Open-ended (extended/constructed) response questions as predictors of success on subsequent state mathematics examination: the influence of mathematical awareness and conceptual knowledge

Kathy Ann Gullie

University at Albany, State University of New York, kp9854@albany.edu

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Elementary Education Commons, and the Science and Mathematics Education Commons

Recommended Citation

This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive. Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
OPEN-ENDED (EXTENDED/CONSTRUCTED) RESPONSE QUESTIONS AS PREDICTORS OF SUCCESS ON SUBSEQUENT STATE MATHEMATICS EXAMINATION: THE INFLUENCE OF MATHEMATICAL AWARENESS AND CONCEPTUAL KNOWLEDGE

by

Kathy A. Gullie

A Dissertation
Submitted to the University at Albany, State University of New York
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

The Department of Educational Theory and Practice
School of Education
2011
Open-ended (Extended/Constructed) Response Questions as Predictors of Success on Subsequent State Mathematics Examinations: The Influence of Mathematical Awareness and Conceptual Knowledge

by

Kathy A. Gullie

COPYRIGHT 2011
DEDICATION

For Pop

The one who instilled in me an understanding of love, honesty, respect, humor and good manners, and the value of hard work...
ABSTRACT

This study investigated the predictive ability of students’ responses to open-ended, constructed/extended questions in third and fourth grade mathematics content sub-categories on subsequent fifth grade mathematics achievement proficiency levels. Open-ended, extended/constructed response questions reflected content as outlined by the National Council of Teachers of Mathematics (2007) in Statistics and Probability, Number Sense, Numeration and Operations, Geometry, Algebra, and Measurement. Archival data from over 300 third, fourth and fifth grade students in one public school in New York State during the 2003-2009 academic school years were utilized. Predictor variables included Statistics and Probability, Number Sense, Numeration and Operations, Geometry, and Algebra in grades 3 and 4, and Measurement in grade 4.

Results of a statistical discriminant function analysis indicated that, overall, open-ended response question predict fifth grade proficient/non-proficient outcome levels. There was a variation in the predictability of these types of response questions based on grade level and item content specific subscale. At grade three, the variables of statistics and probability and geometry contributed most to overall fifth grade outcomes. In fourth grade, Number Sense, Numeration and Operations as well as Algebra contributed significantly to overall fifth grade achievement. This study adds to current literature investigating the importance of open-ended/constructed response questions and their use in mathematics education. It provides evidence that the constructed response question is an important predictor of and possible contributor to student outcomes on achievement tests in mathematics.
The study has possible implications in several areas. First, it submits that if open-ended response questions predict performance on students’ future mathematics achievement then educators should consider how this type of format can be used to support student mathematical knowledge, including using open-ended constructed response questions for incorporating formative assessment as well as changes in instructional design of mathematics curriculum. Second, future research should consider investigating the extent to which work with extended response type of activities contribute to greater understanding and higher performance in mathematics so that students will acquire mathematical understanding and awareness. Finally, these results contribute to the field of assessment by documenting the value of open-ended responses items on student overall achievement in mathematics.
This dissertation was the journey of a lifetime. The pages of this dissertation hold far more than the culmination of ten years of study. These pages reflect a relationship with the many generous and inspiring people I have met as I traveled through my graduate work. I cherish each person’s contribution to my development as a scholar and acknowledge that I would not have finished without their help and guidance.

To my advisor: Dr. Vicky Kouba, a gracious, thoughtful mentor who, with her warmth and understanding, boosted my confidence by providing me with opportunities to share ideas about mathematics with an equal voice during our working relationship on this dissertation. She demonstrated that rigorous scholarship can be accessible to everyone, even those like me, advanced in age.

To my committee members: Dr. Dianna Newman for her kind support and patience, allowing me to realize that scholars have a responsibility to use the privileges of academia to imagine and create a more knowledgeable world and Dr. Robert Bangert-Drowns for his guidance, encouraging words, thoughtful criticism, and time and attention during his busy semesters. These three awesome guides represent important sides of my education, supporting the development of my personal philosophy, intellectual discipline and clarity and logical consistency.

To all my professors who I have been especially fortunate to know through coursework that encouraged me to question, think, research, and analyze within multiple contexts: Dr. Abbe Herzig, Dr. Jane Agee, Dr. Carol Rogers, Dr. Robert Yagelski, Dr. Judith Langer and Dr. Karen Swan.
To the Department of Educational Theory and Practice staff for assisting me with the administrative tasks necessary for completing my doctoral program: Rosie Renzi, Roberta Hans and Barbara Brunner. I also gratefully acknowledge the institutional support of the University and, in particular, the Evaluation Consortium, which I have received while working through this program.

And finally, to my invaluable network of supportive, forgiving, generous, loving family and friends; to my family, who are my loudest cheerleaders, my children who now can say their mother is a geek, to my fellow graduate students (especially Dr. Wendy Neifeld-Wheeler) who have been an enormous asset and to my consortium colleagues and my friends without whom I could not have survived or finished the process. You are all part of this. Thank you.
TABLE OF CONTENTS

TITLE                      Page
DEDICATION ......................................................... ii
ABSTRACT  ........................................................ iii
ACKNOWLEDGMENTS ......................... .................. vi
TABLE OF CONTENTS  ........................................ viii
LIST OF TABLES AND FIGURES .............................. ix

CHAPTER 1: INTRODUCTION
 Introduction ..................................................... 1
 Purpose .......................................................... 6
 Significance of the study..................................... 7
 Research Question .............................................. 9
 Theoretical Framework ...................................... 10
 Method .......................................................... 13
 Summary ....................................................... 17

CHAPTER 2: REVIEW OF THE LITERATURE
 Assessment and instruction .................................. 18
 Mathematics reform .......................................... 24
 Children’s thinking ........................................... 28
 Conceptual and procedural knowledge .................. 31
 Cognition and mathematics ............................... 35
 Children’s discourse and explanation of solutions ...... 41
 Test design and item format .............................. 47
 Summary ....................................................... 52

CHAPTER 3: METHOD
 Purpose .......................................................... 54
 Sample .......................................................... 55
 Instruments ..................................................... 56
 Research Design .............................................. 61
 Assumptions ................................................... 62
 Data analysis .................................................. 63
 Procedure ..................................................... 64

CHAPTER 4: RESULTS
 Descriptive analysis .......................................... 66
 Analysis 1: Grade 3 ............................................ 67
 Analysis 2: Grade 4 ............................................ 72
 Summary ....................................................... 76

CHAPTER 5: DISCUSSION
 Introduction ..................................................... 78
 Summary of results .......................................... 78
 Implications for theory ...................................... 81
 Implications for practice and research ................. 83
 Conclusion .................................................... 90

REFERENCES ..................................................... 92
# LIST OF TABLES AND FIGURES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Summary of Data Sources from Tests 2003-2009</td>
<td>58</td>
</tr>
<tr>
<td>Table 2</td>
<td>Relationship of Predictors of Success in Mathematics</td>
<td>65</td>
</tr>
<tr>
<td>Table 3</td>
<td>Results of Grade Three Discriminant Function Analysis for Test Predictor Variables</td>
<td>68</td>
</tr>
<tr>
<td>Table 4</td>
<td>Third Grade Means and Standard Deviations for Predictor Variables</td>
<td>69</td>
</tr>
<tr>
<td>Table 5</td>
<td>Grade Three Predicted Classification of Group Membership</td>
<td>71</td>
</tr>
<tr>
<td>Table 6</td>
<td>Results of Grade Four Discriminant Function Analysis for Test Predictor Variables</td>
<td>73</td>
</tr>
<tr>
<td>Table 7</td>
<td>Fourth Grade Means and Standard Deviations for Predictor Variables</td>
<td>74</td>
</tr>
<tr>
<td>Table 8</td>
<td>Grade Four Predicted Classification of Group Membership</td>
<td>75</td>
</tr>
<tr>
<td>Table 9</td>
<td>Comparison Summary of Grade Level Predictor Variables in Grade 3 and 4</td>
<td>77</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Figures</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Function at Group Centroids for Grade 3</td>
<td>70</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Function at Group Centroids for Grade 4</td>
<td>74</td>
</tr>
</tbody>
</table>
Open-ended (Extended/Constructed) Response Questions as Predictors of Success on Subsequent State Mathematics Examinations: The Influence of Mathematical Awareness and Conceptual Knowledge

CHAPTER 1: INTRODUCTION

The United States Department of Education, concerned with students’ achievement levels in mathematics in the United States, is developing multiple initiatives and changing policy to support the improvement of student achievement outcomes. As a result, government policy changes have increased awareness of mathematics achievement problems in our education system and additionally augmented the connection between research and practice. Initiatives including *Race to the Top (2009)* and the *Reauthorization of the Elementary and Secondary Education Act, (2001)* focus on supporting underperforming schools with programs that experts suggest will increase teacher knowledge and improve student test scores, but there is no consensus as to what really works to accomplish this goal.

Additionally, little evidence exists that the initiatives have been working as students still are leaving school with superficial or disconnected mathematics information, lacking a comprehensive conceptual base and achieving below expected levels. In an effort to support a strong educational assessment plan for mathematical conceptual understanding for children at an early age, this research looks at educational testing for mathematics in New York State and provides an analysis of data that could impact the design of testing for the future.
Understanding students’ mathematical knowledge development is at the heart of improvement in mathematics learning and teaching (Bahlmann & Walter, 2006). Assessment provides valuable information that can be used to “promote growth, modify programs, recognize student accomplishments, and improve instruction” (National Council of Teachers of Mathematics, 1995, p. 27). It is commonly argued that if students understand and use solid, conceptually grounded mathematics, then they should pass state-mandated tests (Chauvot, 2006). While many in education may believe this, the curriculum and teaching methodologies used in classrooms have not followed. Despite what we know, the need to improve mathematics scores on standardized tests has not focused on conceptual nor inferential understanding, but instead on “procedural” learning strategies commonly utilized on standard supported tests.

According to Lia and Waltman (2008), many measurement professionals have argued for example, that differential access to test preparation as well as the use of inappropriate practices contribute to preparation that raises students’ observed scores without actually improving their achievement in the broader domain. It is generally agreed that knowledge of concepts and knowledge of procedures are positively correlated when the two are learned in tandem rather than independently (Hiebert & Lefevre, 1986; Rittle-Johnson & Siegler, 1998; Star, 2000). There continues to be disagreement on the order in which children come to acquire these types of knowledge and how knowledge of one affects the acquisition of the other. In a reviews of the literature, Rittle-Johnson and Siegler (1998) and Star (2000) conclude that there is no fixed order in the acquisition of mathematical skills versus concepts; however, Pesek and Kirsner (2000) state that this situation replicates the circumstance they believe characterizes many mathematics classes.
in that most instructional time is spent on routine exercises to consolidate rote or procedural knowledge; much less emphasis is given to students’ intuitive and sense-making capabilities” (p. 538). Mack (1990) found that students’ knowledge of rote procedures frequently interfered with their attempts to build on their informal knowledge and that previously acquired rote procedural knowledge tended to focus on symbolic manipulations and did not seem to consider the validity of responses. Pesek and Kirscner, (2000) suggested that the influence of such rote procedural knowledge could be overcome, but only with a great deal of time and effort. As a result of their studies, Pesek and Kirscner, (2000) indicate that initial rote learning of a concept can create interference to later meaningful learning. They put forward three possible mechanisms of interference which create these problems: cognitive interference, attitudinal interference, and metacognitive interference. They claim these issues disrupt developing new competencies, even if only subconsciously (p. 537).

Multiple stakeholders including policy makers, school districts, and educators are focused on students in the United States not achieving at expected levels in mathematics. For example, Jacobs (2010) discussed a study by William Schmidt, Teacher Education and Development Study in Mathematics (Center for Research in Mathematics and Science Education, 2010) in the United States, which found that prospective U.S. elementary and middle-school math teachers are not as prepared as those from other countries. "A weak K-12 mathematics curriculum in the U.S., taught by teachers with an inadequate mathematics background, produces high school graduates who are at a disadvantage. When some of these students become future teachers and are not given a strong background in mathematics during teacher preparation, the cycle continues” (p. 2).
As a result, the present educational situation repeats and allows students to leave school with a set of facts, procedures, or formulas that are understood in a shallow or disconnected way, lacking in a supportive conceptual base. Researchers suggest that problems of underachievement and failure begin long before students reach secondary grade levels (Carnine, Jones, & Dixon, 1994, p. 36) and believe a limited conceptual base can prevent children from thinking critically and drawing from previous knowledge as well as relevant skills to solve problems they encounter in mathematics (O’Neil & Brown, 1998, p. 332).

Providing solutions for these interference mechanisms are time consuming and unnecessary if conceptual understanding is taught first or underpins the procedural practice. According to Kazemi (2002) results from the National Assessment of Educational Progress (NAEP) have repeatedly shown that students have difficulty with non-routine problems that require them to analyze problems, not just solve them. Alternately, the uses of open-ended response tasks can help solve these difficulties by providing teachers opportunities to consider how students solve problems, develop understanding of those concepts addressed by alternative strategies, and assess whether or not students can flexibly use their understanding in multiple contexts. Additionally, the use of open-ended response tasks and analysis of alternative strategies allow teachers to gain greater insight into students’ thinking and provide opportunities for students to learn as they exercise personal agency (Bahlmann & Walter, 2006), thereby making the student’s process transparent.

According to Kazemi (2002), assessment can be a source of insight into student learning, but it requires that we pay close attention to children’s thinking (p. 221). Davis
and Forster (2003), through an analysis of student responses to open-ended response questions, suggest that there is increased awareness worldwide that ability to communicate mathematics effectively is an important outcome of mathematics education and according to social constructivist theory, “…mathematical understanding develops through communicating ideas with others…” (Cobb, Boufi, McClain, & Whitenack, 1997; p. 36).

Assessments can provide transparency in demonstrating student thought processes and, as researchers suggest, there exists a need for more complex forms of assessments to better gauge the complexity of student thinking (Glaser & Silver, 1994; Resnick & Resnick, 1992). A discussion of open-ended, constructed/extended response question tasks as a useful tools for ensuring a more complex form of assessment has been presented by researchers describing the value of different type of test item formats and the importance of designing assessments that will help improve student understanding and achievement. Policy makers presently ignore much of the possible interpretation and uses of the information they already collect, thus the overarching goal of this study is to raise important questions for policy makers around educational testing and to argue for better test item formatting and uses of test results focusing on analysis for use in the support of early learning in children’s mathematics.

The sample used for this investigation focuses on students from an elementary school identified by the 2009 New York State Report Card as an elementary level school in an urban or suburban school district with high student needs in relation to district resources. This particular sample was chosen because the school is considered a central school district made up of both a large urban group with specific demographic needs as
well as a substantial number of students from a more affluent suburban area with alternative issues. The elementary school serves students in grades Kindergarten through five, is considered a school in good standing and has a 90% stability rate (i.e. how many students remain in a district or school throughout the school year). It is defined as school which meets the needs of 46% of students receiving free and or reduced lunch, indicating an economic needs demographic (New York State Report Card, 2009). The student racial and ethnic makeup is composed of 16%, Black or African American, 8% Hispanic or Latino, 68% White and 8% Multiracial. One hundred percent of teachers in the school are identified as Highly Qualified, certified to teach in the subject area, and show subject matter competency (New York State Report Card, 2009).

**Purpose**

The purpose of this study is to broaden the growing body of knowledge pertaining to student achievement in mathematics. It investigated the degree to which student responses on open-ended, extended/constructed response questions on content specific sub categories of New York State mathematics testing items in third and fourth grade predict success on students’ subsequent proficient/ non-proficient levels on overall fifth grade mathematics achievement test results. Open-ended, extended/constructed response questions reflected content as outlined by the National Council of Teachers of Mathematics (2007) in statistics and probability, number sense, numeration and operations, geometry, algebra, and measurement. Third grade predictor variables included statistics and probability, number sense, numeration and operations, geometry, and algebra. Fourth grade predictor variables included statistics and probability, number sense, numeration and operations, geometry, algebra and measurement.
Significance of the Study

Tests are traditionally developed with various types of items including multiple choice and open-ended or extended, constructed response questions. A study by Kazemi (2002) suggests that when students answer simple or complex specific multiple-choice questions rather than conceptual, extended response questions, their attention is often drawn to the choices themselves rather than to the specific problem or question. They are not forced to think through the problem first and, thus, their choices are often based on an incorrect generalization they have made about problem-solving. For example, Kazemi (2002) stated that it became apparent, through their analysis of the whole class, that the first-graders (they talked with) were ready to be challenged and move their problem solving approaches beyond their fundamental strategies, but also noted through compiling interview data, that students were very comfortable with a specific kind or content of word problem and struggled with others (p. 205).

