N. Farnan (Eds.), *Content area – Reading and learning instructional strategies (*2nd *Ed.)* (pp. 227-244). Boston: Allyn and Bacon.

Elementary and Secondary Education Act of 1965, P.L. No. 89-10. (1965).

Educational Research Service. (2001). What can schools do to close the achievement gap? *ERS On the Same Page Series*. Retrieved: December 12, 2006, from http://www.ers.org/otsp/otsp3.htm.

Erpenbach, W. J., Forte-Fast, E., & Potts, A. (2003). *Statewide educational accountability under NCLB: Central issues arising from an examination of state accountability workbooks and U.S. Department of Education Reviews under the No Child Left Behind Act of 2001*. Washington, DC: Council of Chief State School Officers.

Federal Registrar Part II. (2003). *Title I – Improving the achievement of the disadvantage: Final Rule (34-CFR Part 200).* Washington, DC: U.S. Department of Education.

Fenema, E., & Leder, G. (Eds.). (1990). *Mathematics and gender*. New York: Teacher College Press.

Forte-Fast, E., & Erpenbach, W. J. (2004). *Revisiting statewide educational accountability under NCLB: Central issues arising from an examination of state accountability workbooks and U.S. Department of Education Reviews under the No Child Left Behind Act of 2001*. Washington, DC: Council of Chief State School Officers.

Fuchs, L., & Fuchs, D. (2002). Mathematical problem-solving profiles of students with mathematics disabilities with and without comorbid reading disabilities. *Journal of Learning Disabilities, 35*(6), 563-573.

Gonzales, P., Guzman, J., Partelo, L., Pahlke, E., Jocelyn, J., Kastberg, D., & Williams, T. (2004). *Highlights from the Trends in International Mathematics and Science Study (TIMSS) 2003(NCES 2005-005).* National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Greenberg, E., Skidmore, D., & Rhodes, D. (2004) *Climates for learning: Mathematics achievement and its relationship to schoolwide student behavior, schoolwide parental involvement, and schoolwide morale.* Paper presented at the American Educational Research Association. San Diego.

Gutierrez, R. (2002). Beyond essentialism: The complexity of language in teaching mathematics to Latina/o students. *American Educational Research Journal, 39*(4), 1047-48.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning.* Baltimore, MD: University Park.

Heaton, J. (1998). Secondary analysis of qualitative data. *Social Research Update, Issue 22*. Retrieved: December 10, 2006, from http://www.soc.surrey.ac.uk/sru/SRU22.html.

Hegarty, M., Mayer, R., & Monk, C. (1995). Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *Journal of Educational Psychology 87, (1),* 18-32.

Hofstetter, C. (2003). Contextual and mathematics accommodation test effects for English-language learners. *Applied Measurement in Education, 16*(2), 159-188.

Hombo, C. (2003). NAEP and No Child Left Behind: Technical challenges and practical solutions. *Theory Into Practice, 42*(1), 59-65.

Honaker, J., & King, G. (2006). What to do about missing values in time series cross-section data? Retrieved: March 13, 2007, from http://gking.harvard.edu/amelia.

Horton, N. J., & Kleinman, K. P. (2007). Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American Statistician, 61*(1), 79-90.

Huntsinger, C., Shaligram, C., Jose, P., Larson, S., & Krieg, D. (2000). Mathematics, vocabulary, and reading development in Chinese American and European American children over the primary school years. *Journal of Educational Psychology, 92*(4), 745-760.

Inman, D (2005). *Natural language processing: Can computers understand language?* Retrieved: October 10, 2006, from http://www.aaai.org/AITopics/html/natlang. html.

Jitendra, A., DiPipi, C., & Perron, N. (2002). An exploratory study of schema-based word-problem-solving instruction for middle school students with learning disabilities: Emphasis on conceptual and procedural understanding. *The Journal of Special Education 36*(1), 23-38.

Johnson, E.G. (1992). The design of the National Assessment of Educational Progress. *Journal of Educational Measurement, 29*(2), 95-110.

Kelly, R. R., & Mousley, K. (2001). Solving word problems: More than reading issues for deaf students. *American Annuals of the Deaf*, *146*(3), 215-262.

Keyes, D. (Spring, 2007). Classroom caste system. Washington Post.com. Retrieved: April 9, 2007, from: http://www.washingtonpost.com/wpdyn/content/article/ 2007/04/08/AR2007040800925.html

Kober, N. (2001). *It takes more than testing: Closing the Achievement Gap*. Report of the Center on Educational Policy. Washington, DC.

Koponen, T., Mononen, R., Rasanen, P. & Ahonen, T. (2006). Basic numeracy in children with specific language impairment: Heterogeneity and connections to language. *Journal of Speech, Language and Hearing Research, 49*(1), 47-57.

Kosstermand, P., & Lester, F. K., Jr. (Eds.) (2004). *Results and interpretation of the 1900-2000 mathematics assessment of the National Assessment of Educational Progress* (pp. 305-306). Reston, VA: NCTM.

Lachance, A., & Confrey, J. (1995). Introducing fifth graders to decimal notation through ratio and proportion. Paper presented at the Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Columbus, OH.

Ladson-Billings, G. (1997). It doesn't add up: African American students' mathematics achievement. *Journal of Research in Mathematics Education*, *28*(6), 697-708.

Lai, K., Griffin, P., Mak, A., Wu, M., & Dulhunty, M. (2001). Modeling strategies in problem-solving. Paper presented at the Annual Conference of the Australian Association for Research in Education, Perth, Australia.

Lee, J. (2004). Multiple facets of inequity in racial and ethnic achievement gaps. *Peabody Journal of Education, 79*(2), 51-73.

Lee, S., & Xu, L. (2003). On local influence analysis of full information item factor models. *Psychometricka, 68*(3), 339-360.

Lemke, M., Sen, A., Pahlke, E., Partelow, L., Miller, D., Williams, T., Kastberg, D., & Jocelyn, J. (2004). *International Outcomes of Learning in Mathematics Literacy and Problem Solving: PISA 2003 Results from the U.S. Perspective. (NCES2005-003).* National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Leong, C., & Jerrod, W. (2001). Effects of consistency and adequacy of language information on understanding elementary mathematics word problems. *Annals of Dyslexia, 51*, 277-298.

Levine & Reed, M., M.D., & Reed, M., M.D. (2000). *Developmental Variation and Learning Disorders* 2nd Edition. Cambridge, MA : Educators Publishing Service Inc.

Lubienski, S. (2000a). A clash of class cultures? Students' experiences in a discussion-intensive seventh-grade mathematics classroom. *Elementary School Journal, 100,* 377-403.

Lubienski, S. (2000b). Problem solving as a means toward “mathematics for all”: An exploratory look through a class lens. *Journal for Research in Mathematics Education, 31*(4), 454-482.

Lubienski, S. (2004). *Instruction, achievement, and equity: Intersections of race/ethnicity and SES in NAEP data.* Paper presented at the annual meeting the American Educational Research Association in San Diego, CA.

