8-1-2003

Mathematics Instructional Methods and Their Effects on Student Mathematics Scores

Ruth Huovinen

MATHEMATICS INSTRUCTIONAL METHODS AND THEIR EFFECTS ON STUDENT MATHEMATICS SCORES

by

Ruth Huovinen
Bachelor of Science, University of North Dakota, 1971
Master of Education, University of North Dakota, 1992

A Dissertation

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Grand Forks, North Dakota
August
2003

This dissertation, submitted by Ruth Huovinen in partial fulfillment of the requirements for the Degree of Doctor of Philosophy from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

Marjorie A. Bock
Chairperson

Richard D. Landry

Donald K. Lemon

This dissertation meets the standards for appearance, conforms to the style and format requirements of the Graduate School of the University of North Dakota, and is hereby approved.

Joseph N. Benoit
Dean of the Graduate School

July 18, 2003
Date

PERMISSION

Title	Mathematics Instructional Methods and Their Effects on Student Mathematics Scores
Department	Teaching and Learning
Degree	Doctor of Philosophy

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work or, in her absence, by the chairperson of the department or the dean of the Graduate School. It is understood that any copying or publication or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my dissertation.

Signature Ruth Huovinen

Date July 14, 2003

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The writer expresses sincere appreciation to the many people who contributed to the success of this study.

Appreciation is expressed to Dr. Marjorie Bock for her encouragement throughout the preparation of the study. Thanks also to Dr. Richard Landry for his assistance with the statistical interpretation of the data.

Thanks are given to Dr. Paul Johnson, Mr. John Salwei, Mrs. Ellen Knudson, and the Bismarck Public Schools for assistance with the study.

The writer wants to honor the love and support her family and friends by dedicating this work to them.

My two sons, Michael, and his wife, Suzi, and Andrew, my two sisters, Ruby and Mary Ellen, along with countless friends and co-workers were always the cheerleaders providing encouragement for me to continue. My parents, George and Ellen, although gone from this earth, continue to provide a sense of strength for me.

My school children for the past 32 years have challenged me to be the best I can be for them. Teaching is a noble and humble profession that allows a person to touch the future through the lives of today's children. May it always remain that way.

ABSTRACT

Mathematics instruction is undergoing an evolution in the United States. The National Council of Teachers of Mathematics (NCTM) advocates a change from the previous emphasis on computational skills to a greater emphasis on problem solving. The pre-service training most teachers completed did not include extensive exposure to strategies of this type. For that reason, school districts may choose to utilize staff development programs to assist teachers in acquiring these new strategies. This study examined one such staff development program, Developing Mathematical Ideas (DMI), and its effects on the mathematics achievement scores of the fourth grade students in a school district. This study examined the students' mathematics achievement scores to identify changes that might have occurred as a result of the staff development.

The study examined data from the Comprehensive Tests of Basic Skills (CTBS) mathematics achievement scores of 16 elementary schools for each year from 1999 to 2002. There were approximately 800 students included in the assessments in each of those years. Schools having 50% or more of their staff involved in the training sessions were identified as schools with trained teachers. In addition, scores of low socioeconomic schools, schools with more than 25% of the students receiving free and reduced lunch, were studied to discover any significant changes in their scores after the staff development sessions. Because the composite mathematics scores might not reflect changes in specific skill areas affected by the staff development, the eight subscores on the mathematics assessment were examined separately.

Although the composite mathematics scores for the district as a whole improved from 1999 to 2002, results indicated that the comparison of the composite mathematics

scores for schools with trained and untrained teachers did not show a significant difference. The mathematics scores for the low socioeconomic schools also did not indicate a significant difference. Five subtest sections increased during that time, but, again, the subscores of the schools with trained teachers were not significantly different than the schools with untrained teachers.

CHAPTER I

INTRODUCTION

"Math wars" are not new to American education, but rather have cycled through reform discussions since the turn of the century. Crisis events such as World War II and the space race have brought heightened criticism of mathematics education. Even in times of relative calm there have existed differing voices on how much mathematics should be taught and how it should be taught (Grouws & Cebulla, 2000). As Grouws and Cebulla state, throughout history discussions of reform focused on two points: first, applications of mathematics and the extent to which mathematics is linked to other content areas in the curriculum, and second, "learning with meaning" and the degree to which instruction focused on teaching mathematics meaningfully.

At the turn of the century, math was viewed as a means of developing mental discipline and as a necessary component of a classical education (Grouws & Cebulla, 2000). Thorndike's view of behaviorism (early 1900s) questioned the idea of transfer of learning. It was no longer believed that learning mathematics would strengthen the brain and result in improved reasoning in other content areas. Mathematics became an isolated drill and practice subject. It was believed that this drill and practice would build bonds, or connections, within the brain and create accuracy in mathematics. Accuracy had value in the pre-computer world of business. The bonds that were developed through repetition could then form an organized hierarchy of habits to help solve novel problems (Thorndike, 1970).