A careful review of student strategies on open-response tasks warrants a new interpretation of the open-response task as an assessment tool. Open-ended items offer opportunities for students to demonstrate their mathematical thinking, reasoning processes, problem-solving, and communication skills. Due to the fact that open-ended items induce more cognitive strategies (O’Neil & Brown, 1998) and invite a wider range of solutions and solution methods than more traditional assessment items such as multiple choice questions, open-ended items are more effective at revealing students' conceptual understanding in mathematics (Sanchez & Ice, 2004). Additionally, multiple choice formats were not used as predictor variables in this study, as factors complicate there use and interpretation such as the probability of a random correct response (Kuechler &
Simkin, 2010), a focus on procedure and de-emphasis of problem-solving skills (Kazemi, 2002), and reduction in support for conceptual knowledge development.

The outcomes of this study will permit researchers and practitioners to look further into understanding the nature of early arithmetic development and the close, iterative relationship between conceptual and procedural development in children’s mathematical cognition (Canobi & Bethune, 2008).

An additional goal of this research is to provide an alternative framework for the role of conceptual and procedural development in mathematics, particularly in the areas that influence teachers' instructional decisions and their design of curriculum and strategies that support children's long-term mathematical knowledge. Changes or reframing the uses of assessment items can promote changes in instruction (Kuhm, 1993). Teachers are the core agents of any changes that are successfully implemented in the classroom; therefore, movements in education regarding redesigning curriculum as the key to change must be fully supported by teachers in the process of reform, as well as through the implementation and evaluation phase. The connection between student achievement, use of assessment results, and changes in teacher practice fosters learning of worthwhile academic content for all students. School communities as a whole can use assessment results in a formative way to determine how well they are meeting instructional goals and how to alter curriculum and instruction so that goals can be better met (Wolf, Bixby, Glenn, & Gardner, 1991). Focusing on the relationship between student achievement, uses of assessment results and changes in teacher practice, the review of literature for this research investigates the issues behind how achievement information can help teachers support children’s’ conceptual learning in mathematics and
introduces an architecture of possible changes that might theoretically create new meaning appropriate to the topic.

The foundation for this research begins with the perceived relationship between student achievement and uses of achievement tests scores earned on multiple subscales of open-ended response questions in grades three and four and it continues with an investigation of the ability of those scores to forecast student proficiency in total mathematics achievement in fifth grade. Although the focus of this research is the practical applications of mathematical learning such as curriculum design, the research specifically addresses the problems educators confront when designing curriculum and discussing children’s mathematics conceptual understandings. “The need to raise proficiency and understanding in mathematics has never been greater or more important, children must develop an understanding of mathematics that will empower them to solve problems effectively, make reasoned judgments and be reflective in their thoughts” (NCTM, 2007, p. 1). If researchers believe that the strong conceptual understanding exhibited through reasoned responses to open-ended questions develops knowledgeable learners in mathematics, then the support of conceptual learning in our curriculum should develop proficient mathematicians while promoting achievement and proficiency on mathematics tests. This creates a much needed link between assessment, instruction, curriculum design in Elementary Mathematics, and the stakeholders representing these areas.

**The Research Question**

Based on the above information, this study seeks to address the following research questions: To what extent does third grade and fourth grade student performance
on five categories of New York State Mathematics Testing Items in third and fourth grade predict success on students’ subsequent overall proficient/non-proficient levels on fifth grade mathematics achievement test results? Are there specific areas of content or patterns of specific test subscales at each grade level that predict mathematics proficiency?

Theoretical Framework

This research draws on several theoretical and conceptual orientations that focus on understanding knowledge and practice in mathematics. Duebel (2003) suggests that the individual cognitive trends we envision today and practice in our classrooms, which are derived from Piaget's theory, “emphasize the constructivist activity of individuals as they try to make sense of the world” and the “socio-cultural trend emphasizes the socially and culturally situated context of cognition” as expressed by Vygotsky (Duffy & Cunningham, 1996 as cited in Deubel, 2003, p. 67). This focused movement underscores conceptual understanding and the active role of students in their own learning. The present research is designed with the assumption that individual student’s learning, cognition and knowledge, as well as on classroom culture and practices in mathematics, is situated and socially sensitive assigning socio-cultural theory and situated meta-cognition as the overarching theoretical position of the study.

According to Schmittau (2004), theories connected with constructivism must be credited with having moving the field of mathematics education in the United States in a positive, direction. The learning of mathematics is considered an active construction of meanings by making sense of mathematics problem situations, combined with the development of inquiry, reasoning, explanation, justification, argumentation, and
intellectual autonomy. Within socio-cultural theory, which emphasizes human intelligence within our society or culture, individual cognitive gains occur first through interaction with social environment and then a personal internalization of that interaction (Bruner, 1962; Vygotsky, 1962; Hsiao, 2000; Wells, 1999).

The socio-cultural perspective describes language and other semiotic tools as important in the construction of proximal zones for learning, during which socially shared meaning-making rouses new perspectives and possibilities yet to be discovered (Vygotsky, 1962, 1978; Wells, 1999). On the basis of such an idea, the study suggests that effective mathematical classroom discourse is organized so that students learn to explain ideas and solutions to mathematics problems, rather than focusing entirely on whether answers are correct. This socio-cultural perspective allows for the fostering of cognitive growth and building conceptual knowledge. The social nature of the process supports the interaction of students with each other in order to formulate and evaluate questions, solve problems, as well as develop hypotheses, conjectures, and explanations. Students propose and evaluate evidence as well as present examples and argument that lead to understandings based on mathematical reasoning (Olson, & Loucks-Horsley, 2000).

Also, as a function of socio-cultural theory, mathematics exhibits itself through socio-mathematical norms. These norms can be described as intrinsic aspects of a classroom’s mathematical micro-culture, the students’ knowledge present in the situation, and ways of judging “what counts” as an acceptable mathematical explanation (Yackel & Cobb, 1996). This view is associated with ideas of intuition and a priori knowledge as defined by the philosophy of Kant (Kant, 1965; Roth & Hwang, 2006). Additionally,
cultural assumptions in mathematics are connected with certainty; for example, being able to get the correct answer quickly is usually shaped by school experience. Thus mathematics has come to mean following the rules laid down by the teachers and knowing mathematics means remembering and applying the correct rule when the teacher asks questions. Mathematical truth is determined when the answer is ratified by the teacher. Beliefs about how to do mathematics and what it means to know mathematics in school are acquired through years of watching, listening, and practicing (Ball, 1988; Lampert, 1990; Schoenfeld, 1985b; Stodolsky, 1985). Beliefs about mathematics also impact perspectives on all new information as it is introduced and amassed over time.

The construct of situated metacognition is important to this framework as it refers to the awareness that learners have about their general academic strengths and weaknesses, the cognitive resources they can apply to meet the demands of particular situational tasks as well as their knowledge and skill about how to regulate engagement in tasks in order to optimize learning processes and outcomes (Lave & Wagner, 1991; Perry, Van de Kamp, Mercer, & Nordby, 2002; Winne & Perry, 2000). Within pedagogy for critical consciousness (Friere, 1970 as cited in Cliff & Miller, 1997), teachers initiate activities to pose problems and initiate a critical dialogue. Through this problem posing approach to mathematics inquiry, those involved are invited to investigate and reflect in order to develop an in-depth, conceptual understanding of mathematics that exhibits itself as something other than an answer on a test. It is often only with great difficulty that we ascertain the deep (conceptual) structure of an algorithm. It is, however, essential that we do so. Vygotskian psychology holds that the understanding is not complete until the origin and full development of a concept is comprehended (Schmittau, 2004).
Although beliefs about mathematics, in general, impact perspectives on the conceptual understanding of mathematics, the context in which teachers instruct influences the methods teachers use when preparing students for end-of-the-year achievement tests required by most states. Previous studies have shown that testing encourages teachers to narrowly focus the specific content to be covered (Stecher, 2002; Abrams, Pedulla, & Madaus, 2003; Finbarr & Anthony, 2003), increase the amount of time teaching students test-taking skills, and spend less time on conceptual and in-depth instruction of the content (Abrams, et al., 2003; Taylor, Shepard, Kinner, & Rosenthal, 2002).

Standardized tests have changed the pace and content of instruction, where relentless drill practice for students is instilled and instruction focuses on test content or test-taking skills while ignoring subject areas that are not addressed on the test (Ballard & Bates, 2008). As state tests become increasingly integral to the culture of how districts, schools, teachers, and students are evaluated, educators must be cognizant of the way test item format impacts classroom instruction and student performance. With this in mind, it the present research examines the role of open-ended response test items in predicting outcomes of classroom instruction for elementary students.

**Method**

The present research utilizes data from one school in one upstate New York State school district. The data encompasses longitudinal archival test result for approximately 300 elementary students who were enrolled in grades 3, 4 and 5 from 2003 -2009. Participation was voluntary for the school district and the school, but was not for students. Assent was part of a statewide program and thus participation by students was
required and not voluntary. To address Institutional Review Board (IRB) data sets did not use identification respondent information. Collection of final test results was assembled after approval of the dissertation proposal.

**Sample**

As previously noted, the sample for this study is made up of approximately 300 students in grade three through 5, from a school identified in the New York State Report Card (2009) as an elementary level school in an urban or suburban school district with high student needs in relation to district resources. Comparable schools in this group are in the middle range of student needs for elementary level schools in these districts.

**Instrumentation**

This study used components of two assessment tools to assess the dependent and independent variables based on mathematics scaled scores obtained from the administration of either the Test of New York State Standards (TONYSS; created by Riverside Publishing Company, 2006) and the New York State Testing Program, Mathematics Tests from 2003 -2009.

**Research Design**

A correlational statistical discriminant analysis was employed to assess the ability of the independent test variables in grade three (statistics and probability, number sense, numeration and operations, geometry, and algebra) and grade four (statistics and probability, number sense, numeration and operations, geometry, algebra and measurement) that best discriminate among students in two groups, those who attain a non-proficient grade (Group 1: non-proficient at level 1 or 2) and a proficient grade (Group 2: proficiency at level 3 or 4) a on their mathematics achievement tests in grade 5.
Discriminant analysis permits a sophisticated predictive view (Burns & Burns, 2008) of the test results by providing depth of information based on the analysis of individual student results tracked over the course of three years.

**Analysis**

A statistical discriminant analysis was used to determine if third and fourth grade mathematics extended response subscales were useful in discriminating between proficient and non-proficient fifth grade performance and to what degree specific subscales were useful discriminators. Discriminant analysis is a useful tool for detecting the variables that allow the researcher to differentiate between pre-existing groups with a better than chance accuracy (Starsoft, 2011).

In this study the variables include:

- **Dependent variable:** Students’ categorization into Group 1, non-proficient level (scores of 1 and 2) or Group 2, proficient level (scores of 3 and 4) on the overall grade 5th grade New York State Testing Program Mathematics Test.
- **Independent variables:** Student scores in grades 3 and 4 on Test of New York State Standards (TONYSS) in mathematics and the New York State Testing Program Mathematics Test open-ended, extended/constructed response questions in five subscale areas of statistics and probability; number sense and numeration; geometry; algebra; and measurement.

The choice of this approach is based on both the theoretical and methodological reasons. First, efforts for increasing student achievement must be based on a systematic and plausible procedure for assessing accountability (Walker, Gosz, & Huinker, 2005); however, studies that compare unadjusted average student performances on standardized
tests across successive cohorts (McCaffrey et al., 2003; Walker, et al. 2005) and cohort-to-cohort accountability systems do not take into consideration pre-existing differences between successive classes at the individual level (Walker, 2005). Such designs do not account for individual students’ incoming conceptual knowledge and educational experiences that are certain to impact performance on achievement tests, nor do they consider the growth of students from one year to the next to address previous knowledge and performance and to possibly augment the limited information provided by a one-year snapshot of achievement. This current research chooses to follow the growth and performance of students over time. Second, the discriminant function allows for the development of best linear weighting for predictor variables that will maximize the differences among two or more groups (Pyryt, 2004) and produces discriminant function coefficients for each predicator variable that would indicate the importance of each variable (Smith, 1979).

**Limitations**

This research is limited in generalizability based on the sample size and the use of one specific environment; data collected from one school in one district may not be applicable to other situations. Additionally, the analysis looks at two sets of test data configurations, which similar and aligned, are not exactly the same. These limitations, however, do not preclude utilizing results to guide other studies patterns amongst student achievement in mathematics.
Summary

According to Kamezi (2002), teachers practice an array of discrete mathematical procedures they anticipate will be covered on the end-of-the-year tests. What is particularly striking is that some teachers perceive they must put aside their efforts to elicit children’s thinking and instead perform triage on the skills they have not yet “covered.” Teachers’ anxieties about “coverage” are heightened because of the sheer volume of distinct skills students are expected to master at each grade. Observing teachers turn to hours of computational practice, one wonders instead what a student might learn if students were asked to tell us how they approached seemingly straightforward problems that they encounter on tests (Kamezi, 2002, p. 205).

Additionally, although many of students can manipulate equations, find derivatives, and apply algorithms they fail to comprehend qualitative descriptions of real world everyday problems. Such disparities indicate an educational system that focuses on manipulative skills at the expense of real understanding (National Center for Educational Statistics, 2005).

This research hypothesizes that the early support of student involvement in more extended or open-ended response activities in the school mathematics curriculum would not only provide an stronger avenue for conceptual learning, but would also support the development of student's mathematical awareness and conceptual knowledge, thus resulting in a more knowledgeable student able to perform better on any required test.
CHAPTER 2: REVIEW OF THE LITERATURE

Assessment and Instruction

Understanding students mathematical knowledge development is at the heart of improvement in mathematics learning and teaching (Bahlmann & Walter, 2006) and assessment provides valuable information which can be used to “… promote growth, modify programs, recognize student accomplishments, and improve instruction…” (NCTM, 1995, p. 27). Some research results suggest that high-stakes testing has the potential to substantially improve student learning (Krueger, 2000), but may also lead to some strategic responses on the part of teachers. Educators and policymakers need to understand and account for possible increases in teaching test-specific skills among younger students as some research suggests that tests scores may go up without there being a commensurate gain in learning (Shepard, & Dougherty, 1991).

Originally, traditional standardized tests of achievement were used to report to parents and monitor district and state trends. Early on, according to Goslin (1967), teachers paid little attention to the outcomes of testing and instruction was not affected by such assessment. Measurement-driven instructional practices started to increase as educational reformers intentionally worked toward improving instruction through testing measures (Shepard et al., 1991). The utility of tests to evaluate teachers and give students practice with the type of item formats that appear on the test has become the focus of mathematics education from the 1970’s to the present.

While the architects and supporters of state-wide testing systems clearly mean for these accountability systems to improve instruction and learning, most state testing systems are ill adapted for this purpose Critics (Amrein & Berliner, 2002) argue that state
accountability pressures actually lead school personnel to replace meaningful instruction with a narrow test-prep procedural focused curriculum. Often reported after a school term ends and usually limited to statistical measures, these tests results do very little to inform instruction or to detail specific deficiencies of students’ mathematical thinking (Foster & Noyce, 2004). The key to developing students who know math may lie in realization that if we are to use tests to evaluate student achievement, then the test needs to be looked at more closely. Standards-based instructions, aligned with the NCTM principles, were created with a large influence from the mathematics education community (Latterell, 2005; Schoenfeld, 2004).

Standards-based instruction is thought of as devoting more time to problem solving, reasoning, and communicating thinking in order to construct personal meaningful understandings of mathematical concepts (Baxter, Woodward, & Olson, 2001; Brooks & Brooks, 1993; Jones et al., 2003). In contrast, traditional instructional methods place the teacher at the center of all activities (Jones et al., 2003). Typically, this traditional method of instruction includes a dominance of practices such as lecture, lecture and discussion, direct instruction, and the use of assessments focusing on lower-level knowledge (e.g., multiple-choice questions); (Brooks & Brooks, 1993; Eggen & Kauchak, 2001; Jones et al., 2003; Vogler & Burton, 2010).

A constructivist based reform movement in mathematics education in the United States has been spearheaded by the NCTM whose initial (1989) and revised standards (2000) have become the prototype for the mathematics standards adopted by the various states (Schmittau, 2004). These state standards, in turn, form the basis for curriculum, and accordingly determine what is taught in the classroom (Schmidt, Houang, & Cogan,
2002; Schmittau, 2004) and the passage of “No Child Left Behind Act” ensures that test-based accountability continues to be a pervasive force in elementary and secondary education. It also becomes less clear how to evaluate high-stakes testing as a school reform strategy based on test results because the achievement gains are driven largely by increases in skills emphasized on exams, particularly among younger students (Jacob, 2005). According to Jacob (2005), an assessment of policy “depends largely on how one values these skills and how much one believes that there has been a decrease in other skills that are not assessed on standardized achievement exams…” (p. 782).