Lubienski, S. (2006). Examining instruction, achievement and equity with NAEP mathematics data. *Education Policy Analysis Archives 14(14*). Retrieved: June 26, 2006, from http://epaa.asi/edi/epaa/v14n14/.

Lubienski, S., & Lubienski, C. (2006). What NAEP can tell about school achievement. *Education Week, 25*(26), 28-30.

Lubienski, S., McGraw, R., & Strutchens, M. (2004). NAEP findings regarding gender: Mathematics achievement, student affect, and learning practices. In P. Kosstermand & F. K. Lester, Jr. (Eds.), *Results and interpretation of the 1900-2000 mathematics assessments of the National Assessment of Educational Progress* (pp. 305-306). Reston, VA: NCTM.

Lubienski, S., & Shelley, M. (2003). *A closer look at U.S. mathematics instruction and achievement: Examinations of race and SES in a decade of NAEP data*. Paper presented at the American Education Research Association, Chicago.

MacGregor, M. (1990). Reading and writing in mathematics. In Bickmore-Brand, J (Eds.), *Language and Mathematics (*100-107).Portsmouth, NH: Heinemann.

Mayer, R. (1982). Memory for algebra story problems. *Journal of Educational Psychology, 74*, 199-216.

Merich, E., & Griffith, J. (1992). *International mathematics and science assessment: What have we learned?* Research and Development Report for the National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Moss, J., & Case, R. (1999). Developing children’s understanding of the rational numbers: A new model and experimental curriculum. *Journal for Research in Mathematics Education, 30*(2), 122-148.

Moses, R., & Cobb, C. (2001). *Radical equations: Civil rights from Mississippi to the Algebra Project.* Boston: Beacon Press.

National Assessment Governing Board. (2003). *Mathematics framework for the 2003 National Assessment of Educational Progress (College Board Rep. RN91084001)*. Washington, DC: U.S. Government Printing Office.

National Assessment for Educational Progress (2005). The Nation's Report Card. Retrieved [6/18/05] from http://nces.ed.gov/nationsreportcard/naepdata/

National Center for Educational Statistics. (1999). *NAEP 1999 trends in academic progress: Three decades of student performance*. U. S. Department of Education. Washington, DC: U.S. Government Printing Office.

National Center for Educational Statistics. (2005).*The condition of education 2000-2005*. U.S. Department of Education. Washington, DC: U.S. Government Printing Office.

National Commission on Excellence in Education. (1983). *A nation at risk*. Washington, DC: U.S. Government Printing Office.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA. Retrieved: November 17, 2004, from http://standards.nctm.org/document/prepost/project.htm.

National Council of Teachers of Mathematics (2006). *Curriculum focal points for pre-kindergarten through 8th-grade: A quest for coherence*. Reston, VA. Retrieved: March 28, 2006, from http://www.nctmmedia.org/cfp.

National Education Association. (2007). *No Child Left Behind' Act/ESEA.* Retrieved: April 12, 2007, from: http://www.nea.org/esea/index.html.

National Science Board. (2004). *Science and engineering indicators: Elementary and secondary education (Chap.1).* National Science Foundation. Retrieved: February 27, 2005, from http://www.nsf.gov/sbe/srs/seind04/ci/c1i.htm.

Neidorf, T., & Binkley, M. (2003). *Comparing NAEP, TIMSS and PISA in mathematics and science*. National Center for Educational Statistics. Washington, DC: U. S. Government Printing office.

No Child Left Behind Act, 2001 H.R., Pub.L. No. 107-110 Stat. 1425. (2002).

No Child Left Behind Act Guide to "Frequently Asked Questions." (2005). Retrieved [5/6/2005] from

Nowak, M., Komarova, N., & Niyogi, P. (2002). Computational and evolutionary aspects of language. *Nature*, *417*(6), 611-617.

Okpala, C., Okpala, A., & Smith, F. (2001). Parental involvement, instructional expenditures, family socioeconomic attributes, and student achievement. *The Journal of Educational Research, 95*(2) 110-115.

Pachtman, A., & Riley, J. (1978). Teaching the vocabulary of mathematics through interaction, exposure, and structure. *Journal of Reading*, 22, 240-244.

Pape, S. (2004). Middle school children's problem-solving behavior: A cognitive analysis from a reading comprehension perspective. *Journal for Research in Mathematics Education, 35*(3), 187-219.

Perkins, D. N., & Simmons, R. (1988). Patterns of misunderstanding: An integrative model for science, mathematics, and programming. *Review of Educational Research,* 58(3), 303-326.

Piaget, J. (2001). *The language and thought of a child.* London: Routledge Classics. (Original work published in 1926).

Quilici, J., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology, 88,* 144-161.

RAND Mathematics Study Panel. (2003). *Mathematical proficiency for all students: Toward a strategic research and development program in mathematics education*. Santa Monica, CA: RAND.

Resnick, L. (1987). *Education and learning to think.* National Academy Press, Washington, DC.

Resnick, L. (2005). Getting back on course. Retrieved: February 15, 2006, from http://aft.orgpubs-report/american-educaor/issues/spring 05/resnick.htm.

Reyes, B. J., Dingman, S., McNaught, M., Regis, T., & Togashi, J. (2006). *What mathematics are fourth graders in the U.S. expected to learn?* Columbia, MO: University of Missouri, Center for the Study of Mathematics Curriculum.

Roger, A., Kokolis, G., Stoeckel, J., & Kline, D. (2002). *National Assessment of Educational Progress 2000 Mathematics Assessment Secondary-Use Data Files Data Companion.* National Center for Education Statistics, Potomac, MD.

Rowan, B., Correnti, R., & Miller, R. (2002). *What large-scale survey research tells us about the teacher effects on student achievement: Insights from the Prospects Study of elementary schools*. Consortium for Policy Research in Education (CPRE Research Report–RR-051).

Schulz, E., Lee, W., & Mullen, K. (2005). A domain-level approach to describing growth in achievement. *Journal of Educational Measurement, 42*(1). 1-26.

Smith, T. (2004). Curricular reform in mathematics and sciences since *A Nation at Risk. Peabody Journal of Education*, 79(1), 105-129.

Sowder, J., Wearne, D., & Martin, G., & Struchens M. (2004). What do 8th-grade students know about mathematics? Changes over a decade. In P. Kosstermand & F. K. Lester Jr. (Eds.), *Results and interpretation of the 1900-2000 mathematics assessments of the National Assessment of Educational Progress* (pp. 305-306). Reston, VA: NCTM.

Stephenson, W. (1997). Factors as operant subjectivity. *Operant Subjectivity, 1,* 3-16.

Swanson, C., & Stevenson, D. (2002). Standards-based reform in practice: Evidence on state policy and classroom instruction from the NAEP state assessments. *Educational Evaluation and Policy Analysis*, 24(1), 1-27.