When World War II heightened the nation's awareness of the need for trained scientists and engineers, mathematics education was again examined. Through the use of technology such as nuclear power, war became more than human muscle. Educated scientists were needed to create the machines of war. Mathematics, viewed as a prerequisite for engineering, was needed to become educated. Learning mathematics became patriotic, something done for the "good of the country" (Grouws & Cebulla, 2000). After the war, the emphasis on mathematics lessened and Thorndike's emphasis on drill and practice returned.

Beginning in 1940, the voice of William A. Brownell (1928) was heard in mathematics discussions. He disagreed with Thorndike and stressed the need for mathematics to teach the meaning of why an algorithm worked. He did not believe in the mere mindless repetition of procedures (Kroll, 1989). His voice, however, was drowned out when Sputnik shot into orbit in 1957. The United States had to catch up with the Soviets. "New math" or "modern math" catapulted into classrooms throughout the nation. This "new math" of the 1960s and 1970s was based on set theory and focused on math language and symbols (Grouws & Cebulla, 2000). Students were taught foundations and abstract connections (Kroll, 1989). By 1973, when Morris Kline wrote *Why Johnny Can't Add: The Failure of the New Math,* there were widespread cries for a change in mathematics instruction. Kline (1973) advocated making mathematics meaningful by using physical objects and real-life experiences. Other critics of that time focused on the need to return to the basics. The basics were defined as proficiency with algorithms. From the mid 1970s to the mid 1980s, mathematics instructional activities returned to paper-and-pencil computations (Grouws & Cebulla, 2000).

Although the National Council of Teachers of Mathematics (NCTM) was founded in 1920, they were not a major player in the early "math wars." By 1980, they had begun to take an active role. They advocated the idea that basics in mathematics should be about more than drill and practice and should include problem solving. They continued to introduce reforms with the 1989 *Curriculum and Evaluation Standards for School Mathematics* (National Council of Teachers of Mathematics, 1989) and the 2000 *Principles and*

Standards for School Mathematics (National Council of Teachers of Mathematics, 2000). The focus of mathematics as defined by the NCTM was real-life problem situations (Grouws & Cebulla, 2000).

Marilyn Burns (1989) saw the progression of changes in mathematics over these past decades as related to changing job requirements. Jobs in the pre-computer era required computational proficiency, hence the emphasis on drill and practice. As Burns noted, "The ability to reason with numbers, to judge reasonableness of results, and to make effective decisions based on numerical information, [and] . . . to use the numerical data to solve problems [is of great importance]" (p. 124). Certainly this has remained constant to date.

Mathematics Reform and Psychological Learning Theory

Leading learning theorists have influenced the evolution of mathematics curriculum at different times in history (Bergeron & Herscovics, 1990). At certain times, the focus has been on small segments of knowledge; at other times, the overall "big picture" has been the focal point (Burns, 1989, 1992) of mathematics education. Parental and societal expectations of student performance have likewise evolved throughout the years. "Traditional expectations were far different from the goals and methods of new programs" (Davis, 1990, p. 94). At one time, competence in mathematics was defined as accuracy in adding long columns of figures. At other times, students were expected to memorize lengthy formulas (Burns, 1989). Today students are asked to investigate problems, develop solutions, and defend their answers by explaining how the answer was formulated (Schifter, Bastable, & Russell, 1999a, 1999b). Kroll (1989) created a table to highlight the relationships between phases of mathematics education and psychological learning theories. She included the drill and practice, meaningful arithmetic, and new math phases. Thus, her table includes the major mathematics education phases based on processes emerging from different theories of educational psychology at various times in the twentieth century.

Table 1. Relationships Between Phases of Mathematics Education and Psychological Learning Theories.

Phase	Main Theorists	Focus	How Achieved
Drill & Practice (approx. 1920-1930)	Thorndike	Facility with computation	Rote memorization of facts and algorithms Break all work into series of small steps
Meaningful Arithmetic (approx. 1930-1950s)	Brownell	Understanding of arithmetic ideas and skills Application of math to real-world	Emphasis on mathematical relationships Incidental learning Activity-oriented approach
New Math (approx. 1960-1970s)	Bruner	Structure of the discipline	Study of mathematical structures Spiral curriculum Discovery learning

The drill and practice phase from 1920-1930 was a reflection of the behaviorist movement of that time. Edward Thorndike (1932) applied the theory of behaviorism to learning. He advocated identifying precise goals and measuring changes (Peltzman, 1998). Education was viewed as an efficiency-engineering situation. There was a direct connection between the *stimuli*, such as 8+4, and the *response*, 12. Combinations were presented in no particular sequence. For example, 3+4 was not followed by 3+5 or 4+3, but rather something completely unrelated like 6+5. All facts were learned by themselves, independent of all others (Brownell, 1928). Necessary skills were identified, taught, tested, and re-taught if needed.

In a classroom applying the drill and practice theory to mathematics activities, one would see fourth grade students using different sets of flash cards to learn addition, subtraction, multiplication, and division facts. These flash cards would be in separate groups, not integrated to show how an addition fact could be used to solve a subtraction fact. The focus would be on speed and repetition of the isolated facts. Timed tests were used to evaluate students.