Additionally, "Data-Driven Curriculum Reform" as explained by Nancy Love (1999) is "defining a desired practice is important both for building teachers’ understanding and commitment to the change and for assessing progress” (p. 1). A key premise to reform before facilitators of change can determine their next intervention and before judgments can be drawn about student outcomes is that there must be systematic information about implementation and use of the innovation in classrooms (Hall et al., 1999).

An item-level analysis by Krueger (2000) indicated that the improvements on the high-stakes tests were driven largely by an increase in specific skills (i.e., computation skills) in the case of mathematics. Additionally, Jacob’s research indicated that at the elementary level no improvement on questions involving estimating for problem solving nor the effective use of various strategies to solve word-problems, and very little (if any) improvement on items involving multiple-step word problems, measurement, and interpreting relationships shown in charts, graphs, or tables (Jacob, 2004). Despite the
growing importance of high-stakes testing, there is limited empirical evidence on its veridical impact or value for improvement in student learning.

Traditional theories of learning that focus on knowledge reproduction rather than knowledge understanding and use are inadequate when learning approaches advanced knowledge domains where tasks become more complex, dynamic, and ill-structured. According to Jakworski (2007), acknowledging the importance of communication and interaction and their resulting reciprocal discourses to growth of knowledge have appended the word “social” to constructivist, emphasizing the importance of language and discourse in communicative environments through which learning occurs. Discussion, negotiation, and argumentation in inquiry and investigation practices underpin knowledge growth in teaching and educational situations (e.g., Cobb & Bowers, 1999; Jaworski, 2007; Lampert, 1998; Wood, 1999).

According to the NCTM (2007), "traditional curricula tend to rely on direct explication of the to-be-learned material as well as careful sequencing and the accumulation of lower-level skills before presenting students with the opportunity to engage in higher-order thinking, reasoning, and problem solving with those skills." (p. 1). Hiebert (1999) outlines the traditional approach to solving problems in U.S. classrooms as an introduction of a procedure and then assignment of problems on which students are to practice the procedure. Problems are viewed as applications of already learned procedures. This method of instruction allows instructors to cover a greater breadth of material in a shorter period of time (Thomas, Mergendoller, & Michaelson, 1999); thus the focus is procedural and explicit.
In contrast, standards-based materials rarely explicate concepts for students; rather, they rely on students’ interaction with reformed designed tasks to expose them to the concepts and related skills. During this alternative process, after a concept has been introduced and its features explored by students through activity and discourse, the curriculum and teacher serve to apply definitions, standard labels, and standard procedural techniques (NCTM, 2007).

Examinations of traditional mathematics curricula show that the development of instructional approaches used in today’s classrooms are based on the construction of mathematics as a discipline rather than based on child development and ways of knowing mathematics. Traditional mathematics, according to Draper and Siebert (2004), is the act of selecting the appropriate skill from one's mathematical toolbox and performing a series of computations or symbol manipulations to produce the same answer as the teacher or from the back of the textbook. Quoting facts, correctly performing computations, memorization and skill practice are of utmost importance in learning mathematics in this process. The authors also say that proponents of this “instrumentalist view of mathematics” (p. 930) give little attention to understanding the important mathematical concepts that underlie facts and procedures. The assumption is that the procedures are the concepts, or that the concepts will become self-apparent once enough facts and procedures have been mastered (Draper & Siebert, 2004). This suggestion that concepts will become self-apparent is the misconception that many teachers cling to when they develop their plans and design their curriculum as research indicates through an analyses of key pedagogical features of the lesson materials, that tasks did involved hands-on activities or real-world contexts and technology but rarely required students to provide
explanations or demonstrate mathematical reasoning. (Silver, Mesa, Morris, Star, & Benken, 2009).

In a review of discussion practices in schools, Boaler (2000) described how Amber Hill School's practices of school mathematics that included interpreting cues, seeking structured help, and memorizing school procedures which influenced students' cognitive forms and impacted their broader identities as procedural learners. The discussions adversely affected student knowledge production within and outside school. In contrast to the compared differentiated, project based mathematics at Phoenix Park School; "The teachers at Amber Hill explained mathematical methods clearly and the students received opportunities to practice these methods. The students were confident in their use of school methods in the classroom and it is inconceivable to think that they understood none of these methods, yet the students suggested that they were unable to use any of their school mathematical methods in real world situations" (Boaler, 2000, p. 8). When given open-ended tasks, Phoenix students outperformed students in Amber Hill because they were willing to derive meaning from the problem, and were able to select an appropriate procedure or adapt one to fit a new situation. At Amber Hill, the program emphasized individual workbooks and textbooks and classrooms were characterized by a search for correct answers rather than understanding, competition, individual work, and the transmission of algorithms and procedures. The knowledge of Amber Hill students was inert and they could not apply their knowledge. Boaler (2002) concluded that in contrast, Phoenix students learned how to use their knowledge and were able to select appropriate procedures or adapt to fit new situations.
Furthermore, the reasons students gave for the disparate performance was not related to inadequate understanding, but to the differences in the constraints and affordances provided by the two environments. Students’ confusion and inability to think mathematically in certain situations stemmed from the procedural forms of knowledge they had developed that were inferior to the more conceptual and flexible forms of knowledge that their counterparts at Phoenix Park School had developed (Boaler, 1998, p.7).

If we, as educators, are to understand more about students’ use of mathematical knowledge, we need to consider extending our focus beyond the concepts and procedures that students learn to practice in classrooms and enable educators to consider the macro context within which students engage mathematics. As outlined by Zevenbergen (1996) this paradigm shift would mean a revision of the patterns of work in the classroom and possibly a substantial change in the teacher perceptions of the "naturalness" of cognitive development within the situations perceptions are produced (p 108- 109). “When teachers explain mathematical methods clearly to a group of students, or allow students to construct an understanding of mathematics in other ways, students often engage in a process of learning mathematics, but they simultaneously learn about a set of practices that is school mathematics” (Boaler, 2000, p. 7).

Mathematics Reform

Determining the effect of the any type of program on student achievement is no easy feat. Within a school, it is ultimately the classroom teacher that has the most direct role in impacting student achievement and ideally, if we were to look at what every individual student is saying, we might be able to come away with a clear picture of what
the mathematics reform moment is doing to the environment and activity within a classroom. In reality this is not possible and thus it leaves researchers and practitioners with a substantial gap of understanding.

According to Ross, McDougall and Hoagaboam-Gray (2002), the chief characteristics of math education reform that emerged from their review of mathematics reform studies and NCTM policy statements (1989, 1991, 2007) include a broader scope implementing less commonly taught constructs; equal access to all forms of mathematics; complex tasks; open-ended problems embedded in real-life contexts with multiple solutions; and a focus on the construction of mathematical ideas through students' discourse. In such a reform setting the teacher acts as a co-learner and creator of a mathematical community; mathematical problems are undertaken with the aid of manipulatives and ready access to mathematical tools; the classroom is organized to encourage student-student interaction; assessment is authentic in nature; mathematics is considered dynamic discipline rather than a fixed body of knowledge; and the development of student self-confidence in mathematics discourse is not as important as achievement (Ross, McDougall, & Hoagaboam-Gray, 2002).

Ross, McDougall, and Hoagaboam-Gray’s (2002) analysis supports the NCTM (2007) contention that talk is important when designing curriculum where studies showed that when faced with multi-step problems, students frequently attempted to solve the problems by randomly combining numbers instead of implementing a solution strategy step by step. "The process of encouraging students to verbalize their thinking—by talking, writing, or drawing the steps they used in solving a problem— is consistently effective" (p. 4). In part, this procedure may be effective because the impulsive approach
to solving problems taken by many students with mathematics difficulties was addressed. According to Ross et al., (2002) the research results of these students were impressive, with an average effect size of 0.98. Teachers had students practice verbalizing a solution and good deal of time went into how to solve, for example, the different types of subtraction problems by using part-whole relationships. This verbalization appeared to help anchor the students both behaviorally and mathematically. Schoenfeld (2002), in a review of test results in Pittsburgh, Pennsylvania, illustrated how student scores on concepts and problem solving increased with the implementation of the new curriculum and continued to rise as teachers became increasingly familiar with this new curriculum. In 1997, roughly 10% of the students taught through traditional methods met or exceeded the standards for concepts or problem solving; in 2000 roughly 25% of Pittsburgh’s (now reform) students met or exceeded those standards. Schoenfeld (2002) also notes that problem solving was not a focus of the traditional curriculum. However, it was in the area of skills that the most surprising data emerge, as "traditional measures of skills such as the Iowa Test of Basic Skills (ITBS) show that the reform curricula compared favorably with traditional curricula with regard to skills" (p.16).

Trafton, Reys, and Wasman (2001) advise that the emphasis in mathematics’ reform is on intellectual engagement; that is, student learning has to go beyond the learning of specific concepts and skills and should include the development of a mathematical disposition, a flexibility in exploring mathematical ideas, a perseverance in working on mathematical tasks, including an interest as well as an inventive nature in doing mathematics. Students need to reflect on their thinking and value mathematics and its role in our culture. It is through inquiry, reflection, and discourse about children’s’
thinking that will allow educators to have a better understanding of student's mathematical understandings. The question remains, then, how do we make sure students have these “reform” opportunities and what should they look like?

According to a number of reports including Ross, McDougal, and Grey (2002) reform in mathematics education contributes to higher student achievement (p. 131). Senk and Thompson (2003) and Newman (2007), suggested that students who used standards-based curricula generally perform as well as other students on traditional measures of mathematics achievement and do better on assessments of conceptual understanding and ability to use mathematics to solve problems. In their final report for a national Mathematics Science Partnerships (MSP) grant, Newman, (2007) completed longitudinal cohort-based analyses that traced students across grades, teachers, and instructional modes which showed statistically significant improvements on the percent of students passing New York State exams as well as district sponsored tests across all grade levels for students of teachers involved in the professional development. According to researchers, teachers’ uses of more complex strategies increased over the course of the MSP grant and there was a decrease in the use of traditional strategies. An increase in blended instruction and student-centered or student-directed approaches was also evidenced (Newman, 2007).

These study results are consistent with the strategies for learning described by Loucks-Horsley and others (1996; 1998) which outlined four important characteristics of activity; content-specific material, inquiry-based learning, collaborative grouping, and establishing learning communities. Empowering teachers with knowledge and skills to create an effective classroom environment facilitates the transfer of learning and in this

Trafton, Reys, and Wasman (2001) state that a primary concern in all curriculum reform is the inclusion of knowledge, understandings, processes, and skills that constitute competency in a field. Most long-time observers of U.S. mathematics programs have been struck by a traditionally narrow focus on skills and procedures in mathematics education. A change in that focus comes with standards-based materials and an emphasis on connections and links to related mathematical ideas and applications through problems and tasks. Additionally, Semper (1996) asserts that the reform standards emphasizing the practice of inquiry, is a way of developing thinking, processing, and operating in the world. The key is having an initial curiosity about something and a framework to ask questions.

The need for a personal interest from students in what is being examined is why inquiry cannot be taught as a process skill, demonstrating that the “standards-based curricula are working in classrooms in ways their designers intended for them to work” (Kilpatrick, 2003, p. 472). It should also be noted, however, that these studies did not satisfy skeptics who demand that independent researchers examine students’ learning using more rigorous methodologies (Senk, 2003).

Children’s Thinking

In the research literature on the nature of ‘thinking’ and how it might be taught, the terms ‘critical thinking’ and “cognition” are used to describe competencies which seem to be applicable to teaching mathematics in context. The term ‘critical thinking’ is used in a select body of research to describe reasonable, reflective thinking, focused on a
task, people or belief (Ennis, 1993, p.180). Pithers and Soden (2000) support other researchers (Facione, Sánchez, Giancarlo, Facione, & Gainen, 1995), using the construct of critical thinking as a disposition, which refers to the necessary attitudes that enable the thinker to apply particular skills when a situation calls for their use under the umbrella of metacognition. The connection between critical thinking skills and critical thinking dispositions can be traced back to the Greeks, whose emphasis on habits of the mind held parallel importance with physical, moral, and social habits (Facione, 1997).

The importance of the idea that intellectual character of students’ willingness and intention to think and learn nurtured simultaneously as they are taught skills and information should not to be underestimated (Ruff, 2005). Ennis’s (1975) view of critical thinking involves broad dispositions, transferable over various domains such as being ‘open-minded’, ‘drawing unwarranted assumptions cautiously’ and ‘weighing the credibility of evidence’ (Pithers & Soden, 2000). These abilities and dispositions occur within a global perspective in which thinking is conceptualised as a type of reasoned argument with an explicitly social dimension (Kuhn, 1991).

The concept of children’s thinking has a relationship to collaboration and cooperation in mathematics education. For example, Gokhale (1995) makes this connection when she states that a collaborative learning medium in mathematics provided students with opportunities to analyze, synthesize, and evaluate ideas cooperatively. The informal setting of the study facilitated discussion and interaction (Gokhale, 1995). After conducting a statistical analysis on test scores, she found that students who participated in collaborative learning had performed significantly better on the critical thinking test than students who studied individually, while both groups did equally well on the drill and
practice test. Though a small study on college students, these results are in agreement with the learning theories proposed by other researchers who are proponents of collaborative learning (Johnson, Johnson & Smith, 1991; Panitz, 1997). Implications of Gokhale’s research are described as support for collaborative learning which fosters the development of critical thinking through discussion, clarification of ideas, and evaluation of others' ideas. Since both collaborative and traditional methods of instruction were found to be equally effective in gaining factual knowledge, collaborative learning is more beneficial for enhancing critical thinking and problem-solving skills.

A four-year longitudinal study by Fennema et al. (1996) provides evidence that knowledge of children's thinking is a powerful tool that enables teachers to transform this information and use it to formatively change instruction. The findings coalesce into a convincing argument that an effective way to improve mathematics instruction and learning is to help teachers understand the mathematical thought processes of their students. Additionally, the results suggest that there is increasing evidence that knowledge of children’s thinking has a powerful influence on teachers as they consider instructional change (Fennema, et al., 1996).

As teaching demands that educators establish routines to guide the basic physical and social interactions of the classroom, thinking routines need to be established to help guide students’ learning and intellectual interactions. (Leinhardt & Greeno, 1986; Leinhardt, Weidman, & Hammond, 1987). Thinking routines provide the structures through which students collectively, as well as individually, initiate, explore, discuss, document, and manage their thinking in classrooms (Ritchhart, Palmer, Church, & Tishman, 2006, p. 1).
Many studies show that an open and creative state of consciousness can be induced in the short term but do not tell us as much about the development of mindfulness as a trait (Ritchhart & Perkins, 2000). In their study *Thinking Routines: Establishing Patterns of Thinking in the Classroom*, Ritchhart et al. (2006) investigated teachers’ use of routines, their adaptations, refinements and selection, situating thinking routines within a larger context. The idea was to develop thoughtful classrooms and nurture students’ thinking dispositions, providing an analysis of the key structural and epistemological characteristics of thinking routines. Their results suggested that through the epistemic messages embedded in the routines; both students and teachers come to view and approach thinking and learning differently. As a result of using routines teachers were able to lead students beyond superficial responses into more thoughtful explanations. For example, a third grade teacher used routines to get a better sense of her students’ understandings and to consider the shape of future lessons. “As students develop models of and language for thinking it is demystified and made visible. We have seen that thinking routines are more than strategies that cultivate students’ ability or that simply engage them in interesting activities” (p.39); they provide a diagnostic situation, revealing students’ conceptions or misconceptions and ideas that impact future instructional direction (Ritchhart et al., 2006, p. 41).

**Conceptual and Procedural Knowledge**

“Recognizing the potential problems caused by procedural teaching and learning, the National Council of Teachers of Mathematics (NCTM) and many researchers have urged that conceptual understanding be developed before computational fluency (Bezuk

To improve long-term conceptual learning, Gray and Tall (2002) suggested that the whole curriculum must be framed with an awareness of the abstraction process to produce thinkable concepts. Support for this position can be found in a recent research brief created by NCTM (2007), which suggested that if students understand their arithmetic such that they are able to explain and justify the properties they are using as they carry out calculations. As such, students will have learned some fundamental foundations of algebra and will be able to express general, algebraic properties about the number system without the necessity of using traditional algebraic functions and notations used at higher grade levels (Carpenter, Franke, & Levi 2003). This finding indicates that though students do not use algebraic notation in their answers, they are still able to be aware of and express general algebraic properties about the number system, which lays the foundation for later success in more complex mathematics including high school algebra (NCTM, 2007).