Tate, R. (1997). Race-ethnicity, SES, gender, and language proficiency trends in mathematics achievement: An update*. Journal for Research in Mathematics Education, 28*(6), 652-679.

Tatsuoka, K., Corter, J., & Tatsuoka, C. (2004). Patterns of diagnosed mathematical content and process skills in TIMSS-R across a sample of 20 countries. *American Educational Research Journal*, *41*(4), 901-926.

The College Board. (1999). Reaching the top: A report on the national task force on minority achievement. A National Task Force Report. Towers, J. (1992). Twenty-five years after the Coleman Report: What should we have learned? *The Clearinghouse, 65*(3), 138-140.

Turner,R needs initial. (2004). Motivational influences on student participation in classroom learning activities. *Teachers College Record, 106*(9), 1759-1785.

Vinovskis, M. (1998). Overseeing the “Nation’s Report Card”: The creation and evolution of the National Assessment Governing Board

Volger, K. (2002). Standards-based reform in practice: Evidence on state policy and classroom instruction form the NAEP state assessments. *Educational Evaluation & Policy Analysis, 24*(1), 39-50.

Wakefield, D. (2000). Math as a second language. *The Educational Forum, 64,* 272-279.

Wenglinsky, H. (2004). Closing the Racial Achievement Gap: The Role of Reforming Instructional Practices. *Educational Policy Analysis Archives*, *12*(64). Retrieved: January 26, 2005, from http://epaa.asu.edu/epaa/v12n64.

Wilkins, J., & Xin, M. (2002). Predicting Student Growth in Mathematical Knowledge Content Knowledge. *The Journal of Educational Research, 95*(5), 288-298.

Wong, K., & Nicotera, A. (2004). Brown v. Board of Education and the Coleman Report: Social science research and the debate on educational equality. *Peabody Journal of Education*, *79*(2), 122-135.

Xin, Y., Jitendra, A., & Deatline, A. (2005). Effects of mathematical word problem-solving instruction on middle students with learning problems. *The Journal of Special Education, 39*(3), 181-192.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and likert scale (ordinal) item scores.* Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Appendix A

Researcher's Qualifications

Yvette Marie Fagan

[redacted]

[redacted]

VITA

EDUCATIONAL BACKGROUND

Ed.D.	Learning and Instruction, Emphasis in Educational Statistics, University of San Francisco, San Francisco, CA (An Analysis 2005 NAEP Mathematics Achievement Items by Content Strand, Problem Type, and Language Complexity-December, 2007)
M.A.	Curriculum and Instruction (Special Education), University of San Francisco, San Francisco, CA, 1996
B.S.	Journalism, Ohio University, Athens, OH, 1976

LICENSURE AND CERTIFICATION

2001	Cross-cultural, Language and Academic Development (CLAD) Certificate, California Teaching Commission
1993	Special Education Learning Handicapped, University of San Francisco, San Francisco, CA
1991	Multiple Subject Ryan Clear Credential, University of San Francisco, San Francisco, CA

PROFESSIONAL EXPERIENCES

1/07-5/07	Adjunct Faculty, Learning and Instruction Department (Technology & Curriculum Integration for Special Education), University of San Francisco, San Francisco, CA
9/06-12/07	Teaching Assistant, Learning and Instruction Department (Advanced Statistics – Doctoral Course), University of San Francisco, San Francisco, CA
9/06-6/07	Adjunct Faculty, Learning and Instruction Department (Master Thesis Special Education), University of San Francisco, San Francisco, CA

1/06-5/06	Adjunct Faculty, Learning and Instruction Department (Technology & Curriculum Integration for Special Education), University of San Francisco, San Francisco, CA
9/05-5/06	Adjunct Faculty, Learning and Instruction Department (Master Thesis Special Education), University of San Francisco, San Francisco, CA
9/05-12/05	Teaching Assistant, Learning and Instruction Department (Advanced Statistics – Doctoral Course), University of San Francisco, San Francisco, CA
5/05-8/05	Adjunct Faculty, Teacher Education Department, Curriculum & Instruction (Math and Science), University of San Francisco, San Francisco, CA
5/05-7/05	Teaching Assistant, Learning and Instruction Department (Educational Research Methods – Masters Course), University of San Francisco, San Francisco, CA
5/04-8/04	Adjunct Faculty, Teacher Education Department, Curriculum & Instruction (Math and Science), University of San Francisco, San Francisco, CA
5/04-7/04	Teaching Assistant, Learning and Instruction Department (Educational Research Methods – Masters Course), University of San Francisco, San Francisco, CA
5/03-8/03	Adjunct Faculty, Teacher Education Department, Curriculum & Instruction (Math and Science), University of San Francisco, San Francisco, CA
5/03-8/03	Adjunct Faculty, Teacher Education Department, Curriculum & Instruction (Exceptional Children), University of San Francisco, San Francisco, CA
5/02-8/02	Adjunct Faculty Teacher Education Department, Curriculum & Instruction (Exceptional Children), University of San Francisco, San Francisco, CA
2001-Present	Co-Lecturer: Learning and Instruction Department (Special Education), University of San Francisco, San Francisco, CA
2000-Present	Fifth Grade Teacher: Dr. William L. Cobb Elementary School, San Francisco Unified School District, San Francisco, CA
6/00-8/00	English/Humanities Teacher (Grades 9-10): Bayview Learns Community School, Thurgood Marshall High School, San Francisco Unified School District, San Francisco, CA

4/00-8/00 Curriculum Coordinator (Grades 9-12): Bayview Learns Community School, Thurgood Marshall High School, San Francisco Unified School District, San Francisco, CA

1998-2000 Fifth Grade Teacher /Lead Teacher (Grades 3-5): Edison Charter Academy, San Francisco Unified School District, San Francisco, CA

1998-2000 Mathematics & Science Coordinator: Edison Charter Academy, San Francisco, Unified School District, San Francisco, CA

6/97-8/97 Gifted & Talented Summer Program Social Studies / Science Teacher (Grades 6-8): Benjamin Franklin Middle School, San Francisco Unified School District, San Francisco, CA

1996-1999 Lecturer: Adult Education (Classroom Strategies for Teaching Reading, K-12) University of California Berkeley Extension, San Francisco, CA

1995-1998 Reverse Full Inclusion Teacher, (Grades 3-5) Bryant Elementary School, San Francisco Unified School District, San Francisco, CA

1991-1995 Special Day Class Teacher (Grades 2-5) Bryant Elementary, San Francisco Unified School District, San Francisco, CA

6/91-8/91 Kindergarten Teacher Bryant Elementary School, San Francisco Unified School District, San Francisco Unified School District, San Francisco, CA San Francisco, CA

PROFESSIONAL PRESENTATIONS

Fagan, Y. M. (2007, April). Building Expository Skills. Presentation at The Council for Exceptional Children, Louisville, KY.

Bui, Y. N., & Fagan, Y. M. (2007, April). Reading Comprehension Strategies. Presentation at The Council for Exceptional Children, Louisville, KY.