The meaningful arithmetic phase from 1930-1950s reflected Brownell's approach to mathematics. Brownell (1928) saw mathematics as having two layers, number facts and the rationale of the number system. He saw that students could learn to repeat number names and number combinations, but be ignorant of the principles that underlie the number system. They failed to see mathematics as an instrument of thought. Brownell thought that facts should be presented with *meaning attached.* The mental process used in deriving the answer needed to be identified. Effective, more mature methods of thinking were introduced so that a child could move beyond immature methods such as counting to acquire an answer (Brownell, 1928).

If a person observed the meaningful arithmetic theory in a classroom, students could be using other facts to solve a new problem (e.g., using 4+3 to solve 3+4). Facts would be presented in number families (e.g., 6+4=10, 4+6=10, 10-6=4, and 10-4=6). In addition, students would be encouraged to see number patterns (e.g., patterns of 10 as in 7+5, 17+5, and 27+5) (Brownell, 1928). Students would be taught general procedures and left to apply these to new materials. Assessment would identify a student's transfer of a mathematical procedure, such as patterns of 10, to a new group of numbers (Grouws & Cebulla, 2000).

The new math phase from 1960-1970s reflected Bruner's (1966) work with learning. There was an increased focus on the foundations of mathematics and the underlying logical structure (Grouws & Cebulla, 2000). Knowledge became more than just repeating facts. It was "deriving the unknown from the known" (Olson & Bruner, 1996, p. 14). Bruner thought that much of the work of learning occurs within a social context. Discovery learning occurred as the students worked in groups on problems. Bruner's belief in building new learning on previous learning, a spiral curriculum model, strongly influenced new math. For example, the algebraic concept of an unknown quality, such as *n* *or x,* was introduced in the elementary grades and revisited each year.

When a person visited a new math class, students could be seen working on mathematics theory, such as base 2 or base 7 numbers. Notations and set theory were used to describe solutions (Grouws & Cebulla, 2000). Much of the information was presented in a college prep lecture format with students working in large groups. Examinations basically tested the student's ability to solve equations. There was a focus on how new math would further the design and development of computers (Grouws & Cebulla, 2000).

Since Kroll's (1989) original publication, mathematics education has undergone continued change. The following information, if added to her table, would update it to include the current mathematics phase and psychological theorists.

Table 2. Suggested Changes to Update Kroll's Table: Relationships Between Phases of Mathematics Education and Psychological Learning Theories.

Phase	Main Theorists	Focus	How Achieved
Inquiry Learning (1989-Present)	Vygotsky Piaget	Constructing understanding Achieving standards	Hands-on activities Manipulatives Group work Deriving meaning from hands-on exploration

The inquiry learning phase (i.e., from the late 1980s to date) reflects Vygotsky and Piaget's views of constructivism. In a constructivist classroom, students encounter new and varied learning experiences that challenge their previous understandings. These new experiences create *disequilibrium* (Bybee & Sund, 1982; Piaget, 1978). As the students work together on activities, they modify their prior knowledge and reestablish *equilibrium* (Wakefield, 1997). While Piaget emphasized the importance of creating disequilibrium, Vygotsky (1962) emphasized the value of discussion with other students: (a) to refine the understanding of an experience and (b) to provide a link between language and cognition.

If a person were to visit a classroom using inquiry learning, students would be working with partners or in small groups engaging in discussion and problem solving. Different groups would be using different objects to solve a problem. For example, if the students were trying to determine how many rows, with eight chairs in each row, they would need to seat 200 students, they might be using blocks, drawings, or real chairs to solve the problem. After working in small groups, they would gather to discuss their reasoning and solutions. Student work was often collected in portfolios and evaluated with scoring rubrics (Nunes & Bryant, 1996; Schifter, 1996).

In addition, Piaget's stages of learning (Elkin, 1969; Furth, 1970; Jacobs, 1984; Piaget, 1978; Sund, 1976) have influenced the development of standards-based mathematics curricula. In standards-based curricula, past learning is used to support new learning. For example, understanding of *conservation of number* is used to support the understanding of number symbols. Many inquiry learning programs utilize standards (e.g., the 1989 NCTM standards) as the foundation of their curriculum. Proponents of inquiry learning mathematics believe standards such as problem solving, data analysis, and application of mathematics to new situations are critical to participation in the changing society (Grouws & Cebulla, 2000; Kamii, 1990, Reys, 2001).

Inquiry Learning and the Bismarck School District

In the late 1990s, the Bismarck Public Schools searched for changes to their mathematics instruction. They reviewed the historical changes in mathematics and looked for a staff development program based on inquiry learning, the constructivist model, and the 1989 NCTM standards. Based upon their commitment to this inquiry learning philosophy, the Bismarck Public Schools selected a staff development training program called Developing Mathematical Ideas (DMI) (Schifter et al., 1999a, 1999b).

Purpose of the Study

This quantitative study explored the teaching of mathematics and the training of elementary teachers to teach mathematics. Specifically, it examined the effects of the staff development Developing Mathematical Ideas (DMI) on fourth grade student performance. Student performance was measured using the mathematics subtest of a standardized assessment, the Comprehensive Tests of Basic Skills (CTBS) (McGraw-Hill, 2001). It examined the mathematics scores of the general population as well as within subgroups of the student population.