Similarly, the current body of research related to the development of algebraic reasoning at the elementary school level emphasizes that arithmetic can be conceptualized in algebraic ways and that building an understanding of algebra begins within the learning of arithmetic. The research also describes ways in which this emphasis can be capitalized on to encourage young students to make algebraic generalizations without necessarily using algebraic notation. These studies point to promising avenues for developing the conceptual underpinnings of students’ later activities in algebra (NCTM, 2007).
Using algebra with elementary children has been the focus of much review (Grandau, 2005; Mitchelmore & Prescott, 2006; Rogers, Kibler, Wenger, & Truxaw, 1989). Research supports the contention that it is not the content, but rather the process of conceptualization in the abstract form that allows students access to better mathematical thinking and reasoning. In his research, Ennis (1971) discussed that primary grade children vary greatly in their degree of competence in conditional logic, and the principles differed considerably among themselves in their degree of difficulty. This line of reasoning is also supported in a study by Mulligan, Mitchelmore, and Prescott (2006), where children during a simple counting task of multiples of two, were able to count aloud using the pattern correctly but could not show the corresponding pattern in units partitioned on a number line. Similarly, partitioning and visualizing in equal-sized units proved to be difficult across a range of tasks. Children who had an advanced awareness of pattern and structure excelled across most conceptual areas and showed strong indications of early algebraic reasoning (Mulligan, Mitchelmore, & Prescott, 2006). “The entire seamless process of successive problem-solving evolves from the meaning of multiplication and division and the properties that these actions possess, all the way to the standard algorithms for multi-digit multiplication and long division. Thus, the procedural is rendered conceptual and the conceptual becomes procedural; it is impossible to say where one ends and the other begins” (Schmittau, 2004, p. 31)

Spiro and Jehng (1990) consider advanced knowledge acquisition in an ill-structured domain a fertile area in which to pursue constructivist conceptual theories. Traditional learning designs fail where conceptual complexity is a dominant paradigm. It is common for
students not to develop the central concepts of advanced knowledge acquisition, perhaps due to oversimplification or reductive bias (Spiro, 1991) resulting from treating advanced (ill-structured) learning domains as though they were introductory (and well-structured), thus removing complexity and developing an inclination to learn things as simpler or easier than they really are (Wilson, Jonassen, & Cole, 1993). According to Spiro, (1991), errors introduced through reductive bias can compound each other when learners attempt to reassemble the "big picture" (p. 26). In well-structured domains, it is often desirable to focus on general principles. Using this strategy "... leads to misunderstandings in ill-structured domains, where across-case variability and case-sensitive interaction of principles vitiates their force" (Spiro, 1991, p. 27). Spiro and his colleagues propose cognitive flexibility as a remedy for this situation. Flexibility in thinking can be supported through open-ended items which offer opportunities for students to demonstrate their mathematical thinking, reasoning processes, and problem-solving and communication skills. Due to open-ended items inviting a wider range of solutions and solution methods than more traditional assessment items, open-ended items are better at revealing students' understanding of mathematics (Sanchez & Ice, 2004), which can inform effective instruction of mathematics.

Additionally, the research of Schmittau (2004) focuses on the constructivist framework described by Davydov's curriculum and suggests that genetic analysis reveals that preliminary work on quantitative relations (i.e., relationships between quantities) must precede the development of the concept of number. According to Davydov, children should not encounter number until well into the second semester of the first grade.
“The foundational groundwork for this concept consists of increasingly refined activities of comparison and measurement, with major emphasis on continuous quantity. This represents another departure from the United States elementary curriculum, which begins with number and emphasizes discrete quantity almost exclusively. In addition, number in Davydov’s curriculum is developed from measurement rather than counting, as this constitutes a more adequate basis for the development of the real number system and the operations of addition, subtraction, multiplication, and division” (Davydov, 1975a; Schmittau, 1994; 2003b; in Schmittau, 2004).

That is, present educators talk about numbers as a quantity, but fail to make sure that children understanding the ideas of “more” and “less” and the relationships of number quantity important in understanding the idea of number in-depth before they have to use them in algorithms.

**Cognition and Mathematics**

Individuals often take for granted the thinking skills that they employ when solving problems related to everyday tasks. It is not until a person’s thinking is blocked or misconceptions lead one to the wrong answers, that one becomes aware of the importance of thinking skills. As a learner approaches the advanced knowledge acquisition stage in mathematics, learning theories dominating introductory acquisition domains are inadequate because skills obtained in this stage (advanced knowledge acquisition stage) of learning are limited in their transferability (Jonassen, 1991; McGrath, 2001; Spiro et al., 1988). In order to avoid "reproductive thinking" (i.e., specific rules and routines that students develop early on in their formal education)
learners need to develop a cognitive flexibility, and fluid representations of the domain that reflect the dynamics of the real world in which they will use the knowledge (Wertheimer, 1945, as cited in Anderson, 2000). Spiro and his colleagues (1991) developed the Cognitive Flexibility Theory (CFT) in the early 1990’s. This constructivist instructional theory was originally designed for ill-structured learning situations and complex knowledge areas. By “cognitive flexibility” Spiro (1991) means: “… the ability to spontaneously restructure one's knowledge, in many ways, in adaptive response to radically changing situational demands (both within and across knowledge application situations)” (p. 165). The key features of constructivism in cognitive flexibility theory are developing an understanding beyond the presented information, requiring not the retrieval of prepackaged schemas, but rather the creation of new structures drawn from a diverse array of existing mental representations and bringing together the appropriate sources, one of which is flexibly using preexisting knowledge in order to construct new knowledge (Spiro et al., 1991).

A number of the studies refer to cognitive flexibility centered on the use of technology specifically hypertext (Grabinger, 1996; Spiro, Feltovich, Jacobson, & Coulson, 1991; Wilson, Jonassen, & Cole, 1993); yet, these research studies can be applicable to definitions centered not on technology and media, but on a general hypothesis that cognitive flexibility allows for authentic and realistic experiences where learners can construct their own meaning through multiple perspectives and approaches to solving mathematical problems. This increased repertoire and adaptability of thinking perspectives serves for teachers, as a stable platform for understanding, when encountering new mathematical ideas and situations, new students, new materials, and
new tools, so that expertise is generative (Franke, Carpenter, Levi, & Fennema, 2001) and not a matter of mere routine. The need for the ability to think flexibly is documented in a number of studies that focus on students' development of misconceptions and misunderstandings related to mathematical constructs.

The study *Upper Elementary School Pupils’ Difficulties in Modeling and Solving Nonstandard Additive Word Problems Involving Ordinal Numbers*, conducted by Verschaffel, De Corte, and Vierstraete (1999) revealed that “Extensive experience with traditional arithmetic word problems induces in pupils, a strong tendency to approach problem solving in a mindless, superficial, routine-based way in their attempts to identify *the correct* arithmetic operation needed to solve a word problem” (p. 265). Additionally, the researchers claim that students do not critically consider the possible problematic modeling assumptions underlying their proposed solutions in relation to the context involved in the word problem.

This mindless, superficial, routine-based mathematics approach is a consistent concern in research that suggests that students need to be cognitively involved in the process of mathematical thinking, reasoning, and solution finding. Experts advocate for research that investigates understanding the nature of the thinking processes that underlie student's erroneous answers to problems, the related insufficiencies in other aspects of a genuine mathematical disposition (namely conceptual weaknesses), insufficient mastery of valuable heuristic methods, a lack of metacognitive awareness about the nature and the causes of one’s own misleading patterns of thinking, and a distrust of “weak” informal solution methods (Lester, Garofalo, & Kroll, 1989; The Cognition and Technology Group
at Vanderbilt, 1997; Okazaki & Koyama, 2005; Verschaffel & De Corte, 1997, 1999; Verschaffel et al., 1998).

A number of examples of successful standards-based programs that support meta-cognitive awareness as an important construct in learning mathematics are exemplified by the idea of “inquiry mathematics” (Cobb & Bauersfeld, 1995; Richards, 1996; Yackel, 2003). The theoretical basis for the standards is closely associated with Piaget’s theory of constructivism (Steffe & Kieren, 1994; Yackel, 1996) and supported through classroom activities characterized by problem solving, listening, discussion, mathematical justification, and “dialogical encounters” in which “one begins with the assumption that others have something to say to us and to contribute to our understanding” (Bernstein, 1992, p. 337).

The importance of conceptual knowledge skills in the development of reform mathematics can be seen in a number of research reports. Polhemus, Holmes, Jennings, Olson and Rubenfeld (2004), members of the Center for Initiatives in Pre-College Education (CIPCE) research program at Rensselaer Polytechnic Institute, suggested that the conceptual knowledge a student must possess to accurately interpret and answer mathematical problems on the New York State assessments requires a sense of inquiry that is difficult to acquire using traditional instructional techniques of mindless memorization and computation. To provide middle school students with a well-rounded foundation that will support the deconstruction of such questions CIPCE used new learning technologies in consort with inquiry-pedagogy to encourage students’ exploration and meaning making in mathematics. Researchers proposed that creating problem-solving situations designed to elicit the discovery of new conceptual connections
and new mathematical understandings, would positively affect student performance in mathematics (Polhemus et al., 2004). Researchers have also suggested students deep exploration into the activity using collected data allowed students the opportunity to dialogue about relationships and outcomes that were embedded in the realm of algebraic and proportional reasoning. Specifically, in CIPCE’s research, the problem solving, reasoning, communication, connections and representation standards were met when the student groups made predictions and their justifications supported the need to document and communicate meaning that can be interpreted intelligently by others. This research highlighted the fact that teachers viewed student interest and creativity as overall positive outcomes from the inquiry experience (Polhemus et al., 2004).

According to Ridgway, Titterington, and Sherman-Mc Cann (1999), inquiry is "an approach to teaching that involves a process of exploring the natural world, that leads to asking questions and making discoveries in the search of new understandings. It is a method of approaching problems to scientifically address matters encountered in everyday life and is based on the formation of hypotheses and theories and on the collection of relevant evidence. There is no set order to the steps involved in inquiry, but children need to use logic to devise their research questions, analyze their data, and make predictions" (Ridgway, Titterington, & Sherman-Mc Cann, 1999, p.38). When using inquiry methods of investigation, children learn that any question is reasonable and these questions are compatible with different hypotheses and facets of learning, suitable for mathematics as well as science (Wiggins & McTighe, 1998).

Polhemus et al. (2004) also stated that "learners who engage in inquiry approaches become more comfortable with the uncertainty of knowing, and they become
more aware of their own mentalities” (p. 19). Inquiry approaches facilitate a “learning orientation” to schoolwork, and it supports teachers’, as well as children’s, theories of intelligence as “incremental” rather than fixed (Ahmavaara & Houston, 2007, Polhemus, 2004). In inquiry-based pedagogies, teachers try to limit children’s efforts to get to the correct answer immediately in order to give them time to develop understanding and close the gap between not knowing and knowing (Duckworth, 1998). They try to open up learners’ minds by opening up the problem so there is time and there is space to try out, to question, to explore with the support of others, peers and teachers, who honor and enjoy such activity (Polhemus et al., 2004, pp.2-3).

An investigation by Hitz and Scanlon (2001) compared the effectiveness of project-based experiential learning method and the traditional classroom method in mathematics teaching. Their results suggested that educational experiences that are relevant and meaningful are the most effective pathways to learning. The data from their study supported the literature to the extent that when students are able to develop a connectedness through real-life experiences, they are able to apply that learning to other situations while developing conceptual understandings. The higher achievement scores, as indicated by the post-test results, for the traditional methodology students immediately following the unit of instruction supported the research on mastery of isolated skills and the development of knowledge to perform successfully on standardized tests (Hitz & Scanlon, 2001). Also, post-test results indicated a greater percentage of knowledge retained by those students in the project-based experiential learning method and reflected the prior research of Dewey (1933), Piaget (1984), and Gardner (1999). Results showed that project involved students demonstrated the ability to use their experiences
successfully over time. The researchers suggested that data collected show a higher level of achievement on the unit tests immediately following instruction for those students who were taught through the traditional classroom method of instruction, however, when comparing the percentage of gain or loss between the posttest and unit test scores for each unit of instruction, a higher percentage of retention for students who were taught through the project-based method was found (Hitz & Scanlon, 2001).

The second part of the study showed a decline in scores between the unit test and the post-test. The decline was much higher for the traditional group compared to the project-based group (Hitz & Scanlon, 2001). Project-based learning is designed to be an integral part of the curriculum where students receive a real or a potentially real problem and devise practical solutions through inquiry from the research they do (Hitz & Scanlon 2001). When students learn through experiences and inquire or investigate problems to answers or create solutions they are given the opportunity to build objects that interest them, while engaging in such activities, students will construct mental patterns that hold personal meaning that can be transferred and applied to other situations. This type of situated learning supports long-term retention and use of concepts learned in the process.

Children’s Discourse and Explanation of Solutions

According to the socio-cultural perspective, during social exchanges among members of a community there is a joint construction of “zones of proximal development” (Vygotsky, 1978, p. 84). “This construction takes place through guided participation” (Rogoff, 1990; Rogoff & Wertsch, 1984; Vygotsky, 1978). Additionally, mathematics students who participate in a learning community are better at applying
mathematics to ‘real world’ situations and have higher scores on tests using problem situations that did not ‘cue’ students on which procedures to use (Boaler, 1998, p. 9).

Through investigations of interaction and social exchange of information, the terms ‘collaborative learning’ and ‘cooperative learning’ are often used interchangeably to refer to educational activities in which students work together towards a common intellectual goal, founded primarily in constructivist theory. Panitz (1997) has been searching for the "Holy Grail of interactive learning, a distinction between collaborative and cooperative learning definitions" (p.5). According to Panitz (1997) as well as Johnson, Johnson and Smith (1991), collaborative and cooperative learning differ in some fundamental ways that influence how group activities are organized and presented to the class. That is, collaboration is a philosophy of interaction, whereas cooperation is a structure of interaction designed to facilitate the accomplishment of a specific end product or goal through people working together in groups. Collaborative learning highlights the contributions of individual group members, stresses the sharing of authority, and leads to consensus building on topics without a clear right and wrong answer. This process is immersed in a socio-cultural system of activities where participants get involved with different degrees of responsibility and action (Cole, 1996; Rojas-Drummond & Fernández, 2000). In both type of learning situations dialogue or "talk" among students is stressed; teacher intervention is not. Group governance and group processing remain in the hands of the students (Panitz, 1997).

While the distinction might be important to those who design mathematics curriculum, the answer to the debate over how to construct a social situation within which students can work is ultimately left to the classroom teacher and their interpretation of
what the reform movement defines as interaction. There are a number of studies which focus on the interaction of student and how such interaction impacts student achievement. For example, with cooperative learning, the instructor designs a task and a group structure for accomplishing the task, including the assignment of roles to group members. Cooperative learning is often thought of as a subset of collaborative learning that has students interacting under specific conditions set up by the teacher: positive interdependence, face-to-face interaction, individual accountability, collaborative skills, and group processing (Johnson, Johnson, & Smith, 1991). Additionally, efforts have been increasingly directed towards understanding how participants in learning communities engage in “distributed cognition” for solving problems and re-creating culture (Solomon, 1993; Rojas-Drummond & Fernández, 2000). The theory of distributed cognition (Hutchins 1995a) has been used to analyze and evaluate the flow of representations in real-world cooperative work settings. Furthermore, a number of researchers have emphasised the role of social interaction and discourse in outcomes of diverse educational practices (Brown & Campione, 1996; Cole, 1996, 1998; Wertsch et al., 1995; Rojas-Drummond & Fernández, 2000).

The definition of collaboration and cooperation becomes insignificant when one considers the specific activity of children and listens to children’s voices about those activities. In a classroom where students interact but must eventually be graded, the student's concern rests on the grade rather than the process. As teachers, we need to shift our students' focus to process if we want student to look at collaboration and cooperation as useful, valuable tools.

According to Johnson and Johnson (1999), classroom learning improves
significantly when students participate socially, interacting in face-to-face collaborative learning activities with small groups of members. This position, supported by the research of Chernobilsky, Dacosta and Hmelo-Silver (2004) is a clear link between the quality of group interaction and construction of knowledge. Researchers (Dillenbourg 1999; Johnson & Johnson, 1999; Staton, 2001; Vygotsky, 1978) contend that social interaction between peers established by the way in which interaction, synchronization, coordination and negotiation are developed among the group members, is built within a community and fundamental to achieving learning.