Andrews, L., Fagan, Y. M., & Salazar, L. (2006, November). Poster Session at The Council for Exceptional Children, Teacher Education Division, San Diego, CA

Evans, S., Reid, T., Sanchez, E., Fagan, Y. M., & Salazar, L. (2006, November) .Addressing the Faculty Shortage in Special Education, Panel Presentation at The Council for Exceptional Children, Teacher Education Division, San Diego, CA

Bui,Y. N., Simpson, R., & Fagan, Y. M. (2006, November). Utilizing Culturally Diverse Strategies to Effect Achievement with Minority Students With Disabilities. Presentation at The Council for Exceptional Children, Division of Learning Disabilities, San Francisco, CA

Fagan, Y. M., Dennen, L., Bosco, V., & Cannon, T. (2006, April). Integrating Standards-based Science Curriculum Through Web-based Instruction: A California Technology Assistance Project, Paper Presentation at American Educational Research Association, Learning Environments Division C, San Francisco, CA

Fagan, Y. M., Salazar, L., & Andrews, L. (2006, April). Intern Teachers Attitude, Aptitude, and Use of Educational Technology in Urban Special Education Classrooms, Paper Presentation at American Educational Research Association, Teacher Education Division, San Francisco, CA

Fagan, Y. M., & Salazar, L. (2005, November). Delivering Standards-based Science Instruction through Technology: A Full Inclusion Model, Kaleidoscope Presentation at Council of Exceptional Children, Teacher Education Division, Portland, ME

Bui, Y.N., & Fagan, Y. M.. (2005, September). Culturally Balanced Educational Practices: Increasing Reading Comprehension for Culturally & Linguistically Diverse Students, Poster Presentation at the LASER Urban Education Research Conference, Tampa, FL

Fagan, Y. M., Dennen, L., Bosco, V., & Cannon, T. (2005, April). The Impact of Internet-based Science Instruction On Student Achievement, Attitude, and Behavior in an Inner-city Fifth Grade Class, Paper Presentation at the University of San Francisco, San Francisco, CA

Fagan, Y. M., Dennen, L., Bosco, V., & Cannon, T., (2005, February) The Impact of Internet-based Science Instruction On Student Achievement, Attitude, and Behavior in an Inner-city Fifth Grade Class, Paper Presentation at the California Technology Assistance Project, Region IV, Alemeda, CA

Fagan, Y. M., & Salazar, L. (2004, November). Designing Technology Curriculum for the Urban Special Education Intern Teacher, Paper Presentation at Council of Exceptional Children, Teacher Education Division, Albuquerque, NM

Evans, S., Lo, L., Smith, S., Fagan, Y., Reid, T., Mayfield, K., Nares-Guzicki, I., Salazar, L., & Kutaka-Kennedy, J. (2004, November). Diversifying the Special Education Doctoral Pool and the Professorate, Panel Presentation at the Council of Exceptional Children, Teacher Education Division, Albuquerque, NM

Fagan, Y. M. (2004, May). The Human Body Experience Poster Presentation at the University of California San Francisco, Partnership Conference and Celebration, San Francisco, CA

Fagan, Y. M. (2003, September). Developing Community-Based Curriculum Workshop Presentation at the Bayview Learns Education Summit, San Francisco, CA

Fagan, Y. M. (2003, May). Project-Based Curriculum Poster Presentation at Urban Special Education Best Practices University of San Francisco, San Francisco, CA

Paik, S., Andreatta, P., Bosco, V., Donahue, L., Elliot, M., Fagan, Y., Healy, W., Hood, J., Lo, L., & McGee, K. (April, 2003). Celebrated Voices & Videography. Poster Presentation at American Education Research Association, Chicago. IL

Fagan, Y. M. (June, 1993). Featured Presenter at Impact II, Teacher Leadership Conference, Salt Lake City, UT

WORKSHOP PRESENTATIONS

Bui, Y., Fagan, Y. (2/07) The Effects of an Integrated Reading Comprehension Strategy for Fifth Grade Students with Learning Disabilities in Inclusive Settings in Low-Performing Urban Schools for the California Association of Resource Specialist and Special Education Teachers Convention, San Francisco, CA

Fagan, Y.M. (1/07). Mathematics Review: Homework Help for Parents. Cobb Elementary, San Francisco, CA

Bui, Y. N., & Fagan, Y. M. (7/05). Special Education in Catholic Schools for SummerWest: The Institute for Catholic Educational Leadership, University of San Francisco, San Francisco, CA

Fagan, Y. M., & Lujan, J. (04/05). Integrating Technology into Curriculum K- 5: Series of Two Workshops: *Multimedia and Core and Core Content Instruction* and *Investigating Online Educational Sources*, Sanchez Elementary School, San Francisco, CA

Fagan, Y. M., Green, A., & Lujan, J. (01/05 – 04/05). Integrating Technology into Curriculum K-5: Series of Five Workshops: *Video Streaming, Interactive Websites*, *Inspiration, Microsoft Word- "The Toolbar"*, and *Website Construction using Online Templates*, Dr. Cobb Elementary, San Francisco, CA

Fagan, Y. M. (04/02). Differentiating Instruction, Dr. Cobb Elementary School, San Francisco, CA

Edfund Staff, Fagan, Y.M. (1995-97). Grant Writing in conjunction with San Francisco Ed Fund, San Francisco, CA

HONORS AND DISTINCTIONS

2006	Outstanding Educator of the Year, the Council for Exceptional Children Division for Learning Disabilities, San Francisco, CA
2006	Merit Scholar, University of San Francisco. San Francisco, CA
2005	Finalist in General Excellence Category for National Leaders in Learning Awards, Cable in the Classroom, Washington D.C.
2004	Teacher Recognition Award, California Technology Assistance Project, Region IV
2002-2005	Doctoral Fellow, University of San Francisco, San Francisco, CA
2001-2005	Scholarship Recipient, Minority Competition of OSEP Multicultural Urban Special Education Grant, University of San Francisco, San Francisco, CA
2003	Participant, National Assessment on Educational Progress Database Training Seminar, Washington D.C.
1999	Featured Model Teacher in PBS Documentary “The Digital Divide.”
1996	Golden Apple Teacher of the Year San Francisco Education Fund sponsored by KRON TV, San Francisco, CA

PROFESSIONAL ACTIVITIES

2004	Curriculum Committee Member, De Young Museum, San Francisco, CA
2000-2005	Advisory Board Member for Bayview Learns Non-Profit Organization, San Francisco, CA
1993-Present	Advisory Board Member for Special Education Credentialing Program, University of San Francisco, San Francisco, CA
1995-1999	San Francisco Education Fund Grant Committee for Team and Middle School Grants, San Francisco, CA

GRANT AWARDS

Fagan, Y. M., & Chinn, C.K. (2000). Awarded $2000 for "Using Collaboration and Inclusion for the Teaching of Science an Urban 5th grade Sped and General Ed classroom. HASS Foundation, San Francisco, CA