The study answered the following eight questions:

First, was there a significant difference on the composite mathematics scores of the CTBS test given to fourth grade students over a four-year period, 1999-2002, during which time the district focused on mathematics staff development activities for the teachers?

Second, was there a significant difference on the composite mathematics scores of the CTBS test given to fourth grade students in low socioeconomic schools over a four-year period, 1999-2002, during which time the district focused on mathematics staff development activities for the teachers?

Third, were the composite mathematics scores of the CTBS test given to fourth grade students significantly different in schools with 50% or more of the teachers being trained in the district sponsored mathematics staff development activities over a four-year period, 1999-2002?

Fourth, were there differences between schools with trained and untrained teachers on the 2002 composite mathematics scores of the CTBS test given to fourth grade students?

Fifth, were there differences between schools with trained and untrained teachers on the 2002 composite mathematics scores of the CTBS test given to fourth grade students adjusting for the 1999 test?

Sixth, were there significant differences between any of the eight gain scores of the mathematics subtests on the CTBS test given to fourth grade students in 1999 and 2002?

Seventh, was there a significant difference between any of the eight gain scores of the mathematics subtests on the CTBS test given to fourth grade students in low socioeconomic schools in 1999 and 2002?

Eighth, was there a significant difference between any of the 1999 and 2002 eight gain scores of the mathematics subtests on the CTBS test given to fourth grade students in schools with 50% or more of the teachers participating in the district sponsored mathematics staff development training?

Definitions

Constructivism. An educational philosophy in which students actively and personally construct knowledge rather than receive it from others (Silver, 1990).

Drill and practice. An educational strategy that focuses on repeated practice of basic mathematics facts presented in isolation (Hiebert, 1990).

Inquiry learning. An intellectual process by which people generate and test ideas they find personally useful to explain the phenomenon and to predict a consequence of similar circumstances, develop meaning for themselves (McCollum, 1978).

Low socioeconomic status. This study relies upon the federal Income Eligibility Guidelines for Child Nutrition and Food Distribution Programs for Free and Reduced Meals to determine this economic status. A formula using the household size and income is used to make this determination. For example, in 2001-2002, a family of four making less than $32,653 would be eligible for reduced meals (North Dakota Department of Public Instruction, 2002).

Meaningful arithmetic. An approach to teaching arithmetic that stressed the meaning or understanding of why an algorithm works (Grouws & Cebulla, 2000).

New mathematics. A math curriculum from the 1960s that made a radical break from the previous methods of instruction. It taught the language and operations of set theory rather than algorithms (Case, 1996).

Number fact families. A set of three related numbers which have four addition and subtraction facts related to them (e.g., 2+4=6, 4+2=6, 6-4=2, 6-2=4) (Underhill, 1972).

Portfolio assessment. A collection of student work samples that provides the focus for assessment of learning and show growth (Allen, 1998).

Poverty. An economic term relating to the level of income for a family. Following the Office of Management and Budget's directive, the Census Bureau uses a set of money income thresholds that varies by family size and composition to detect who is poor. If the family's total income is less than the family's threshold, then that family and every individual in it is considered poor (U.S. Bureau of Census, 2002). In this study, this term is used interchangeably with the term low socioeconomic status.

Rubrics. Tools which delineate the evaluation for a task; typically there are several performance levels, from very low to superior, and there is a brief description of the performance expectations at each level (Ardovino, Hollingsworth, & Ybarra, 2000).

Standards-based curriculum. A mathematics program that (a) stresses the interconnectedness and understanding of skills; (b) presents materials in a coherent manner; (c) develops ideas in depth; (d) promotes sense making; (e) engages students; (f) motivates learning; and (g) includes the five content standards written by NCTM: algebra, geometry, measurement, number and operations, and data analysis and probability (Trafton, Reys, & Wasman, 2001).

Well-being of children. A description of the status of children which includes healthy development across cognitive, emotional, social, and health domains, in addition to the economic status of the family (U.S. Department of Health and Human Services, 1996).

CHAPTER II

REVIEW OF LITERATURE

Mathematics is one of the core curricular areas in elementary schools. For this reason, teaching children math has been a concern to educators for many years. Debate on teaching math has fluctuated from an organized curriculum that presents skills and ideas in an ascending hierarchy to practical, realistic learning. Is mathematics the memorization of math facts using flash cards, or is it application of those facts? What educational aspects influence mathematics instruction? This chapter will review literature that is connected to mathematics instruction in four main areas. First, the Developing Mathematical Ideas (DMI) staff development program will be reviewed. Next, effective models for staff development training will be detailed. Then, student assessment will be presented. Last, educational concerns of students living in poverty or identified as low socioeconomic students will be included.