To be successful at this type of social interaction a number of researchers, Santoro, Borges and dos Santos (2000); Stahl (2002); Price, Rogers, Stanton and Smith (2003), argue that activities have to satisfy three distinct requirements: (1) clearly define common concepts and terms, (2) be sufficiently well-structured to provide a foundation for the subsequent development of new and increasingly more refined concepts and (3) should enable alternative designs of particular models and systems to be explicitly presented, compared and evaluated within the framework (in Johnson & Johnson, 1999). While we might define these constructs through the perspectives of a number of researchers, we need to know exactly what such ideas might look like within a mathematics classroom.

Ben-Yehuda, Lavy, Linchevski and Sfard (2005) investigated mechanisms of failure in mathematics by looking at communicative approaches to cognition, which describes thinking as an activity of communication and learning mathematics as an initiation to a certain type of discourse. In a search for factors that impede students' participation in arithmetic communication, researchers examined the arithmetical
discourse of two 18-year-old girls with long histories of learning difficulties. The authors defined discourse as any act of communication that was done and was not constrained by specific rules; rather they were meta-discursive in nature.

Communication and discourse are central to the current vision of desirable reformed mathematics teaching (NCTM, 1989; 1991). This view of mathematics teaching is quite different from traditional mathematics where instruction has been pre-designed and traditionally delivered; a process defined as narrowly focused on skills and procedures (Semper, 1996).

In reform mathematics, students are expected to engage in doing mathematics while participating actively in a "discourse community." For the purposes of this research discourse, "discourse communities" are those in which students feel free to express their thinking and take responsibility for listening, paraphrasing, questioning, and interpreting one another’s ideas in whole-class and small-group discussions (Clark, Jacobs, Pittman, & Borko, 2005). Thus, the new role envisioned for mathematics teachers is one intimately tied to issues of communication. Evidence exists that the acquisition of useful knowledge is related to quality of discourse (Wells, 1999). It is understood that a community of inquiry is a bi-directional process that is shaped by the interactions of its members, when embedded in the grounded rules of the classroom (Rojas-Drummond & Fernández, 2000).

According to a study conducted by Rojas-Drummond, and Fernández (2000), one important aspect of exploratory talk within a cooperative of collaborative situation is the child's capacity to take turns and distribute the tasks at hand among the group in an even way, thus limiting "disputational" talk during the activity. To evaluate communication
within this study, collective reasoning and problem-solving abilities, children were administered an adapted version of the Raven’s Test of Progressive Matrices, before and after intervention. The intervention was described as a four-month training program that promoted children’s use of exploratory talk, so that they could express and share their reasoning while they solved logical and social problems jointly (Rojas-Drummond & Fernández, 2000). The researchers carried out a three-way Analysis of Variance with experimental treatment and tests as the main independent factors. The dependent variable was the score obtained by each triad of students in the small-group version of the Raven’s test. Researchers found highly significant effects for the main factors and for the interaction between these factors (p < 0.0005) indicating the need to interpret their effects to interpret their effects in combination (p. 664). The results also illustrated an intricate relationship between language and reasoning in social contexts such as small groups of peers solving problems jointly. They point to subtle relationships between language used as a tool to facilitate collective reasoning, children’s capacity to solve problems while using language as a tool, and children’s sensitivity to the context of the problems they were attempting to solve, with a noticeable reduction in "disputational" talk. Students from the experimental groups showed a marked increase in performance from the pre- to the post-test, while students from the control groups showed relatively smaller improvement suggesting a “situatied” view of cognition (Rojas-Drummond & Fernández, 2000, p. 665).

In another study of sixth graders’ mathematical discourse, Seymour and Lerher (2006), explored a teacher's ability to increasingly recognize patterns of student talk and reasoning as they acted in increasingly contingent and productive ways to transform
students’ mathematical activity. Researchers (Choppin, 2007) suggest that discursive features of classrooms influence the ways in which students engage with mathematical content. Patterns of discourse that are dialogic provide opportunities to generate ideas which then become the focal point for collective reflection. In the study these opportunities for students to learn mathematics became building blocks for the development of mathematical understanding. The study suggested that when these types of discourse are in place, teachers could attenuate them to the variation in students’ present level of conversation. The development of discourse strategies over a two-year period allowed the teacher in the study to employ a more expansive, more flexible repertoire of means for supporting student reasoning, as well as continue to innovate in response to her student’s ideas. This connection between teachers learning and student discourse development is also seen in a number of other studies (e.g. Cobb & Bauersfeld, 1995; Cochran, Mayer, & Mullins, 2007; Richards, 1996; Yackel, 2003).

Although there is not a clear cut, agreed upon definition of what mathematical discourse looks like, (There is more to discourse than meets the ears…(Sfard, 2000) there is consensus on the need for social interactions with mathematical ideas which have the power to elicit unnoticed aspects of learning and to develop in-depth understanding.

**Test Design and Item Format**

The relevancy of methods for assessing and interpreting student understanding is not limited to pre-test preparation of mathematics, but should also address the type and format of the achievement testing as well. A review of discussions on mathematics item formats indicated conflicting perspectives on the value and purpose of item types and formats used in designing achievement tests that are evaluative of what students know
about mathematics. According to New York State Education Department (NYSED, 2009), the test items are “indicators used to assess a variety of mathematics skills and abilities. Each item is aligned with one content performance indicator for reporting purposes but is also aligned to one or more process performance indicators, as appropriate for the concepts embodied in the task. As a result of the alignment to both process and content strands, the tests assess students’ conceptual understanding, procedural fluency, and problem-solving abilities, rather than solely assessing their knowledge of isolated skills and facts” (p. 5). The two principal forms of knowledge assessment used by the NYSED (2009) are multiple-choice questions and constructed, extended, and open-ended response questions.

The widespread perception is that open-ended, constructed/extended response questions test a deeper understanding of the subject material; are better at evaluating the respondent’s integrative skills; assumed to have a zero probability of being answered correctly by guessing, and reflect higher working memory loads using more cognitive complexity (Kuechler & Simkin, 2010). Constructed response questions require respondents to create their own answers and response to these questions varies within the test. A final reason favoring extended response tests is the greater likelihood of structural fidelity. That is, “the degree to which examination questions require solving problems similar to those encountered in the actual work situations of a given field” (Kuechler & Simkin, 2010, p. 57). This concern is particularly important to those more interested in competent mathematicians rather than simply good test takers. Those with a contrasting view suggest that answers to open ended/constructed response questions tend to be more
difficult to grade, are more subjective, take more time, and most importantly, require substantial prerequisite knowledge and skill of both the teacher and the student (p. 57).

All variations of multiple choice questions have in common is the requirement that the test taker select an answer from a small list of possibilities. Researchers suggest that multiple choice items provide multiple difficulty levels, develop problem solving skills (reasoning for test taking), and are considered more flexible and uniform, while assessing a broad range of learner knowledge in a short period of time (i.e., efficient). However, these items always have some probability of being answered correctly by guessing and so must be described by three-parameter models where the third parameter is the probability of a random correct response (Kuechler & Simkin, 2010).

In their research on the comparability of multiple choice and constructed response questions, Kuechler and Simkin (2010) found that testing results were consistent with previously suggested theoretical perspectives which state that multiple choice question sections decreased in scores with increasing knowledge level, and differences in means scores were strongly significant. Thus, multiple choice questions did not correlate well with constructed response items.

In the current study, the decision to investigate open-ended, constructed/extended response question tasks as assessment tools for predictive power resulted from the comparisons made by researchers when trying to describe the value of type of test items. For example, one perspective is that students make sense of mathematics by exercising personal agency. Personal agency is the freedom and responsibility to choose to act (Walter & Gerson, 2006). Open-response tasks allow students to exercise personal agency, hence eliciting not only what students know, but also how students inquire about
and explore concepts. Open-response tasks require “students to explain their thinking and thus allow teachers to gain insights into the ‘holes’ in their understanding” (Moon & Schulman, 1995, p.30).

Research on the impact of reform curriculum and testing outcomes on instruction is necessary if, as suggested by Martinez (1999), testing item formats vary in their typical cognitive demand and in the range of cognitions they sample. Multiple-choice items, for example, often elicit low-level cognitive processing whereas constructed-response items more often evoke complex thinking (Martinez 1999). Teachers tend to gravitate towards one "correct" way to teach, which is to "reduce a task to simpler components and drill it repeatedly until pupils have mastered it" (Smith, 1991, p. 11). Furthermore, multiple choice testing tends to lead to multiple choice teaching, which lessens the likelihood that teachers will assist students in developing conceptual understanding of the content (Smith, 1991). Additionally, students failing to interpret information as the test maker intended can result in an "incorrect" response, even if the test-taker's response is potentially valid. The term "multiple guess" has been used by researchers when test-takers attempt to guess rather than determine the correct answer when viewing choices. In contrast, an open-ended response item allows the test-taker to make an argument for their viewpoint and potentially receive partial credit; however, the use of multiple choice formats remain popular due to its utility, efficiency and cost effectiveness. It is possible, according to Shepard (1991), for test scores to go up without there being any commensurate gain in learning. Recent research by the NAEP suggested, however, that students’ performance has improved on basic math skills but there has been no gain or decrease in higher order, advanced skills which they attribute to patterns in instruction.
that exhibit the negative influence of standardized testing on teaching and learning (Forgione, 1998). Cannell (1987) suggested that high stake testing in all 50 states produced inflated and spurious test results, which further reflects the impact of measurement driven instruction.

According to Verschaffel, De Corte, and Vierstraete (1999), weak performances in reasoning may be related to insufficiencies in a number of aspects of a genuine mathematical disposition, namely, conceptual weaknesses such as lack of understanding of number concept, the notions of addition and subtraction, insufficient mastery of valuable heuristic methods (like making a diagram and thinking of an analogous problem), lack of metacognitive awareness about the nature and the causes of one’s own fallacious patterns of thinking, and a distrust of “weak” informal solution methods (like counting) in cases of doubt about the applicability of “strong” formal methods (p 281).

While Verschaffel, De Corte, and Vierstraete (1999) do not explain which flaws in pupils’ knowledge bases are precisely and directly responsible for their thinking behavior and incorrect responses, the authors suggest a number of possible explanations such as: "Students approached these non-routine problems in a superficial and mindless way, without seriously taking into account the particularities of the problem situations that make the application of a routine-based solution method problematic" or "[I]t might be that pupils’ inaccurate thinking processes and weak performances on these non-routine problems are the “fruits” of years of exposure to a mathematics culture and practice that repetitiously reinforces children’s expectations that if there are two numbers in the problem, the answer can always be found by adding, subtracting, multiplying, or dividing these two numbers" (p. 282). They further propose that better results on these problem
types can be expected only if pupils regularly confront problems and experiences with rich and varied opportunities to explore, discuss, as well as reflect on various ways of modeling and solving such problems. These results also indicate a need to address this issue before these strong formal methods in problem solving of non-routine problems by encouraging students to look at problem solving in a less formal, more flexible way at an early age.

**Summary**

While the previously discussed studies have outlined a number of important findings about dialogue, discourse, and inquiry methodologies that support conceptual knowledge, the impact of these constructs is usually considered through a limited lens of a few teachers and a limited number of students within a restricted amount of time. Although Wenger (1998) assumes that learning “has a beginning and an end; that it is best separated from the rest of our activities; and that it is the result of teaching.” (p. 3), project staff (Blythe, Allen, & Powell, 1999) chose to explore at the intermediate and integrated function of teaching math. "We can only begin to see and understand the serious work that students undertake if we suspend judgment long enough to look carefully and closely at what is actually in the work rather than what we hope to see in it." (Blythe, Allen, & Powell, 1999).

In a study entitled *How Long Do Teacher Effects Persist?*, Spyros Konstantopoulos (2007) undertook the task of investigating whether teachers and their instruction type effects student achievement and if such impact persists over time. The results suggest that teacher effects in early grades endure over time up to grade three. Due to the random assignment of teachers and students to classrooms in Konstantopoulos’s
(2007) experiment, results suggest that there is strong evidence about the durability of teacher effects. Therefore, teacher choices about type of instruction significantly affect student achievement not only in the current or the following year, but in subsequent years as well. That is, teacher effects are cumulative.

Senk and Thompson (2003) and the findings of the larger-scale comparative studies conducted by external reviewers (i.e. Newman, 2007) propose that students taught using standards-based curricula tended to hold their own on tests of computational skills and outperform students taught with conventional curricula on tests of thinking, reasoning, and conceptual understanding. This pattern of findings, in support of reform initiative and development of proof of student understanding, has prompted some researchers evaluate the overall efficacy of standards-based curricula (Schoenfeld, 2002).

The purpose of this research is to broaden the growing body of knowledge pertaining to student achievement in mathematics as it pertains to the extent that third grade and fourth grade student performance on five categories of New York State Mathematics Testing Items in third and fourth grade predict success on students’ subsequent overall fifth grade mathematics achievement test result. It additionally seeks to understand if there are specific areas of content or patterns of specific test subscales that predict mathematics proficiency? The research investigates the predictive power of performance on open-ended constructed response questions at 3rd and 4th grade. This is a first step toward a sustained body of research which seeks to define the connection between student achievement, use of assessment results, changes in teacher practice and students mathematical awareness and conceptual knowledge.
CHAPTER 3: METHOD

The purpose of this chapter is to present the rationale and methodology used for this study, by describing the research design, participants, and instrumentation. The chapter also describes the dependent and independent variables, outlines the research procedures, discusses data analysis, and specifies the assumptions and limitations of the study.

Purpose

This longitudinal study investigates the degree to which student responses on open-ended, constructed/extended response mathematics questions in third and fourth grade predict success on students’ subsequent overall fifth grade mathematics achievement test results. Predictor variables consisted of student responses to open-ended, constructed/extended response items in grade three and four on item subscales of statistics and probability, number sense and numeration, geometry, algebra, and measurement. Constructed, open-ended question formats are considered by some researchers as contributors to mathematical awareness and understanding, and lend themselves to understanding students’ conceptual knowledge (Moon & Schulman, 1995), supporting the development and exhibition of reasoning skills which allow teachers insights into what students actually know. The aim of the research is to examine the role of mathematical awareness and understanding and its relationship to student achievement in early mathematics education and to broaden the growing body of knowledge pertaining to student achievement in mathematics.
Sample

The sample for the current study includes 332, students in one New York State elementary school third, fourth and fifth grades from the 2003-2009 in one school district. The sample includes students of all ability levels, including students receiving special education services (NY State Report Card, 2009). The district and building’s participation was voluntary; the students involvement in the tests were part of a statewide program hence ex-post facto and mandatory.

The Similar District Groups within a Need/Resource Capacity identified the building as an elementary school in an urban or suburban school district with high student needs in relation to district resources; that is, the school draws students from poorer urban areas and more affluent suburban area as well as students in multiple academic, cultural, and financial categories (New York State Report Card, 2009). At the time of the data development, the school enrolled approximately 560 students per year and was composed of culturally mixed races: 16% Black or African American, 8% Hispanic or Latino, 68% White, and 8% Multi-racial as well as 1.5% English Language Learners. The school spent approximately $11,896 per pupil and served students in grades Kindergarten through 5. Forty-six percent of the school’s students received free and or reduced lunch. One hundred percent of teachers in the school were identified as Highly Qualified, certified to teach in their subject area, and with subject matter competency. The attendance rate of the school was 93% and the suspension rate 10%. The overall accountability status of the district was in good standing at all levels and subject areas including Math, English, and Science and graduation rate (New York State Report Card, 2009).
Instruments

The study utilized components of two instruments: the New York State Testing Program Mathematics Test (grade four, 2003-2009; grade three and five, 2006-2009) and Test of New York State Standards (TONYSS), achievement tests modeled for and applicable to the New York State standards in mathematics (grades three and five, 2003–2005). The specific scores used included the fifth grade overall proficiency outcomes; non-proficient (scores of 1 and 2) and proficient (score of 3 and 4) and third and fourth grade open-ended, extended/constructed response test questions. The open-ended questions included in grade 3 assessed the categories of statistics and probability, number sense and numeration, geometry, and algebra. The content-specific variable of measurement was eliminated in 2005 on the third grade test as a constructed response item, thus it could not be used as a variable for the third grade analysis. In fourth grade the categories assessed were statistics and probability, number sense and numeration, geometry, algebra, and measurement. Total test point values for content-specific categories ranged from 6-14 points depending on year and grade level. Students were allowed eighty five to one hundred fifty minutes to complete the open-ended section (NYSED, 2009). The constructed response categories reflected the National Council of Teachers of Mathematics (NCTM) and the New York State standards and cover mathematics content using the following blueprint:

Statistics and probability.