Fagan, Y. M., & Henry, R. (2005). Awarded $20,000 (In-kind) for Computer Technology by the University of San Francisco Retired Computers Program

Fagan, Y. M., & Chinn, C. K. (2005). Awarded $2500 for "Fostering Academic Empowerment Through the Use of the Internet as an Educational Tool" by Jordan Fundamentals Grant Program, Nike Corporation

Fagan, Y. M. (2004). Awarded $1000 for Technology Innovation in the Classroom by California Assistance Technology Project, CTAP Region IV

Fagan, Y. M. (1998). Awarded $2000 for "Rules of the Game," Math, social studies and technology by San Francisco Ed Fund, San Francisco,

Fagan, Y. M. (1997). Awarded $2000 for "Puppetry Productions," Literacy, art and social studies by San Francisco Ed Fund, San Francisco, CA

Fagan, Y. M. (1996) Awarded $2000 for "Machines in Motion," Physics, math, science, technology, literacy by San Francisco Ed Fund, San Francisco, CA

Zita, A., Davis, V., Fagan, Y. M. (1996) Awarded $50,000 for Networking ISDN Lines by Pacific Bell, San Francisco, CA

Fagan, Y. M. (1994-98) Awarded $10,000 for "Adventures in Outdoor Education," Environmental education by American Youth Hostels, San Francisco, CA

Zita, A, Chou, E, Fagan, Y. M. (1994-98) Awarded $35,000 for Community based partnerships for educational reform by Service Learning/AmeriCorps, San Francisco, CA

Fagan, Y. M. (1993-98) Awarded $20,000 for Swimming Instruction for Special Education Students (Grades 2-5) by Embarcadero, YMCA, San Francisco, CA

Fagan, Y. M (1993) Awarded $1000 Art of Writing," Literacy, art, technology by San Francisco Ed Fund, San Francisco, CA

Fagan, Y.M., Allison, E. (1993) Awarded $2000 Exploration of Ecosystems through Collaboration," Science, literacy, social studies by San Francisco Ed Fund, San Francisco, CA

Fagan, Y.M., Farrow, K. (1992-93) Awarded $14,000 for Art-In-Schools by Wells Fargo Foundation, San Francisco, CA

Fagan, Y. M. (1992) Awarded $1000 for "Multiculturalism Through Art," Art, social studies, literacy by San Francisco Ed Fund, San Francisco, CA

Appendix B

Computer Security Act of 1987

Public Law 10-235

U.S. DEPARTMENT OF EDUCATION
INSTITUTE OF EDUCATION SCIENCES

NATIONAL CENTER FOR EDUCATION STATISTICS

February 21, 2007

Yvette Fagan
University of San Francisco
1706 McAllister
San Francisco, CA 94115

Dear Ms. Fagan:

Thank you for submitting your application to access secure NAEP items. The application was reviewed by the National Center for Education Statistics (NCES), and your request has been *granted.*

NCES will provide access to all requested main NAEP items. This includes all 2005, grade 8 mathematics items.

Laura Bufford will contact you to arrange a viewing session of the approved items, in accordance with NCES procedures for providing researchers with access to secure NAEP cognitive items. Please review the guidelines for access that you received with the application form in advance of the session.

On behalf of NCES, I thank you for your patience and cooperation.

Sincerely,

Peggy Carr

Peggy G. Carr, Ph.D.
Associate Commissioner
Assessment Division

Cc: Marilyn Binkley

WASHINGTON D.C. 20006-

12/05/95

LICENSE FOR THE USE OF INDIVIDUALLY IDENTIFIABLE INFORMATION PROTECTED UNDER THE NATIONAL EDUCATION STATISTICS ACT OF 1994, AS AMENDED, AND THE PRIVACY ACT OF 1974

WHEREAS, the National Center for Education Statistics (NCES) in the Office of Educational Research and Improvement (OERI) of the United States Department of Education has collected individually identifiable information, the confidentiality of which is protected by the Privacy Act of 1974, 5 U.S.C. 552a, and sections 408 and 411 of the National Education Statistics Act of 1994 20 U.S.C. 9001 et seq., as amended, and

WHEREAS, NCES wishes to make the data available for statistical purposes to requestors qualified and capable of research and analysis consistent with the statistical purposes for which the data were provided, but only if the data are used and protected in accordance with the terms and conditions stated in this license, upon receipt of such assurance of qualification and capability, it is hereby agreed between

University of San Francisco,

(Insert the name of the agency or organization to be licensed)
hereinafter referred to as the "Licensee", and NCES that:

I. INFORMATION SUBJECT TO THIS AGREEMENT

A. All data containing individually identifiable information (including Schools in the National Assessment of Educational Progress) collected by or on the behalf of NCES under sections 408 and 411 of the National Education Statistics Act of 1994, as amended, that are provided to the Licensee and all information derived from those data, and all data resulting from merges, matches, or other uses of the data provided by NCES with other data are subject to this license and are referred to in this license as subject data.

B. Subject data under this license may be in the form of computer tapes, diskettes, CD-ROMs, hard copy, etc. The Licensee may only use the subject data in a manner and to a purpose consistent with:
(1) the statistical purpose for which the data were supplied, (Licensee's description of the research and analysis which is planned is attached and made a part of this license - Attachment No. 1.)
(2) the limitations imposed under the provisions of this license and,
(3) sections 408 and 411 of the National Education Statistics Act of 1994, as amended, and 5 U.S.C. 552a, which are attached to and made a part of this license (Attachment. No 2.)

II. INDIVIDUALS WHO MAY HAVE ACCESS TO SUBJECT DATA

A. There are four categories of individuals that the Licensee may authorize to have access to subject data. The four categories of individuals are as follows:

1. The Principal Project Officer (PPO) is the most senior officer in charge of the day-to-day operations involving the use of subject data and is responsible for liaison with NCES.

2. Professional/Technical Staff (P/TS) conduct the research for which this license was issued.

3. Support staff includes secretaries, typists, computer technicians, messengers, etc. Licensee may disclose subject data to support staff who come in contact with the subject data in course of their duties only to the extent necessary to support the research under this license.

4. An independent researcher is an individual who has satisfied the requirements specified in paragraph II.C. of this license.

B. Licensee may disclose subject data to only seven (7) P/TS unless NCES provides written authorization for a larger number of P/TS.

C. Licensee may disclose subject data to individuals who desire to do independent research, under the following conditions:

1. The independent researcher submits an application for access to subject data to NCES directly, or through the Licensee.

2. NCES provides written approval for the Licensee to disclose subject data to the independent researcher.

3. The Licensee completes the affidavit procedures in paragraph IV.B. of the license.

III. LIMITATIONS ON DISCLOSURE

A. Licensee shall not use or disclose subject data for any administrative purposes nor may they be applied in any manner to change the status, condition, or public perception of any individual regarding whom subject data is maintained. (Note: Federal Law pre-empts any State law that might require the reporting or dissemination of these data for any purpose other than the statistical purposes for which they were collected.)