Literature on the Developing Mathematical Ideas Staff Development Program

Developing Mathematical Ideas (DMI) evolved from the inquiry learning movement that was prominent in mathematics education. In the DMI program, Piaget and Vygotsky's theories of constructed learning were applied to teachers' analyses of student learning. The role of the teacher was not to present procedures and formulas, but rather to listen and examine the students' explanations, and to determine their understanding, or misunderstanding, of mathematical concepts (Schifter et al., 1999a, 1999b; Silver, 1990). According to Silver, Kilpatrick, and Schlesinger (1995), "Mathematics is learned through a process of communication, not just listening, but speaking" (p. 9). DMI strongly supported the role of communication and learning (Schifter et al., 1999a, 1999b). Consequently, DMI was a program of professional development that integrated the study of mathematics and the study of students' learning so that teachers forged connections between the two (Kilpatrick, Swafford, & Findell, 2001).

In a classroom that employed mathematics strategies learned in DMI, students were faced with situations where they examined the problem, separated the facts presented, and constructed meaning. While the students were learning mathematics, they needed to actively construct knowledge and thereby own it (Davison, Miller, & Metheny, 1999; Rowan & Bourne, 1994; Yackel, Cobb, Wood, Wheatley, & Merkel, 1990). Through the explanation of their thought processes, students displayed their understanding of the basic mathematics procedures as well as their logic in solving the problem. The thinking that was used to find a solution was an important part of assessing whether a child had mastery of a problem.

When using DMI, students and teachers engaged in active listening as they struggled to understand each other's thinking (Schifter et al., 1999a, 1999b). Taber (1998) viewed teacher listening to student reasoning and strategies as vital. She said, "As teachers listened to students' strategies and reasoning, they developed greater appreciation for the reasoning and problem solving abilities of students" (p. 19). The process of verbalizing or writing about their mathematical thinking created a link between language and cognition (Dixon-Krauss, 1996; Vygotsky, 1962). This interaction did not occur without practice; teachers needed to construct practices appropriate to these principles (Sawyer, 1995; Schifter, 1996).

DMI piloted this staff development program with 65 teachers from urban, suburban, and rural communities in Massachusetts, Connecticut, and Vermont. The piloting was conducted at Mount Holyoke College using two courses for practicing teachers. One course, Inquiry in Mathematics Education, was a DMI course for teachers who were new to the ideas involved in the mathematics education reform movement. A second course, Teacher to Teacher, was a DMI course for teachers who had been working for some time to implement teaching practices aligned with the National Council of Teachers of Mathematics (NCTM) standards. This project was supported with grants from the Dwight D. Eisenhower Professional Development Program or the National Science Foundation. The effectiveness of the DMI program on participants was evaluated by researchers using participant journal entries, writing assignments, portfolio analysis, and one-on-one interviews (Education Development Center, 2002).

The major goals of the DMI professional development were to help participants learn (a) to extend their knowledge of mathematics content, (b) to define and select mathematical objectives for their students, (c) to recognize key mathematical ideas with which their students are grappling, (d) to develop strategies to support children's mathematical thinking, (e) to appreciate the power and complexity of student thinking, (f) to ask questions that help students deepen their mathematical understanding, (g) to analyze a piece of curriculum and identify the mathematics students will learn, (h) to make more mathematical connections for themselves, enhancing their ability to help their students do so, and (i) to continue learning about children and mathematics (Education Development Center, 2002; E. Knudson, personal communication, December 11, 2002; B. Livermont, personal communication, December 11, 2002).

Since inquiry learning teaching strategies differed considerably from previous teaching strategies, the Bismarck Public Schools district wanted to prepare the teachers for these changes (E. Knudson, personal communication, December 11, 2002). As Bay, Reys, and Reys (1999) said, "Few kinds of change are more challenging for teachers than changing the curriculum and teaching methods they use" (p. 503). Teachers involved in the DMI staff development attended four-hour class sessions every other week throughout the nine-month school year. These sessions occurred from 8:00 to 11:30 a.m. or 12:30 to 4:00 p.m. During each session, one chapter of Developing Mathematical Ideas, Number and Operations, Parts 1 and 2 (Schifter et al., 1999a, 1999b) was discussed. Between sessions, teachers read the next chapter and completed assignments related to the previous session. For most teachers, this approach represented a paradigm shift from the drill and practice method they were taught as children, as well as during their pre-service teacher training, to an inquiry-based model of teaching (E. Knudson, personal communication, December 11, 2002). Many teachers, especially on the elementary level, had a limited background in mathematics, particularly on the theory of mathematics. Ellen Knudson, the Bismarck Public Schools district math facilitator, said, "The single most important factor contributing to student learning in math is the teacher's understanding of math" (personal communication, December 11, 2002). Consequently, Bismarck Public Schools administrators certainly hoped DMI staff development would address these issues.

Throughout DMI staff development, the teachers were encouraged to construct their own personal knowledge base of mathematical principles. A sample chapter presented four or more case studies of children working on a specific mathematics principle, such as addition. During the class period, the teachers discussed the case studies to identify the mathematical thinking that the students were using to solve the problems. Most of the sessions also included a videotape of a master math teacher modeling a lesson. In addition to reflecting on the readings, the teacher participants completed student interviews and sample lessons. There were a total of 64 class hours plus homework hours during the course of the yearlong staff development (E. Knudson, personal communication, December 11, 2002).