Students collect, organize, display, and analyze data; make predictions that are based upon data analysis; and understand and apply concepts of probability (NYSED, 2010, p. 26).
**Number sense, numeration and operations.**

Students understand numbers, multiple ways of representing numbers, relationships among numbers, and number systems; understand meanings of operations and procedures, and how they relate to one another; compute accurately and make reasonable estimates (NYSED, 2010, p. 26).

**Geometry.**

Students use visualization and spatial reasoning to analyze characteristics and properties of geometric shape; identify and justify geometric relationships, formally and informally; apply transformations and symmetry to analyze problem solving situations; apply coordinate geometry to analyze problem solving situations (NYSED, 2010, p. 26).

**Algebra.**

Students represent and analyze algebraically a wide variety of problem solving situations; perform algebraic procedures accurately; recognize, use, and represent algebraically patterns, relations, and functions (NYSED, 2010, p. 26).

**Measurement (grade 4 only).**

Students determine what can be measured and how, using appropriate methods and formulas; use units to give meaning to measurements; understand that all measurement contains error and are able to determine its significance; develop strategies for estimating measurements (NYSED, 2010, p. 26).

Grades 3 through 5 mathematics TONYSS and the New York State Testing Program Mathematics tests are composed of multiple-choice and constructed-response items differentiated by maximum score points. Multiple-choice items have a maximum
score of 1, and the open-ended/constructed response items have a maximum score of 3 points (NYSED, 2010). According to New York State Education Department (2010), students completing the test were assessed on their ability to communicate and reason mathematically, construct solutions to multiple step problems, interpret data, analyze information and solve problems. As noted previously, prior to federally mandated testing in grades 3-8, the TONYSS was used by the school to provide formative assessment in non-mandated testing grade levels. After 2006, the New York State test was used in all grade levels.

Table 1

Summary of Data Sources from Tests 2003-2009

<table>
<thead>
<tr>
<th>Student Scores***</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort 1</td>
<td>4th NYS</td>
<td>5th</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cohort 2</td>
<td>3rd TONYSS</td>
<td>4th NYS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cohort 3</td>
<td>3rd TONYSS</td>
<td>4th NYS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cohort 4</td>
<td>TONYSS</td>
<td>NYS</td>
<td>NYS</td>
<td>5th</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cohort 5</td>
<td>NYS</td>
<td>4th</td>
<td>NYS</td>
<td>5th</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cohort 6</td>
<td>4th NYS</td>
<td>5th</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cohort 7</td>
<td>NYS</td>
<td>5th</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*TONYSS, Test of New York State Standards  
**NYS, New York State Testing Program Mathematics Test  
***Scores for 3rd and 4th grade include open-ended/constructed response subscales; Scores for 5th grade include complete test results

As summarized in Table 1, the Test of New York State Standards (TONYSS) data was collected for all students in grades three and five from school years 2002-2003, 2003-2004, and 2004-2005. Additionally, the New York State Testing Program
Mathematics Test score data are available for fourth grades from 2003-2008 and for students in grades three and five for 2005-2006, 2006-2007, 2007-2008 and 2008-2009. The results of student achievement in fifth grade on these instruments were represented four levels of proficiency based on New York State Standards. According to the New York State Education Department (2006), the proficiency cut scores (Levels I, II, III and IV) were established during the process of standard setting and reflect the abilities intended included in the New York state Learning standard and Core curriculum (New York State Education Department, 2006). Performance of students on the Grades 3-8 Math Tests is reported as one of four performance level classifications presented below:

**Non-proficient**

Level 1 (Below Basic): Students do not meet the learning standards; their performance shows minimal understanding of key mathematical ideas; and

Level 2 (Basic): Students show partial achievement of the learning standards. Their performance shows partial understanding of key mathematical ideas;

**Proficient**

Level 3 (Proficient): Students meet the learning standards. Students demonstrated proficiency in the standards expected for their grade level; and

Level 4 (Advanced): Students exceed the learning standards for mathematics. Their performance shows superior understanding of key mathematical ideas.

(NYSED, 2006)

The actual test contains both multiple choice questions and performance assessment items (open-ended, constructed/extended response) and is given in the spring
of the school year. The third, fourth, and fifth grade tests are timed tests that are given over a three-day period (NYSED, 2007) and require a level three to meet proficiency.

The Test of New York State Standards (TONYSS) for Mathematics is a standardized test; in this school the test was administered to third and fifth grade students from 2003-2005 when the NYS test was given only to students in grade 3. The TONYSS measure all seven New York State Learning Standards and Key Ideas for Mathematics and also required students to score at a level three or higher to meet state standards (Riverside Publishing, 2006).

The Riverside Publishing service provided participating schools with student records containing demographic information, item data, and scores. These permitted the local district to access and disaggregate data, create special reports, and merge data into a local file. Prior to required State testing the TONYSS was considered a beneficial method for formative assessment for mathematics because it;

The benefits of using the TONYSS described by the publisher included:

- Aligned with New York State Learning standards,
- Had content and test-item formats mirroring the New York State Testing Program, Mathematics Tests, including multiple-choice, short-answer, response, extended-response, and writing tasks,
- Was review by New York state educators to ensure grade-level appropriateness and bias-free content,
- Was used in research studies conducted throughout districts in New York State provide information based on the performance of New York students, and
• Provided reports designed to reflect New York State Learning Standards yield important information to New York state educators and parents (Riverside Publishing, 2006).

The district Superintendent (personal communication, June 20, 2005) stated that the District adopted the TONYSS in grades three and five, because the district decided they were better aligned to the state assessments than other tests and demonstrated equivalence with the New York State Testing Program Mathematics tests. The New York State Education Department’s standardized test was developed as part of New York State Testing Program, Mathematics Test. After 2006, students in grades 3, 4 and 5 were required to take the New York State Testing Program, Mathematics test. As a result of the availability of these two testing formats, equivalent mathematics achievement data were available for a sample of students from 2003-2009 in grades 3 through 5.

Confidentiality (Anonymity)

All collected data was kept confidential and made anonymous when the school district created pseudo numbers for both teachers and students in the study. Any identifying descriptors were removed from all data. Since all student data is archival, there was no direct involvement with students.

Research Design

A correlational multivariate research design using a priori data was utilized for the purposes of this study. A statistical discriminant function analysis was used to determine which variables predict or discriminate between the two groups (non-proficient/ proficient) and provide a measure of association of the predictors and the group categorization and to developed weights for the discriminant function. The analysis
was performed using the discriminate subprogram in PAWS:SPSS (Statistical Package for Social Sciences, version 18).

More specifically, discriminant analysis were used: 1) to investigate which content-specific variables from grades three and four best predict fifth grade proficiency level; 2) to determined the degree of association of third and fourth grade content variables with 5th grade outcomes and 3) to outline the generation of weights that allow for the establishment of predictive functions.

As previously mentioned, the analysis focused on content-specific subscales for statistics and probability, number sense, numeration and operations, geometry, and algebra in grade 3 and statistics and probability, number sense, numeration and operations, geometry, algebra and measurement in grade 4.

**Assumptions of Discriminant Analysis:**

Assumptions for discriminant analysis were examined for data at both grade 3 and 4. Following is a summary of their status for this research study:

*Sample size:* Unequal sample sizes are acceptable. However, the sample size of the smallest group needs to exceed the number of predictor variables (Poulsen & French, 2008). In the case of this study, the sample size of the smallest group exceeds the number of predictor variables for both analyses and meets the discriminant analyses requirement of 20 cases per independent variable (Stevens, 2001).

*Normal distribution:* The assumption that the data (for the variables) represent a sample from a multivariate normal distribution (Tabachnick & Fidell 1996) was met. In these analyses 100% of the values in the population are within one standard deviation of the mean.
Homogeneity of variances/covariances: Discriminant Analysis is very sensitive to heterogeneity of variance-covariance matrices. Before accepting final conclusions, the within-groups variances and correlation matrices were reviewed (Poulsen & French, 2008). The test for homogeneity of variance in multivariate statistics was met through non significant Box’s M scores for both analyses.

Outliers: Discriminant analysis is highly sensitive to the inclusion of outliers. The study procedure included testing for univariate and multivariate outliers for each group, as well as transforming or eliminating them (Garson, 2008). For these analyses there were no outliers included.

Based on these findings it was determined that all assumptions necessary for discriminant analysis were met for the third and fourth grade analyses.

Data Analysis

Discriminant function analysis was selected for this study because it provides the best linear weighting of predictor variables to maximize the differences among two or more groups (Pyryt, 2004) and produces discriminant function coefficients for each predictor variable that indicate the importance of each variable (Smith, 1979). SPSS (Statistical Package for Social Sciences, version 18) was used in this study This subprogram permits a sophisticated predictive view of the test results (Thomas, Marr, Thomas, Hume, Neff & Walker, 1996) by providing depth of information based on the analysis of individual student results tracked over the course of three years. Discriminant analysis allows the researcher to investigate the differences between two or more groups of people with respect to several underlying variables, and provides correlations and prediction weights for identifying groups instead of continuous variables (Garson, 2008;
Thomas et al, 1996). Additionally, discriminant analysis is preferred over logistic regression when the assumptions of linear regression are met since DA has more statistical power than logistic regression, which decreases the chance of Type II errors or accepting a false null hypothesis (Garson, 2008).

**Procedure**

For this study, the independent, predictor variables were the percentage of correct answers on open-ended (extended/constructed) response questions for subscales of statistics and probability, number sense and numeration, geometry, algebra, and measurement in grade 3; and statistics and probability, number sense and numeration, geometry, and algebra, in grade 4 as outlined by the NCTM and the New York State standards (NYSED, 2010). It should be noted that the number of points for open-ended, extended/constructed response questions for the yearly tests were different over the course of the seven years. To correct for the difference, the point values used for each student were based on the percentage of total possible points in each subscale for each year. When variables are standardized, absolute weights can be used to rank variables in terms of their discriminating power, the largest weight being associated with the most powerful discriminating variable (Burns & Burns, 2008). Variables with large weights are those which contribute most to differentiating the groups. The dependent or grouping variable was the level of proficiency on the fifth grade test at achieved level scores of 1 or 2 (not passing or not proficient) and 3 or 4 (passing or proficient) on the overall test scores. A summary of the analyses is presented in Table 2.
Table 2

*Relationship of Predictors to Success in Mathematics*

<table>
<thead>
<tr>
<th>Analysis 1 Grade 3</th>
<th>Independent Predictor Variables*</th>
<th>Dependent Grouping Variable**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-Statistics and Probability</td>
<td>-Group 1: non proficient (1 or 2) on 5th grade mathematics</td>
</tr>
<tr>
<td></td>
<td>-Geometry</td>
<td>-Group 2: proficient (3 or 4) on 5th grade mathematics</td>
</tr>
<tr>
<td></td>
<td>-Number Sense, Numeration and Operations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Algebra</td>
<td></td>
</tr>
<tr>
<td>Analysis 2 Grade 4</td>
<td>-Statistics and Probability</td>
<td>-Group 1: non proficient (1 or 2) on 5th grade mathematics test</td>
</tr>
<tr>
<td></td>
<td>-Geometry</td>
<td>-Group 2: proficient (3 or 4) on 5th grade mathematics test</td>
</tr>
<tr>
<td></td>
<td>-Number Sense, Numeration and Operations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Algebra</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-Measurement</td>
<td></td>
</tr>
</tbody>
</table>

* Mathematics subscales from test results statistics and probability, number sense and numeration, geometry, and algebra, (and measurement, grade 4)

** Results on complete mathematics test

The analyses generate a test of significance of the predictor equation and associated correlational information. Findings also provide weights for the math test content-specific subscales. As part of the process, the analysis provides a classification matrix to determine how well the classification functions predict group membership of cases and a discriminant function equation that can be used to classify students as making proficient (predicted grade of 3 or 4) or non proficient scores (predicted grade of 1 or 2) for overall grade five test results,
CHAPTER 4: RESULTS

This chapter presents the results of the analyses utilized to answer the research questions of this study. Statistical discriminant function analysis was conducted to determine the ability of content specific open-ended, constructed response questions at grades 3 and 4 to predict proficiency levels (proficient or non-proficient) on fifth grade students’ mathematics achievement tests. The content-specific categories, described by the New York State Education Department as strands of content that students should learn (NYSED, 2009) served as predictor variables and consisted of (at grade 3) statistics and probability, number sense, numeration and operations, geometry, and algebra, and (at grade 4) statistics and probability, measurement, number sense, numeration and operations, geometry, and algebra. These categories are assessed by the TONYSS for Mathematics (2003-2005 at grade three) and the New York State Testing Program, Mathematics Tests (2005 -2009 at grade 3, 2003-2009 at grade four and five). The fixed categorical dependent grouping variables were levels of proficiency attained on the fifth grade New York State Testing Program, Mathematics Tests: Levels 1 and 2, categorized as non-proficient and levels 3 and 4, categorized as proficient.

Of the original 560 students available for the study, 228 were eliminated because they left the school, missed the fifth grade testing session, or had results on the fifth grade tests only. Of the remaining 332 cases, 177 students were male, and 155 were female. On the fifth grade outcomes, 31 students earned proficiency scores at level 1, 83 students were at level 2, 183 students were at level 3, and 36 earned a level 4. Of these students 114 students or 34% earned a non-proficient or not passing score and 218 students or 66% earned a proficient or passing score.
A robust analysis was supported by meeting the requirement for discriminant analysis of 20 cases for each independent variable to establish stability for the coefficients and correlations (Stevens, 2001). The analysis was designed in two stages: Analysis 1 \((n=280)\) involved the independent predictor variables of four categories of constructed response questions at grade 3 and dependent variables in two groups: non-proficient (levels 1 and 2) and proficient (levels 3 and 4) at fifth grade. Analysis 2 \((n=221)\) involved the independent predictor variables of five categories of constructed response questions at grade 4 and the dependent variables composed of two groups: non-proficient (levels 1 and 2) and proficient (levels 3 and 4) at grade five. The following sections present the analyses of the data for this study.

**Analysis 1: Grade Three**

To determine whether categories of open-ended, constructed/extended response categories were important in predicting a group membership for proficiency level at grade five, the data were analyzed for grade three \((n=280)\) through a statistical discriminant function analysis. Predictor variables included statistics and probability grade 3, measurement grade 3, number sense, numeration and operations grade 3, geometry grade 3, and algebra grade 3. Analysis of predictor variables revealed that the assumptions of linearity, normality, multicollinearity (or singularity), and homogeneity of variance-covariance matrices were met (Box’s M, \(p>.05\)).

A summary of the discriminant function analyses of predictor variables in third grade is presented in Tables 3 and 4.
Table 3

Results of Grade Three Discriminant Function Analysis for Test Predictor Variables

<table>
<thead>
<tr>
<th>Independent Variables Grade 3</th>
<th>Standardized Canonical Discriminant Function Coefficients</th>
<th>Structured Matrix</th>
<th>Univariate F (1, 278)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistics and Probability</td>
<td>.65</td>
<td>.81</td>
<td>64.49*</td>
</tr>
<tr>
<td>Geometry</td>
<td>.49</td>
<td>.71</td>
<td>49.79*</td>
</tr>
<tr>
<td>Number Sense, Numeration and Operations</td>
<td>.20</td>
<td>.59</td>
<td>33.79*</td>
</tr>
<tr>
<td>Algebra</td>
<td>.01</td>
<td>.49</td>
<td>23.15*</td>
</tr>
<tr>
<td>Canonical R</td>
<td>.51</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chi Statistic \( \chi^2(\text{df}=4) = 83.24^* \)

* \( p < .05 \)  

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions  

One canonical discriminant functions was used in the analysis

As presented in Table 3, the test of the relationship between the third grade predictor variables and fifth grade outcomes was statistically significant, \( p < .05 \) (\( \chi^2 = 83.24, \text{df}=4 \)). This test enables the researcher to see how well the function separates the groups (Burns & Burns, 2008); that is, the open-ended, constructed/extended response categories did not just randomly predict fifth grade test proficiency level. The standardized canonical discriminant function coefficients (reflecting the unique shared variance of each independent variable with the dependent variable (Burns & Burns, 2008) indicated a strong relationship between the independent variable of open-ended response questions and the dependent variable of proficient or non-proficient (Canonical R of .51) and purported that 26% of the variance between groups and predictors can be accounted for by the discriminant function.
The structure matrix in Table 3 shows the relationship between the variables in the model and the discriminant functions. These coefficients serve as structure weights or discriminant loadings that reflect the contribution of one variable in the context of the other variables in the model (Wuensch, 2008): The larger the number the more predictive the power of the construct. Generally 0.30 is seen as the cut-off between important and less important variables (Burns & Burns, 2008). This analysis suggests that at grade three, all categories of Statistics and Probability (.812), Geometry (.713), Number Sense Numeration and Operations, (.588) and Algebra (.486) are important variables in predicting fifth grade proficiency levels.