B. Licensee shall not disclose subject data or other information containing, or derived from, subject data at fine levels of geography, such as school district, institution, or school, to anyone other than NCES employees working in the course of their employment or individuals for whom access is authorized under this license agreement. Licensee may make disclosures of subject data to individuals other than those specified in this paragraph only if those individuals have executed an affidavit of nondisclosure and the Licensee has obtained advance written approval from NCES.

C. Licensee shall not make any publication or other release of subject data listing information regarding individuals even if the individual identifiers have been removed.

D. Licensee may publish the results, analysis, or other information developed as a result of any research based on subject data made available under this license only in summary or statistical form so that the identity of individuals contained in the subject data is not revealed.

IV. ADMINISTRATIVE REQUIREMENTS

A. The research conducted under this license and the disclosure of subject data needed for that research must be consistent with the statistical purpose for which the data were supplied.

B. Execution of affidavits of nondisclosure.

1. Licensee shall provide a copy of this agreement, together with the attached SECURITY PROCEDURES (Attachment No. 3) to each employee of the licensee who will have access to subject data and shall require each of those employees to execute an affidavit of nondisclosure. Licensee shall also provide a copy of the attached SECURITY PROCEDURES, and the abstracted statement of the statistical purpose for which the data were supplied, to each independent researcher approved by NCES who the licensee intends to have access to subject data and shall require each of those researchers to execute an affidavit of nondisclosure.

2. The Licensee must ensure that each individual who executes an affidavit of nondisclosure reads and understands the materials provided to her or him before executing the affidavit.

3. Licensee shall ensure that each affidavit of nondisclosure is notarized upon execution.

4. Licensee may not permit any individual specified in paragraph II.A. to have access to subject data until the procedures in paragraphs IV.B.1. through 3. of this license are fulfilled for that individual.

5. Licensee shall promptly, after the execution of each affidavit, send the original affidavit to NCES and shall maintain a copy of each affidavit at the licensee's secured facility protected under this license.

C. Notification regarding authorized individuals to NCES.

1. Licensee shall promptly notify NCES when any employee who has been authorized to have access to subject data no longer has access to those data.

2. If the terms of an independent researcher's application specify when the researcher's access to subject data terminates and access does terminate on that date, the Licensee need not notify NCES of that fact. However, if the researcher's access terminates on another date, the Licensee shall promptly notify NCES of the date that such access terminates.

D. Publications made available to NCES.

1. Licensee shall provide NCES a copy of each publication containing information based on subject data or other data product based on subject data made available to individuals who have not executed an affidavit of nondisclosure.

2. When publication or other release of research results could raise reasonable questions regarding disclosure of individually identifiable information contained in subject data, copies of the proposed publication or release must be provided to NCES before that disclosure is made so that NCES may advise whether the disclosure is authorized under this license and the provisions of sections 408 and 411 of the National Education Statistics Act of 1994, as amended, and 5 U.S.C. 552a. Licensee agrees not to publish or otherwise release research results provided to NCES if NCES advises that such disclosure is not authorized.

E. Licensee shall notify NCES immediately upon receipt of any legal, investigatory, or other demand for disclosure of subject data.

F. Licensee shall notify NCES immediately upon discovering any breach or suspected breach of security or any disclosure of subject data to unauthorized parties or agencies.

G. Licensee agrees that representatives of NCES have the right to make unannounced and unscheduled inspections of the Licensee's facilities, including any associated computer center, to evaluate compliance with the terms of this license and the requirements of sections 408 and 411 of the National Education Statistics Act of 1994 and 5 U.S.C. 552a.

V. SECURITY REQUIREMENTS

A. Maintenance of, and access to, subject data.

1. Licensee shall retain the original version of the subject data at a single location and may make no copy or extract of the subject data available to anyone except a P/TS or independent researcher as necessary for the purpose of the statistical research for which the subject data were made available to the Licensee.

2. Licensee shall maintain subject data (whether maintained at a mainframe facility, remote terminals, personal computers, or on printed or other material) in a space that is limited to access by authorized personnel.

3. Licensee shall ensure that access to subject data maintained in computer memory is controlled by password protection. For subject data maintained on a mainframe computer, password protection is required at the file level. Licensee shall maintain all print-outs, diskettes, personal computers with subject data on hard disks, or other physical products containing individually identifiable information derived from subject data in locked cabinets, file drawers, or other secure locations when not in use.

4. Licensee shall ensure that all printouts, tabulations, and reports are edited for any possible disclosures of subject data.

5. Licensee shall establish procedures to ensure that subject data cannot be extracted from a computer mainframe, remote terminals or separate PCS by unauthorized individuals.

6. Licensee shall not permit removal of any subject data from the limited access space protected under the provisions of this license as required in the attached SECURITY PROCEDURES, without first notifying, and obtaining written approval from, NCES.

B. Retention of subject data.

Licensee shall return to NCES all subject data, or destroy those data under NCES supervision or by approved NCES procedures when the research that is the subject of this agreement has been completed or this license terminates, whichever occurs first.

C. Compliance with established security procedures.

Licensee shall comply with the SECURITY PROCEDURES attached to this license.

VI. PENALTIES

A. Any violation of the terms and conditions of this license may subject the Licensee to immediate revocation of the license by NCES.

1. The NCES official responsible for liaison with the Licensee shall initiate revocation of this License by written notice to Licensee indicating the factual basis and grounds for revocation.

2. Upon receipt of the notice specified in paragraph VI.A.1 of this license, the Licensee has thirty (30) days to submit written argument and evidence to the Commissioner of NCES indicating why the License should not be revoked.

3. The Commissioner shall decide whether to revoke the license based solely on the information contained in the notice to the Licensee and the Licensee's response and shall provide written notice of the decision to the Licensee within forty-five (45) days after receipt of Licensee's response. The Commissioner may extend this time period for good cause.

B. Any violation of this license may also be a violation of Federal criminal law under the Privacy Act of 1974, 5 U.S.C. 552a, and/or sections 408 and 411 of the National Education Statistics Act of 1994, 20 U.S.C. 9001 et seq., as amended. Alleged violations under the National Education Statistics Act of 1994 are subject to prosecution by the United States Attorney. The penalty for violation of sections 408 and 411 of the National Education Statistics Act of 1994, as amended, is a fine of not more than $250,000 and imprisonment for a period of not more than five years.

VII. PROCESSING OF THIS LICENSE

A. The term of this license shall be for five years. If, before the expiration of this license, the Commissioner establishes regulatory standards for the issuance and content of licenses, the Licensee agrees to comply with the regulatory standards.

B. This license may be amended, extended or terminated by mutual written agreement between the Licensee and the Commissioner, NCES. Any amendment must be signed by a Senior Official specified in paragraph VII.C. of this license, PPO, and the Commissioner and is effective on the date that all required parties have signed the amendment.