Literature on Staff Development

> Our [education's] biggest long term problem is we have no way of getting better, no mechanism built into the teaching profession that allows us to improve gradually over time. (Stigler & Hiebert, 1997, p. 14)

In his study of constancy and change in schools from 1880 to present, Cuban (1988) saw change as belonging to one of two orders. First order change was aimed at correcting deficiencies, such as low test scores. Second order change altered the fundamental way schools were put together (e.g., reorganizing grade levels into a middle school). Staff development is the "mechanism" schools use to promote "first order" change. It addresses three key factors essential to first order change: (a) changes in the materials teachers use, (b) changes in teachers' beliefs, and (c) changes in how teachers teach (Cuban, 1997).

How do schools effectively introduce change in mathematics education? Are there ways to introduce change that make it more likely to occur? To answer these questions, a review of the staff development literature base was conducted (Cooper, 1989). The literature presented in the following pages resulted from five literature review search strategies: (a) computer search of abstract data bases, (b) manual search of abstract data bases, (c) computer search of citations index, (d) review of papers written by others, and (e) communications with an expert in school administration. Two key discussions of staff development organization surfaced, *Standards for Staff Development* (National Staff Development Council, 2001) and "Best Practice in Professional Development for Sustained Educational Change" by Speck (1996). These two will be discussed in detail in the following pages.

The National Staff Development Council Standards

In the research the National Staff Development Council (NSDC) has conducted, it found "the most powerful forms of staff development occur in ongoing teams that meet on a regular basis, preferably several times a week, for the purposes of learning, joint lesson planning, and problem solving" (National Staff Development Council, 2001, p. 1). The NSDC has identified 12 standards to guide successful staff development. These standards fall into three broad categories: (a) *context*, (b) *process*, and (c) *content*. In developing these categories, NSDC considers *context* standards as the structure proceeding the actual staff development sessions when a school would be choosing a program and getting ready to do the training. NSDC classifies the format of the staff development sessions as well as the attitudes toward change being introduced as *process* standards. *Content* standards are viewed by NSDC as the attention paid to what was learned from the staff development and implemented in the weeks and months following the training.

Context standards (National Staff Development Council, 2001) for staff development that improve the learning of all students include (a) learning communities: organizes adults into learning communities whose goals are aligned with those of the school and district,, (b) leadership: requires skillful school and district leaders who guide continuous instructional improvement, and (c) resources: requires resources to support adult learning and collaboration. These *context* standards could be thought of as the foundation of successful staff development. For instance, it would do no good to research and select a mathematics staff development program and then lack the resources to purchase the materials needed for the implementation of the program. In addition, these standards also highlight the value of a large majority of the staff being involved in the training to provide a large base of support for the change. Borman et al. (2000) found similar evidence in their review of models of school improvement. One of the key elements contributing to success was "buy-in" by teachers. The National Staff Development Council (2001) notes that of particular importance is the need for good leadership, a captain of the ship who knows where the ship is going (Speck, 1996). Pinks and Hyde (1992) highlight that same thought: "Administrator understanding and sustained support are critical to staff development" (p. 260).

Process standards (National Staff Development Council, 2001) for staff development that improve the learning of all students include (a) data-driven: uses disaggregated student data to determine learning priorities for participants, monitor progress, and help sustain continuous improvement; (b) evaluation: uses multiple sources of information to guide improvement and demonstrate its impact; (c) research-based: prepares educators to apply research to decision making; (d) design: uses learning strategies appropriate to the intended staff development goal; and (e) learning: applies knowledge about human learning and change to staff development. *Process* standards are the activities that occur during staff development. As staff participate in these activities in the training, they sift through what they are learning and accept or reject that information. Teachers rarely accept everything presented to them. They have to wrestle with the ideas and philosophies in a new program. This might be viewed somewhat like Piaget's principle of *disequalibration* (Jacobs, 1984). Teachers have to take what they already know and adjust it to accept what is being presented in the staff development, thus creating an ownership of the material. Lovitt, Stephens, Clarke, and Romberg (1990) identify similar keys to effective staff development. They recognize the need for teachers to feel a substantial degree of ownership in the staff development program and to make a commitment to the program. Teachers who feel ownership of a program are then willing to work through difficulties to implement the material. If teachers are convinced that implementation of the materials presented in staff development will make a difference for students, they will be more likely to put into practice those ideas (Borman et al., 2000).

Content standards (National Staff Development Council, 2001) for staff development that improve the learning of all students include (a) collaboration: provides educators with the knowledge and skills to collaborate; (b) equity: prepares educators to understand and appreciate all students, create safe, orderly, and supportive learning environments, and hold high expectations for all students' academic achievement; (c) quality teaching: deepens educators' content knowledge, provides them with research-based instructional strategies to assist students in meeting rigorous academic standards, and prepares them to use various types of classroom assessments appropriately; and (d) family involvement: provides educators with the knowledge and skills to involve families and other stakeholders appropriately in the schooling process. *Content* standards might be seen as what happens after training sessions and during the implementation of what was learned in those staff development sessions. Teachers who have internalized the philosophies of the material presented in the staff development are then eager to apply it in their classrooms. This application might represent only minor changes in philosophy or, as in the case of DMI, significant adjustments in thinking about learning. As Cuban (1988) stated, "I was repeatedly struck by the willingness of these teachers to alter routines and try other approaches if it met *their* criteria for classroom change" (p. 88). Teachers want to see students succeed and are willing to try new techniques if they believe it will lead to student success.