Table 4

*Third Grade Means and Standard Deviations by Proficiency level for Predictor Variables*

| Grade 3 | Non-Proficient a | | | Proficient b | | | Comparison between proficiency levels |
|---------|-----------------|----------|----------|-----------------|----------|------------------|
|         | Mean Standard Deviation Standard Error | Mean Standard Deviation Standard Error | | |
| Statistics and Probability | .45 .30 .03 | .76 .30 .02 | .31 |
| Number Sense, Numeration and Operations | .35 .34 .04 | .63 .39 .03 | .28 |
| Geometry | .45 .37 .04 | .75 .32 .02 | .30 |
| Algebra | .46 .28 .03 | .64 .29 .02 | .18 |

\(^a_n=93 \quad ^b_n=187\). Wilks Lambda Test of Equality of Group Means, p<.05

In Table 4 the mean values are provided for each predictors by level of proficiency. A small standard deviation (relative to the mean score) indicated that the majority of individuals tended to have scores that are very close to the mean (Burns & Burns, 2008). The mean or average score on the third grade test extended response questions ranged from 35% to 46% for those who are non-proficient while those meeting
proficiency had rates ranging from 63% to 76% of the possible points available for each category. Proficient students’ mean score in Statistics and Probability was .31 greater for proficient students than non-proficient students. Similar results were evident for Number Sense, Numeration and Operations and Geometry. Algebra differences were low at .18. The Wilks Lambda Test of Equality of Group Means was significant (p<.05) indicating that the function of predictors significantly differentiated between proficient and non-proficient in fifth grade. The univariate F tests presented in table 3 indicate that the difference between means were significant across the groups.

![Figure 1](image)

**Figure 1 Function at Group Centroids for Grade 3**

As shown in Figure 1, the discriminant function separated the non-proficient and the proficient groups. Group centroids indicated the distance (1.25) between variables. Large centroids differences reflect better group discriminability, showing that the groups are mutually exclusive and collectively exhaustive (Burns & Burns, 2008).

Presented in Table 5 is an assessment of adequacy of the predictive equation in classification of proficient and non-proficient levels on the fifth grade achievement test based on third grade students’ scores. The analysis, through cross validation, allows for evaluating the accuracy of the function to predict or classify subjects into the appropriate group and is designed to give an unbiased measure of the model's predictive power (Burns & Burns, 2008; Mertler & Vannatta, 2010). Cross validation is used to
discriminate between different (naturally occurring) groups, and for classifying cases into different groups with a better than chance accuracy.

Table 5

*Grade Three Predicted Classification of Group Membership*

<table>
<thead>
<tr>
<th>Classification Results&lt;sup&gt;b,c&lt;/sup&gt;</th>
<th>Predicted Group Membership</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non Proficient</td>
<td>Proficient</td>
</tr>
<tr>
<td>Cases Selected</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Original count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actual count</td>
<td>Non-Proficient</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Proficient</td>
<td>52</td>
</tr>
<tr>
<td>Actual %</td>
<td>Non-Proficient</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>Proficient</td>
<td>28</td>
</tr>
<tr>
<td>Cross-validated&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actual count</td>
<td>Non-Proficient</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Proficient</td>
<td>54</td>
</tr>
<tr>
<td>Actual %</td>
<td>Non-Proficient</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>Proficient</td>
<td>29</td>
</tr>
</tbody>
</table>

<sup>a</sup> Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case. <sup>b</sup> 74% of selected original grouped cases correctly classified. <sup>c</sup> 74% of selected cross-validated grouped cases correctly classified.

For the purposes of this analysis a “leave one out” methodology of cross validation through SPSS was used. This involves using a single observation from the original sample as the validation data, and the remaining observations as the training data (Statsoft, 2011). This is repeated such that each observation in the sample is used once as the validation data. Repetition by this method created by SPSS analyses reinforces that observed classification is significantly different from expected chance classification (Burns & Burns, 2008).
The classification results (Table 5) reveal that overall, 74% of students were classified correctly into passing or not passing groups. This overall predictive accuracy was supported by a 74% cross-validation rate. A hit ratio that is approximately 25% larger than that due to chance or 50/50 is generally acceptable (Burns & Burns, 2008).

**Analysis 2: Grade Four**

Discriminant function analysis also was used to determine whether categories of fourth grade (n=221) open-ended, constructed/extended response categories were important in predicting group membership (proficiency /non-proficient) at grade five. Predictor variables included Statistics and Probability grade 4, Number Sense, Numeration and Operations grade 4, Geometry grade 4, Algebra grade 4, and Measurement grade 4. The assumptions of linearity, normality, multicollinearity (or singularity), and homogeneity of variance-covariance matrices were met (Box’s M, p>.05).

The discriminant function analyses of predictor variables in fourth grade are presented in Table 6. The test of the relationship between the fourth grade predictor variables and fifth grade outcomes was found to be statistically significant, p<.05, ($\chi^2$ = 120.42, df =5); open-ended, constructed/extended response categories did not just randomly predict fifth grade test outcome. The canonical R was .65 indicating that 43% of the variance between groups and predictors can be accounted for by the discriminant function.
Table 6

Results of Grade Four Discriminant Function Analysis for Test Predictor Variables

<table>
<thead>
<tr>
<th>Independent Variables Grade 4</th>
<th>Standardized Canonical Discriminant Function Coefficients b</th>
<th>Structured Matrix a</th>
<th>Univariate F (1, 219)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Sense, Numeration and Operations</td>
<td>.79</td>
<td>.89</td>
<td>130.30*</td>
</tr>
<tr>
<td>Algebra</td>
<td>.45</td>
<td>.68</td>
<td>74.40*</td>
</tr>
<tr>
<td>Statistics and Probability</td>
<td>-.11</td>
<td>.41</td>
<td>28.07*</td>
</tr>
<tr>
<td>Measurement</td>
<td>-.07</td>
<td>.39</td>
<td>24.72*</td>
</tr>
<tr>
<td>Geometry</td>
<td>.18</td>
<td>.38</td>
<td>23.30*</td>
</tr>
</tbody>
</table>

Canonical R .65

Chi Statistic $\chi^2$ (df=5) = 120.42*

* p < .05 aPooled within-groups correlations between discriminating variables and standardized canonical discriminant functions. bOne canonical discriminant functions were used in the analysis

The structure matrix in Table 6 shows the relationship between variables in the model and the discriminant functions. Based on the .30 cut off, (Burns & Burns, 2008) all variable at grade 4, including Number sense, Numeration and Operations (.894), Algebra (.676), Statistics and Probability (.415), Measurement (.390), and Geometry (.378) predict fifth grade proficiency. The univariate F tests presented in Table 6 indicates that the difference between means are significant across the groups.

The standardized canonical discriminant function coefficients (reflecting the unique shared variance of each independent variable with the dependent variable (Burns & Burns, 2008) indicated a strong relationship between the independent variable of open-ended response questions and the dependent variable of proficiency or non-proficiency. As seen in Figure 2, the discriminant function separated the non-proficient and the proficient groups. Group centroids indicated the distance (1.38) between variables.
The mean values shown on Table 7 document the differences between grouping variables. A small standard deviation (relative to the mean score) indicated that the majority of individuals in fourth grade tended to have scores that are very close to the mean (Burns & Burns, 2008).

Table 7

*Fourth Grade Means and Standard Deviations by Proficiency Level for Predictor Variables*

<table>
<thead>
<tr>
<th>Grade 4</th>
<th>Non-Proficient&lt;sup&gt;a&lt;/sup&gt;</th>
<th></th>
<th>Proficient&lt;sup&gt;b&lt;/sup&gt;</th>
<th></th>
<th>Comparison between proficiency levels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Standard Deviation</td>
<td>Mean</td>
<td>Standard Deviation</td>
<td>Standard Error</td>
</tr>
<tr>
<td>Statistics and Probability</td>
<td>.54</td>
<td>.29</td>
<td>.03</td>
<td>.74</td>
<td>.25</td>
</tr>
<tr>
<td>Number Sense, Numeration and Operations</td>
<td>.40</td>
<td>.19</td>
<td>.02</td>
<td>.71</td>
<td>.20</td>
</tr>
<tr>
<td>Geometry</td>
<td>.64</td>
<td>.24</td>
<td>.03</td>
<td>.79</td>
<td>.22</td>
</tr>
<tr>
<td>Measurement</td>
<td>.58</td>
<td>.28</td>
<td>.03</td>
<td>.76</td>
<td>.27</td>
</tr>
<tr>
<td>Algebra</td>
<td>.37</td>
<td>.31</td>
<td>.03</td>
<td>.73</td>
<td>.29</td>
</tr>
</tbody>
</table>

<sup>a</sup><sup>n=84</sup> <sup>b</sup><sup>n=137</sup>. Wilks Lambda Test of Equality of Group Means, p<.05

The mean or average score on the fourth grade test open-ended, constructed response questions ranged from 37% to 64% for those at the non-proficient level and those meeting proficiency had rates ranging from 71% to 79% of the possible points available for each category. Proficient students’ mean score in Algebra (.36) and Number Sense, Numeration and Operations (.31) were greater for proficient students than non-
proficient students. Similar results were evident for Statistics and Probability (.20).

Geometry and Measurement differences were low at .15 and .18. Wilks Lambda Test of Equality of Group Means was significant (p < .05) indicating that the function of predictors significantly differentiated between proficient and non-proficient in fifth grade.

Presented in Table 8 is an assessment of adequacy in classification of passing and not passing the fifth grade achievement test for grade 4 students.

Table 8

*Grade Four Predicted Classification of Group Membership*

<table>
<thead>
<tr>
<th>Classification Results</th>
<th>Predicted Group Membership</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Proficient</td>
</tr>
<tr>
<td>Cases Selected</td>
<td></td>
</tr>
<tr>
<td>Original count</td>
<td></td>
</tr>
<tr>
<td>Non-Proficient</td>
<td>69</td>
</tr>
<tr>
<td>Proficient</td>
<td>27</td>
</tr>
<tr>
<td>Actual %</td>
<td></td>
</tr>
<tr>
<td>Non-Proficient</td>
<td>82</td>
</tr>
<tr>
<td>Proficient</td>
<td>20</td>
</tr>
<tr>
<td>Cross-validated</td>
<td></td>
</tr>
<tr>
<td>Actual count</td>
<td></td>
</tr>
<tr>
<td>Non-Proficient</td>
<td>67</td>
</tr>
<tr>
<td>Proficient</td>
<td>28</td>
</tr>
<tr>
<td>Actual %</td>
<td></td>
</tr>
<tr>
<td>Non-Proficient</td>
<td>80</td>
</tr>
<tr>
<td>Proficient</td>
<td>20</td>
</tr>
</tbody>
</table>

* a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.
  b. 81% of selected original grouped cases correctly classified.
  c. 80% of selected cross-validated grouped cases correctly classified.

Overall, 81% of all cases were correctly classified as proficient or non-proficient on the fifth grade test and 80% of selected cross-validated grouped cases were also...
correctly classified. Both hit ratios are 25% larger than that due to chance or 50/50 and meet the criteria for observed classification significantly different from expected chance classification (Burns & Burns, 2008).

Summary

A review and comparisons of the analyses for grade 3 and 4 suggest that the predictive power of scores on specific content differs by grade level. Results of the statistical discriminant function analysis suggest that open-ended, constructed/extended response items in both third and fourth grade achievement tests can discriminate between levels of passing on fifth grade mathematics achievement tests. A summary of comparison of results is presented in Table 9. The best predictors of group membership for proficiency levels in grade 3 are: statistics and probability (i.e., collecting, organizing, displaying, analyzing data, making predictions) and geometry (i.e., understanding, visualizing and reasoning, analyzing characteristics, relationships and properties of geometric shapes). In grade 4 the best predictors of fifth grade test performance were: number sense, (i.e., understanding of numbers, multiple representations of quantity, relationships, number systems, and operations and procedures), and algebra (i.e. understanding and analyzing algebraically a wide variety of problem solving situations; patterns, relationships, and functions. It should be noted, however that the fourth grade outcomes for statistics and probability and measurement may be confounding variables as they are in a negative direction. A more in-depth analysis of these variables is warranted.
Table 9

**Comparison Summary of Grade Level Predictor Variables in Grade 3 and 4**

<table>
<thead>
<tr>
<th>Grade 3</th>
<th>Grade 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistic and Probability and Geometry are most predictive of 5th grade outcomes</td>
<td>Number sense, Numeration and Operations and Algebra are most predictive of 5th grade outcomes</td>
</tr>
<tr>
<td>26% of variability is due to the open ended items</td>
<td>43% of variability due to the open ended items</td>
</tr>
<tr>
<td>74% of cases correctly classified</td>
<td>81% of cases correctly classified</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Content</th>
<th>Direction of predictor*</th>
<th>Direction of predictor*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistics and Probability</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Geometry</td>
<td>+</td>
<td>Algebra</td>
</tr>
<tr>
<td>Number Sense, Numeration and Operations</td>
<td>+</td>
<td>Statistics and Probability</td>
</tr>
<tr>
<td>Algebra</td>
<td>+</td>
<td>Measurement</td>
</tr>
<tr>
<td>Geometry</td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>

*Direction based on Standardized Canonical Discriminant Function Coefficients*

Overall, scores on open-ended questions at grade 3 and grade 4 in all subcategories were significant predictors of overall proficiency level outcomes at grade 5, (p<.05). The standardized canonical discriminant function coefficients indicated a strong relationship between the independent variable of open-ended response questions and the dependent variable of proficient or non-proficient for both grade 3 and 4; 26% and 43% respectively of the variance between groups and predictors can be accounted for by the discriminant function.
CHAPTER 5: DISCUSSION

The purpose of this study was to broaden the growing body of knowledge pertaining to student achievement in mathematics by investigating the predictive ability of students’ responses to open-ended, constructed/extended questions in third and fourth grade mathematics content sub-categories on subsequent fifth grade mathematics achievement proficiency levels. Open-ended/constructed response questions were based on New York State mathematics testing items in third and fourth grade and included content specific categories of statistics and probability, number sense and numeration, geometry, algebra, and measurement as outlined by the National Council of Teachers of Mathematics (2007).

The research utilized archival data from third, fourth and fifth grade students during the 2003-2009 academic school years in one public school in New York State. Data were collected via state and school building testing outcome reports and consisted of test results for approximately 330 students. Through a statistical discriminant analysis, students were classified as proficient based on scores of 3 or 4 or non-proficient based on scores of 1 or 2 for overall grade five test results, after completing open-ended items on mathematics tests in grade three and four. The analysis specified areas of content for test subscales that may predict mathematics proficiency level in grade five. This chapter summarizes the major findings of the study and discusses them in context of implications for theory and practice.

Summary of Results

This study adds to current literature of the growing body of knowledge pertaining to student achievement in mathematics by investigating the importance of open-
ended/constructed response questions and their use in mathematics education. It provides evidence that open-ended constructed response formats may have an important relationship to student outcomes on achievement tests in mathematics. More specifically, it suggests that if performance on open-ended response questions is able to predict performance on students’ future mathematic achievement, educators should consider how this type of testing format informs curriculum and methodology at the elementary mathematics level (Schoenfeld, 2002).

Overall, the data analysis in this study suggested that open-ended, extended/constructed response questions in grades three and four do predict proficiency grouping on future fifth grade mathematics test. All content constructs (grade 3 and 4 statistics and probability, grade 3 and 4 number sense, numeration and operations, grade 3 and 4 geometry, grade 3 and 4 algebra and grade 4 measurement) appeared to be important when attempting to discriminate between the proficiency levels in fifth grade student outcomes. More specifically, in third grade, statistics and probability as well as geometry were the strongest predictors of fifth grade proficiency level. These constructs require students to understand mathematics through the collection, organization, displaying, and analyzing data. Students make predictions that are based upon data analysis and understand and apply concepts of probability (NYSED, 2010, p. 26). In the case of geometry mathematics is realized through visualization, spatial reasoning and analyses of characteristics and properties of geometric shapes. Additionally students formally and informally identify and justify geometric relationships, apply transformations, and use symmetry and coordinate geometry to analyze problem solving situations (NYSED, 2010, p. 26). In grade 4 number and numeration and algebra are the most predictive sub-
categories. Students understand numbers through multiple ways of representing numbers, relationships among numbers, and number systems, glean meanings of operations and procedures and how they relate to one another and compute accurately to make reasonable estimates when dealing with number sense and numeration (NYSED, 2010, p. 26). In algebra, students represent and analyze algebraically problem solving situations; perform algebraic procedures accurately; recognize, use, and represent algebraically patterns, relations, and functions (NYSED, 2010, p. 26).