C. The Senior Official (SO) having the authority to bind the organization to the terms of the license, shall sign this license below. The SO certifies, by his/her signature, that -

1. The organization has the authority to undertake the commitments in this license;

2. The SO has the authority to bind the organization to the provisions of this license; and

3. The PPO is the most senior statistical officer for the licensee who has the authority to manage the day-to-day statistical operations of the Licensee.

Susan Evans — 9.30.03

Signature of the Senior Official — Date

Dr. Susan Evans

Type/Print Name of Senior Official

Title: Faculty, Department Chair — Telephone: (415) 422-5892

D. The individual described in paragraph II.A1. as the PPO shall sign this license below. If the SO also acts as the chief statistical officer for the Licensee; viz. as the PPO, the SO shall likewise sign under this paragraph as well as having signed under paragraph VII.C.

9/30/03

Signature of the Principal Project Officer — Date

Dr. Yvonne Bui

Type/Print Name of Principal Project Officer

Title: Faculty — Telephone: (415) 422-2516

E. If Licensee is an agency of a State or a State Contractor that provides data to any agency of a State, the Attorney General of the State shall sign this license below. The Attorney General certifies, by his/her signature, that pursuant to the Privacy Act of 1974, 5 U.S.C. 552a, and sections 408 and 411 of the National Education Statistics Act of 1994, 20 U.S.C. 9001 et seq., as amended, the Attorney General and any successors to that office will not construe any State or local laws to require Licensee to report subject data in individually identifiable form to any agency of the State or to any individual who has not executed an affidavit of nondisclosure which prohibits that individual from using subject data for any purpose other than the statistical purpose for which the subject data were collected.

Signature of the Attorney General — Date

F. The Commissioner of the National Center for Education Statistics issues this license to the University of San Francisco. The license is effective as of the date of the Commissioner's signature below, or such other period specified in the Licensee's request for the license.

Edith McArthur for VP
Commissioner, National Center for Education Statistics
Valena W. Plisko
Associate Commissioner
Type/Print Name of Commissioner, NCES

15 October 2003
Date

NCES License Control Number: 030904785

AFFIDAVIT OF NONDISCLOSURE

Professor
(Job Title)

(Date of Assignment to NCES Project)

University of San Francisco
(Organization, State or local agency or instrumentality)

2130 Fulton St SF CA 94117-1080
(Organization or agency Address)

NAEP Reading - Writing Math, Science 2002-
(NCES Data Base or File Containing Individually Identifiable Information*)

I, Yvonne Bui, do solemnly swear (or affirm) that when given access to the subject NCES data base or file, I will not -

(i) use or reveal any individually identifiable information [including "schools" in the National Assessment of Educational Progress (NAEP)] furnished, acquired, retrieved or assembled by me or others, under the provisions of Sections 408 and 411 of the National Education Statistics Act of 1994 (20 U.S.C. 9001 et seq.) for any purpose other than statistical purposes specified in the NCES survey, project or contract;

(ii) make any disclosure or publication whereby a sample unit or survey respondent (including "schools" in NAEP) could be identified or the data furnished by or related to any particular person or NAEP school under these sections could be identified; or

(iii) permit anyone other than the individuals authorized by the Commissioner of the National Center for Education Statistics to examine the individual reports.

(Signature)

[The penalty for unlawful disclosure is a fine of not more than $250,000 (under 18 U.S.C. 3571) or imprisonment for not more than five years (under 18 U.S.C. 3559), or both. The word "swear" should be stricken out when a person elects to affirm the affidavit rather than to swear to it.]

* Request all subsequent followups that may be needed. This form cannot be amended by NCES, so access to databases not listed will require submitting additional notarized Affidavits.

City/County of San Francisco Commonwealth/State of California
Sworn to and subscribed before me this 28th day of May, 2003. Witness my hand and official Seal.

RUTH R. DIONISIO
COMM. # 1322370
NOTARY PUBLIC - CALIFORNIA
SAN FRANCISCO COUNTY
My Comm. Expires September 27, 2005
CRS 2

(Notary Public/Seal) May 28 2003

My commission expires Sept 27, 2005

11/14/95

AFFIDAVIT OF NONDISCLOSURE

Professor
(Job Title)

(Date of Assignment to NCES Project)

University of San Francisco
(Organization, State or local agency or instrumentality)

2130 Fulton St SF CA 94117-1080
(Organization or agency Address)

NAEP - Reading, Writing, Math, Science, 2002-07
(NCES Data Base or File Containing Individually Identifiable Information*)

I, Susan Evans, do solemnly swear (or affirm) that when given access to the subject NCES data base or file, I will not -

(i) use or reveal any individually identifiable information [including "schools" in the National Assessment of Educational Progress (NAEP)] furnished, acquired, retrieved or assembled by me or others, under the provisions of Sections 408 and 411 of the National Education Statistics Act of 1994 (20 U.S.C. 9001 et seq.) for any purpose other than statistical purposes specified in the NCES survey, project or contract;

(ii) make any disclosure or publication whereby a sample unit or survey respondent (including "schools" in NAEP) could be identified or the data furnished by or related to any particular person or NAEP school under these sections could be identified; or

(iii) permit anyone other than the individuals authorized by the Commissioner of the National Center for Education Statistics to examine the individual reports.

Susan Evans
(Signature)

[The penalty for unlawful disclosure is a fine of not more than $250,000 (under 18 U.S.C. 3571) or imprisonment for not more than five years (under 18 U.S.C. 3559), or both. The word "swear" should be stricken out when a person elects to affirm the affidavit rather than to swear to it.]

* Request all subsequent followups that may be needed. This form cannot be amended by NCES, so access to databases not listed will require submitting additional notarized Affidavits.

City/County of San Francisco Commonwealth/State of California

Sworn to and subscribed before me this 18th day of May, 19 2005. Witness my hand and official Seal.

RUTH R. DIONISIO
COMM. # 1322370
NOTARY PUBLIC - CALIFORNIA
SAN FRANCISCO COUNTY
My Comm. Expires September 27, 2005

CRS 2

(Notary Public/Seal) May 18, 2005

My commission expires Sept. 27, 2005

11/14/95

AFFIDAVIT OF NONDISCLOSURE

Professor
(Job Title)

(Date of Assignment to NCES Project)

University of San Francisco
(Organization, State or local agency or instrumentality)

2130 Fulton St. SF CA 94117-1080
(Organization or agency Address)

NAEP - Reading, Writing, Math, Science 2002-
(NCES Data Base or File Containing Individually Identifiable Information*)

I, Robert Burns, do solemnly swear (or affirm) that when given access to the subject NCES data base or file, I will not -

(i) use or reveal any individually identifiable information [including "schools" in the National Assessment of Educational Progress (NAEP)] furnished, acquired, retrieved or assembled by me or others, under the provisions of Sections 408 and 411 of the National Education Statistics Act of 1994 (20 U.S.C. 9001 et seq.) for any purpose other than statistical purposes specified in the NCES survey, project or contract;

(ii) make any disclosure or publication whereby a sample unit or survey respondent (including "schools" in NAEP) could be identified or the data furnished by or related to any particular person or NAEP school under these sections could be identified; or

(iii) permit anyone other than the individuals authorized by the Commissioner of the National Center for Education Statistics to examine the individual reports.