Of the 12 standards, 3 standards (research-based, learning, and quality) apply to this study and are discussed in further detail. First, the research-based standard highlights the need for thorough study of available materials prior to selecting a program for implementation. When the charisma of a speaker becomes the foundation for staff development, "the fad du jour often does not live up to its promise of improved teaching and higher student achievement" (National Staff Development Council, 2001, p. 6). Teachers have often listened to a compelling speaker at a session in September but failed to implement any of the ideas presented by the following March. Teachers know only too well that innovation after innovation has been introduced in schools. An overwhelming number of these innovations have disappeared in the blink of an eye (Cuban, 1988). There is usually at least one staff member known to say, "I remember when we tried that, it didn't work." While the term research-based is overused, careful review of research claims made by advocates of a particular program can identify well-founded research claims. Teachers recognize a well-developed program and view it as deserving their time and attention (Sparks & Hirsh, 1997; Speck, 1996).

Second, the learning standard focuses on improving teachers' content area knowledge base or instructional strategies or both. Through staff development activities, teachers have ample opportunity to both deepen their understanding of subject content and learn new techniques for teaching. In organized discussions, teachers can clarify their understanding of content. In addition, the staff development should provide time to practice using new skills and techniques with feedback on their performance (National Staff Development Council, 2001). Just listening to the techniques used in a new program is not sufficient. The teachers need to have opportunities to practice the techniques, to build consensus regarding their worth, and to establish a support network (Wagner, 2001). Thus, their understanding develops through active learning, discussion, and problem solving.

Third, the quality teaching standard identifies the need for staff development that encourages teachers to apply their deep understanding of the content areas they teach, use appropriate instructional methods, and utilize various classroom assessment strategies. This might be viewed as the final stage of staff development. At this point, teachers would come to understand what needed to be done, know how to use the techniques effectively and efficiently, recognize when to collaborate with others to solve problems, and know how to include families in the learning experience. Based on this understanding teachers would fully implement the staff development goals and activities.

Best Practice in Professional Development for Sustained Educational Change

In a second key discussion of staff development, Speck (1996) provides a good visual model of staff development. (Refer to Figure 1 to see Speck's model.) This model has five components: (a) needs assessment, (b) focus, (c) planning, (d) professional development activities, and (e) evaluation. According to a leader in the field of staff development, Dr. Angela Koppang (personal communication, May 7, 2003), Speck's model has strongly influenced the field of staff development. There are other similar models, but hers is a good visual representation of the stages involved in staff development.