These constructs are supported in classrooms through activities focused on specific bodies of knowledge and mental processes associated with each sub-category area and the grade level concepts. While the content may look similar across grade levels the metal processes use at each grade level change and become more sophisticated as the grade level increases. In comparing the third and fourth grade predictive value of the subcategories, the differences may be due to the specific content taught and the level or depth of the concept within each category.

Overall, however, outcomes supporting the predictive nature of open-ended questions add to researchers’ contention that the use of open-ended constructional formats improves the development of conceptual understanding and builds long-term conceptual learning (Gray & Tall 2002). Additionally, it hints at a need to address the issues of testing preparation and the impact of testing on classroom instruction and the discussion of what constitutes in-depth understanding in mathematics learning, how it should be supported, and how it is practiced. (Moon & Schulman, 1995; Kuechler & Simkin, 2010) and supports research that suggests the whole curriculum can be framed with an awareness of the abstract, evoking complex thinking skills (Gray & Tall 2002;
Martinez, 2007). While these analyses may be far from inferring support for the development of more complex conceptual thinking and mathematical awareness in classrooms, the first step in this direction is the investigation of the connection and explanation of the relationship between open-ended question and student achievement.

**Implications for Theory**

The ability to communicate mathematics effectively is an important outcome of mathematics education. According to social-constructivist theory, mathematical understanding develops through communicating ideas with others (Cobb, Boufi, McClain, & Whitenack, 1997, p. 36). Researchers argue that many students pass state mandated tests without really understanding the questions they answer by using test wise skills rather than knowledge of mathematics (Rosnick & Clement, 1980). Students then may advance to the next grade level without being grounded in mathematics awareness and conceptual understanding (Chauvot, 2006).

The theoretical assumptions underlying this study viewed the learning of mathematics as an active construction of meanings by making sense of mathematics problem situations, combined with the development of inquiry, reasoning, explanation, justification, argumentation, and intellectual autonomy, conceptual understanding and the active role of students in their own learning (Duffy & Cunningham, 1996; Deubel, 2003). Such active construction can improve students’ meta-cognitive processes within specific situations as learners become aware of their general academic strengths and weaknesses, the cognitive resources they can apply to meet the demands of particular situational tasks as well as their knowledge and skill about how to regulate engagement in tasks in order to optimize learning processes and outcomes (Lave & Wagner, 1991). As research suggests,
open-ended items induce these cognitive strategies (O’Neil & Brown, 1998), and invite a wide range of solutions and solution methods.

Results of the analysis allow researchers to look further into the idea of construction of learning and the theory behind social construction within mathematical contexts through the paradigm of open-ended formats as well as the ideas behind formative assessment as it relates to teachers’ support of students understanding of mathematics. The analyses in this study define a set of variables that exhibit a strong relationship between mathematical knowledge, meaning and the interaction students engage in while problem solving and the type of questions students are able to answer. While traditionally test outcome are summative in nature, the uses of these type of test results also can be perceived as formative, allowing teacher to construct knowledge about their students and to provide solutions based on that knowledge through instructional design and classroom activities in environments are highly dependent on the quality of the mathematical contexts offered by a teacher (Kajander, 2010). The emphasis here is on discourse, dialogue and the quality of classroom talk for construction on knowledge (Hodgen, 2007).

Although the results of this study are limited in scope, they do contribute to the theories behind construction of knowledge and an investigation into the relationship between conceptual and procedural understanding, both for understanding student outcomes and teacher understanding of students’ mathematical knowledge. Outcomes highlight the relevancy of open-ended, constructed response questions to the design of curriculum for construction of knowledge and allow for further speculation for their use as predictors of success in mathematics.
Implication for Practice and Research

Standardized tests have changed the pace and content of instruction, where relentless drill practice for students is instilled and instruction focuses on test content or test-taking skills while ignoring subject areas that are not addressed on the test (Ballard & Bates, 2008). Within the scope of this change, researchers continue the discussion of the impact of the high stakes testing on classroom methodology and curriculum design. Reflection on what is happening in the classroom as a result of testing highlights the impact of the reform movement on classroom instruction. The present research takes this focus to another level by looking at specific content variables of statistics and probability, number sense and numeration, geometry, algebra, and measurement and how they may impact testing outcomes.

Researchers contend that testing impacts how teachers teach by altering instructional practices and the philosophy behind methods they use. (Barksdale-Ladd & Thomas 2000; Jones & Johnston, 2002; Wideen et al 1997; Yarborough, 1999; Stodolsky, 1988; Mathison, 1987). According to Kazemi, (2002), assessment can be a source of insight into student learning, but it requires that we pay close attention to children’s thinking. The research behind this study of open-ended mathematics test items suggests that item format is not irrelevant to math achievement measured in later grades and can inform effective instruction of mathematics. For example, Silver (1993) realized that a larger number than expected students in his study could provide an appropriate interpretation to their computational answer on open-ended response formats if given a chance to explain their reasoning (Silver, Shapiro, & Deutsch, 1993). Though researchers suggest that constructed response items develop cognitive flexibility and in-depth
mathematical understandings (Sanchez & Ice, 2004), teachers are hesitant to make the time and expend the effort that is required when using strategies associated with these types of response questions. The present research suggests that the insight gained by looking at students’ answers on open-ended mathematics questions may in fact be well worth the time, allowing teachers to formatively assess student's strengths and weaknesses in order to develop an appropriate plan for the next steps, ensuring that students have a consistent, conceptually based lesson progression.

As previously mentioned, Kazemi (2002), motivated in part by his work with teachers and how teachers understand and make use of children’s mathematical thinking in making pedagogical and curricular decisions, looked to students to explain their thinking using open-ended formats. Additionally, Silver et al. (1993) realized that a larger number than expected students in his study could provide an appropriate interpretation to their computational answer on open-ended response formats if given a chance to explain their reasoning. These perspectives allow testing to be viewed in a positive light, suggesting that information from achievement tests can impact teachers practice by directly assessing test results to improve instruction (Sanchez & Ice, 2004). If open-ended questions encourage productive construction of knowledge, through newly developed teaching methods and they are also reasonable predictors of later math achievement, then they may be very useful for focusing mathematics curriculum at early elementary grade levels. For example, as Cooney, Badger, and Wilson (1993) suggest, teachers could produce assessment items that provided insight into student understanding and develop conceptual awareness instead of spending more time teaching to the test,
focusing on test-taking skills through drills. (Abrams et al., 2003; Moon & Schulman, 1995; Smith, 1991).

A more in-depth look at the analysis of open-ended items revealed specific content variables as better predictors of outcomes in fifth grade than others. As outlined in the previous chapter, third grade student predictor variables ranked statistics and probability and geometry as having the strongest relationship to fifth grade outcomes. According to NYSED (2009), statistics and probability emphasized the collection, organization, analysis and the development of predictions, as well as application of concepts of probability. Rosenstein, Caldwell and Crown (1996) state that the reason statistics grew as a branch of mathematics was to provide tools that are helpful in analysis and inference in situations of uncertainty, in order to measure high-inference mental skills or abilities. Researchers also suggest that these methods support higher order thinking skills which help develop mathematical awareness and conceptual understanding (Rosenstein, Caldwell, & Crown, 1996). The analysis of for this study at third grade for the sub category of statistics and probability, looks specifically at the reading of graphs, and the alternative views one might see when using graphs. When comparing this to the 4th grade, the level of mathematics understanding increases as students might have to create the graphs and place specific labels in appropriate places using more inferential skills than in the previous year. The idea of open-ended items promoting higher order thinking skills is not a new one, and although this research suggests that the item formats can help informed instructional methods, the specific content strands and the reason behind one predicting outcomes more than another still needs to be investigated in more depth.
In order to understand how children answer questions, the research needs to investigate individual questions and answers as well. One suggestion might be to look more closely the questions themselves and compare the levels of conceptual knowledge and awareness needed to answer them correctly. Specific content areas in both 3rd and 4th grade demonstrated a significant relationship to fifth grade outcomes. What needs to be determined is what it is about those specific content areas that help promote conceptual understanding?

According to Anderson, Reder, and Simon (1997), to promote effective learning it is essential to analyze question responses in detail focusing on the particular procedures and concepts to be learned and in order to provide students with instruction and examples that help them learn the component skills and understandings. The discriminant analysis of open-ended questions pointed to third grade geometry as important for predicting outcomes at fifth grade. According to NYSED (2009), geometry students used visualization and spatial reasoning to analyze characteristics and properties of geometric shapes, identified and justified geometric relationships and applied transformations and symmetry to analyze problem solving situations, all considered a higher level of task by most researchers. In 1988, Tishler found that there was too much emphasis placed on formal symbolism and naming in the elementary school geometry curriculum, while relational understanding and informal deduction were underemphasized. Geometric ideas, however are now investigated at increasing levels of sophistication for thinking about and justifying ideas (Mistretta, 2000), thus allowing for a more developed conceptual framework supporting higher-level thinking skills. Since both of these specific content
areas are exhibiting predictive value for success, it suggests specific relationships that can be investigated for supporting fifth grade outcome.

At the fourth grade level, predictor variables of number sense and algebra were the more significant predictors of fifth grade outcome. However, as suggested above, there is limited research on how and why specific content strands of open-ended items have more of a relationship to future testing outcomes. Additionally, predictor variables of statistics and probability and measurement in fourth grade predict fifth grade outcomes in a negative direction. The results suggest the development of new questions based on whether these are confounding variables and how their relationship to the other independent variables impacted the analyses. How do the content-specific categories differ, what does a question in each content area look like, and what is the constructs relationship to conceptual understanding in mathematics? The answers to these questions with further support the ability of teachers to use test results formatively and to develop pedagogy that will help student increase mathematical understanding and conceptual knowledge. Despite the challenges related to defining higher-order thinking, in mathematics, researchers’, educators, administrators, and evaluators expressed agreement about the value of teaching them for developing conceptual capacity (Ennis, 1993; Glaser & Resnick, 1991; Haladyna, 1997; Kauchak & Eggen, 1998) and that, looking at the subscales of results, educators can glean another level of understanding based on specific content areas where students thinking can be improved.

In a joint position statement in 2002, the National Association for the Education of Young Children and the National Council of Teachers of Mathematics emphasized the
importance of good early mathematics experiences for children. They noted that without
good early instruction, progress to higher-order skills is more difficult (NCTM, 2009b).
Based on the results of this study, further investigation needs to look at the utility of
open-ended response formats in the elementary classroom setting and what teachers can
do in their classrooms at the early grade levels to support students’ ability to understand
and successfully complete constructed response type items. Despite the fact that
important differences of opinion exist about the impact of achievement testing on
students’ mathematical understanding, the impact of constructed/extended response items
for future success in students’ overall mathematics understanding has been fairly ignored
as a part of the discussion.

Results of this study need to be examined looking further into the categories of
response items and why particular content areas are more predictive than others.
Examining individual content items can illuminate where mathematical knowledge is best
supported through the utility of constructed, extended, and open response question
formats. If, as the research suggests (Paris, Lawton, Turner, Roth, 1991; Shepard,
Doughterty, & Boulder, 1991), teachers’ instructional practices are impacted by the tests
their students are required to take, then using open-ended, constructed response type
questions could result in student outcomes supported by teachers’ practices that focus on
mathematics discussion, interpretation and conceptual knowledge development to yield a
commensurate gain in learning (Shepard, et al., 1991). The use of test results, through
conceptually rich explanations of answers facilitates learning because it brings important
student thinking into a public, metacognitive workspace where ideas can be shaped and
refined (Shemwell & Furtak, 2009). Discernible knowledge construction within
discussion-based formative assessment, particularly as this view applies to the learning of challenging mathematical concepts, provides the impetus and support for refining thinking (Sadler, 1989) and allows teachers to develop appropriate pedagogies for supporting the next steps in students’ learning. The importance of discussion and argumentation in classrooms during the development of concepts takes time but reduces the need for exercises that may previously have been introduced by teachers' examples demonstrating procedures. Changes in teaching pedagogy could shift mathematics away from procedural knowledge to that of mathematical thinking and profound understanding (Owens & Perry, 2001).

Many researchers hope to better capture the essence of student knowledge by allowing learners to express their thinking and problem solving strategies to show what they know with less emphasis on selecting the correct response. (Baker, O’Neil & Linn, 1993; Baxter & Shavelson, 1994; Herman, Aschbacher & Winters, 1992; O’Neil & Brown, 1998). Additionally, investigations into the specific content strands and their impact on student content knowledge development can be a next step in the research in multiple areas and has possible implications in several areas. First, it submits that if open-ended response questions predict performance on students’ future mathematics achievement then educators should consider how this type of format can be used to support student mathematical knowledge, including using open-ended constructed response questions for incorporating formative assessment as well as changes in instructional design of mathematics curriculum. Second, future research should consider investigating the extent to which work with extended constructed response type of activities contribute to greater understanding and higher performance in mathematics so
that students will acquire mathematical understanding and awareness. Finally, these results contribute to the field of assessment by documenting the value of open-ended, constructed/extended responses items on student overall achievement in mathematics.

Conclusions

The present research highlighted the predictive value of open-ended, extended /constructed response questions on overall achievement test results by categorical classifications: proficient and non-proficient. While one can propose that this outcome supports teachers in their search for ways to improve student performance and suggests developing curriculum that supports improvement in competing of open-ended, constructed response questions, it also opens a multitude of investigative tracks. According to results, prior performance on constructed response questions can predict group membership of proficient or non-proficient on later overall mathematics achievement test.

In support of the results of this study, a large body of literature on curriculum and instruction suggests that tests, quizzes, and exams are very important in terms of providing feedback to teachers, thus they play a critical role in the assessment system and the type of curriculum teacher embrace (Van de Walle, Karp, & Bay-Williams, 2010). Additionally, open-ended items reveal, to a certain degree, a perspective on mathematical awareness and conceptual understanding and support better insights into how test results can inform teachers’ curriculum design in order to improve students’ mathematical knowledge. Research by Kazemi (2002) supports this viewpoint stating that in his studies, students’ responses to open-ended items provided further evidence that the
interpretation of the problem situation provides significant insight into their solutions, whether correct or not.

Examining the outcomes of student mathematics testing, and its relationship to specific test item formats and content area is necessary in order to allow us to study thinking perspectives as a stable platform for understanding new mathematical ideas and situations and the classroom practices that support them (Franke, Carpenter, Levi, & Fennema, 2001). While it is difficult to develop conclusions regarding the impact of testing on student achievement without addressing teacher effects, influences of intra-classroom heterogeneity, student achievement level, and class size, etc., the results of this study suggest that the development of curriculum using pedagogy that supports the utility of constructed/extended response methods of questioning may be an important vehicle for improving student achievement. Findings such as these raise important questions about children’s mathematical understanding, its relationship to testing and instruction, and merits further investigation.
REFERENCES


http://www.statisticssolutions.com/. Retrieved from
http://faculty.chass.ncsu.edu/garson/PA765/discrim.htm
Gelman, R. (1986). Toward an understanding-based theory of mathematics learning and
instruction, or, in praise of Lampert on teaching multiplication. *Cognition and
Instruction*, 3(4), 349-355.
Education, 50*(1), 29-55.
Evaluating state and local initiatives. *Theory into Practice: Educational
Evaluation: An Evolving Field, 30*(1) pp. 61-68. Taylor & Francis, Ltd.
Technology Education, 7*. Retrieved from
*Handbook of research for educational communications and technology*, (pp. 665-
692). Washington, DC: Association for Educational Communications and
Technology.


Elementary Mathematics Teacher


development for mathematics and science education: A synthesis of standards.
*NISE Brief*. Madison, WI: National Institute for Science Education.

professional development for teachers of science and mathematics*. Thousand

Alliance, TERC. Cambridge, MA. Retrieved from


Practical application and interpretation (4th Ed.).* Los Angeles: Pyrczak.

Metz, K. E. (2004). Children's understanding of scientific inquiry: Their
conceptualization of uncertainty in investigations of their own design. *Cognition
and Instruction, 22*(2), 219-290.


instruction, and curriculum in elementary mathematics*. Portsmouth, NH:
Heinemann.

processes in early mathematics: The Australian pattern and structure


Schoenfeld, A. (1985b). Metacognitive and epistemological issues in mathematical understanding. In E. A. Silver (Ed.), *Teaching and learning mathematical