Robert Burns
(Signature)

[The penalty for unlawful disclosure is a fine of not more than $250,000 (under 18 U.S.C. 3571) or imprisonment for not more than five years (under 18 U.S.C. 3559), or both. The word "swear" should be stricken out when a person elects to affirm the affidavit rather than to swear to it.]

* Request all subsequent followups that may be needed. This form cannot be amended by NCES, so access to databases not listed will require submitting additional notarized Affidavits.

City/County of San Francisco Commonwealth/State of California

Sworn to and subscribed before me this 28th day of May, 2003. Witness my hand and official Seal.

RUTH R. DIONISIO
COMM. # 1322370
NOTARY PUBLIC - CALIFORNIA
SAN FRANCISCO COUNTY
My Comm. Expires September 27, 2005

CRS 2

(Notary Public/Seal) May 28, 2003

My commission expires Sept 27, 2005

11/14/95

AFFIDAVIT OF NONDISCLOSURE

Research Assistant

(Job Title)

University of San Francisco

(Organization, State or local agency or instrumentality)

2130 Fulton St SF CA 94117-1080

(Organization or agency Address)

(Date of Assignment to NCES Project)

NAEP Reading, Writing Math, Science; 2002-

(NCES Data Base or File Containing Individually Identifiable Information*)

I, Yvette Marie Fagan, do solemnly swear (or affirm) that when given access to the subject NCES data base or file, I will not -

(i) use or reveal any individually identifiable information [including "schools" in the National Assessment of Educational Progress (NAEP)] furnished, acquired, retrieved or assembled by me or others, under the provisions of Sections 408 and 411 of the National Education Statistics Act of 1994 (20 U.S.C. 9001 et seq.) for any purpose other than statistical purposes specified in the NCES survey, project or contract;

(ii) make any disclosure or publication whereby a sample unit or survey respondent (including "schools" in NAEP) could be identified or the data furnished by or related to any particular person or NAEP school under these sections could be identified; or

(iii) permit anyone other than the individuals authorized by the Commissioner of the National Center for Education Statistics to examine the individual reports.

Yvette Marie Fagan

(Signature)

[The penalty for unlawful disclosure is a fine of not more than $250,000 (under 18 U.S.C. 3571) or imprisonment for not more than five years (under 18 U.S.C. 3559), or both. The word "swear" should be stricken out when a person elects to affirm the affidavit rather than to swear to it.]

* Request all subsequent followups that may be needed. This form cannot be amended by NCES, so access to databases not listed will require submitting additional notarized Affidavits.

City/County of San Francisco Commonwealth/State of California

Sworn to and subscribed before me this 28TH day of May 19 2003. Witness my hand and official Seal.

RUTH R. DIONISIO
COMM. # 1322370
NOTARY PUBLIC - CALIFORNIA
SAN FRANCISCO COUNTY
My Comm. Expires September 27, 2005

CRS 2

(Notary Public/Seal)

May 28, 2003

My commission expires Sept 27, 2005

11/14/95

Appendix C

An Analysis of the 2005 8th Grade

NAEP Mathematics by Content Strand

Problem Type and Language Category

Appendix Table A.1

An Analysis of the Measurement Strand by Problem Type and Language Categories

Problem Type	Language Category	# of Items	% of Items per Strand
Multiple Choice	CO	1	4
	DM	4	14
	CS	2	7
	OP	2	7
	NG	7	25
	GV	4	14
Constructed Response	CO	0	
	DM	4	14
	CS	0	
	OP	1	4
	NG	1	4
	GV	0	
Extended Response	CO	0	
	DM	1	4
	CS	0	
	OP	0	
	NG	0	
	GV	1	4
		28	15

Note: GV = graphic vocabulary, NG = non graphic vocabulary, OP = operate plan, CS = convert to solve, DM = draw and manipulate, and CO = convert only.

Appendix Table A.2

An Analysis of the Geometry Strand by Problem Type and Language Categories

Problem Type	Language Category	# of Items	% of Items per Strand
Multiple Choice	CO	0	
	DM	6	16
	CS	2	5
	OP	2	5
	NV	4	11
	GV	9	24
Constructed Response	CO	0	
	DM	7	18
	CS	0	
	OP	1	3
	NV	2	5
	GV	4	14
Extended Response	CO	0	
	DM	1	3
	CS	0	
	OP	2	5
	NG	0	
	GV	1	3
		38	20

Note: GV = graphic vocabulary, NG = non graphic vocabulary, OP = operate plan, CS = convert to solve, DM = draw and manipulate, and CO = convert only.

Appendix Table A.3

An Analysis of the Number and Operation Strand by Problem Type and Language Categories

Problem Type	Language Category	# of Items	% of Items per Strand
Multiple Choice	CO	6	6
	DM	1	3
	CS	13	39
	OP	3	6
	NG	7	15
	GV	7	15
Constructed Response	CO	0	6
	DM	3	
	CS	0	6
	OP	5	2
	NG	1	4
	GV	2	
Extended Response	CO	0	
	DM	0	
	CS	1	2
	OP	2	6
	NG	0	
	GV	0	
		48	26

Note: GV = graphic vocabulary, NG = non graphic vocabulary, OP = operate plan, CS = convert to solve, DM = draw and manipulate, and CO = convert only.

Appendix Table A.4

An Analysis of the Algebra Strand by Problem Type and Language Categories

Strand	Problem Type	Language Category	# of Items	% of Items per Strand
Algebra	Multiple Choice	CO	3	7
		DM	0	
		CS	2	5
		OP	8	18
		NG	8	18
		GV	8	18
	Constructed Response	CO	1	2
		DM	1	2
		CS	0	
		OP	4	9
		NG	3	7
		GV	1	
	Extended Response	CO	0	
		DM	0	
		CS	0	
		OP	3	7
		NV	1	2
		GV	1	2
Total strand			44	24

Note: GV = graphic vocabulary, NG = non graphic vocabulary, OP = operate plan, CS = convert to solve, DM = draw and manipulate, and CO = convert only.

Appendix Table A.5

An Analysis of the Data Analysis Strand by Problem Type and Language Categories

Problem Type	Language Category	# of Items	% of Total Items
Multiple Choice	CO	6	29
	DM	0	
	CS	1	3
	OP	2	7
	NG	5	17
	GV	6	21
Constructed Response	CO	0	
	DM	0	
	CS	0	
	OP	0	
	NG	0	
	GV	1	3
Extended Response	CO	0	
	DM	2	7
	CS	2	7
	OP	1	3
	NG	0	
	GV	3	10
		29	16

Note: GV = graphic vocabulary, NG = non graphic vocabulary, OP = operate plan, CS = convert to solve, DM = draw and manipulate, and CO = convert only.