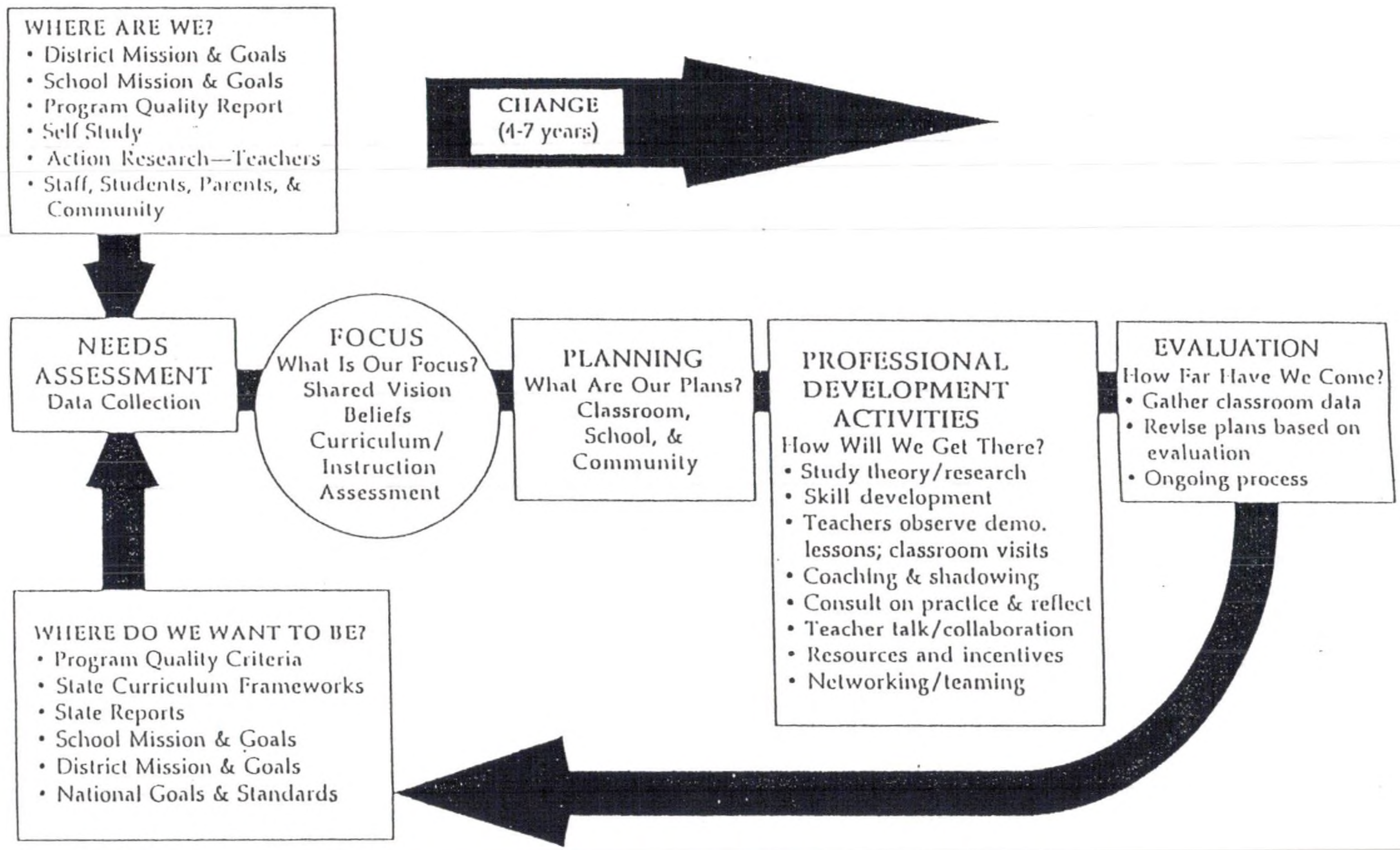

Figure 1. Speck's Model: Essentials of Best Practice in Professional Development for Sustained Educational Change.

Note. From "Best Practice in Professional Development for Sustained Educational Change," by M. Speck, 1996, *ERS Spectrum,* Spring, p. 34.

For effective staff development, schools need to clearly identify their mission and goals (Speck, 1996). They ask themselves “Where are we?” and “Where do we want to be?" To answer these questions data are collected. By utilizing a *needs assessment* and collecting data, the school determines what needs are being met and what needs remain unmet. Through discussions the school develops a shared vision or *focus*. From the vision, *plans* or goals are made. The overriding concept behind all the goals is increasing student learning. This forms the basis for all decisions. Student learning, according to Speck, is the fundamental job of schools. *Staff development activities* help the schools get from where the schools are to where they want to be. These activities might include coaching or shadowing, collaborating, skill training, or receiving consultation. *Evaluation* takes place as these activities occur to help schools assess if they are reaching their goals or vision. Sometimes schools have to revise their plans based on the information gathered while evaluating their work. The entire process is circular with evaluation yielding information that leads to future goal setting.

Speck (1996) lists several key factors that contribute to successful staff development. First, the school, not a district, is the primary unit of change. Ideas that are generated by the building staff are focused on their needs and resources. This develops a feeling of ownership that supports successful staff development, rather than a feeling that “someone else” decided what was needed in the building (Lovitt et al., 1990). In addition, Speck’s remarks also agree with Sparks and Hirsh (1997) that staff development needs to be an ongoing process, not a one-shot approach. In her opinion, significant change takes anywhere from four to seven years.

Speck, in agreement with Pinks and Hyde (1992), also stresses the importance of the principal. Administrative support can encourage or extinguish educational changes. Administrators control the flow of money in the building and can appropriate additional funds to implement a program. In addition to the administration and staff, other stakeholders (e.g., parents or children who currently attend the school or former students) should help to define the mission of the school, set goals for change, and contribute to policies and practices connected with change. The more people involved in the life of the school and its plan for change, the more opportunities there will be for shared support. Often when parents or students appear uncooperative they simple do not understand the school's vision and goals.

Speck (1996) and the National Staff Development Council (2001) highlight several key points from adult learning theory. First, adults as learners (e.g., staff development participants) are committed to learning when the goals and objectives are considered realistic to them. For example, if teachers can visualize stages of implementation of a new program, they will view that change as a process that is achievable given their time constraints. Second, adults want to see that staff development is related to their problems and concerns. If teachers do not believe that a problem exists, they will not see a need to implement any changes. This was particularly evident when schools and communities denied the presence of gangs. After admitting that the problem existed, solutions were developed to address them. And third, if staff help identify staff development needs, they will recognize them as realistic and not something being forced upon them by outside decision makers.

Adult learners have egos that can be harmed by judgment during learning (Speck, 1996). When learning new things, there is always a period of insecurity. (Remember learning to drive a car with your mother in the front seat?) Staff members need to feel safe to make mistakes or ask for additional assistance in using a new method. Providing small group activities to practice the skills they are learning gives teachers the support and encouragement to continue working with the concepts. Coaching and follow-up support provide additional assistance for transitioning learning back to the classroom.

The literature review of staff development (Borman et al., 2000; Cooper, 1989; Cuban, 1988, 1997; Jacobs, 1984; Lovitt et al., 1990; National Staff Development Council, 2001; Pinks & Hyde, 1992; Sparks & Hirsh, 1997; Speck, 1996; Stigler & Hiebert, 1997; Wagner, 2001) has shown that, through well-thought out staff development, schools can implement research-based information into their practice. They can identify their weaknesses, design solutions to meet these needs, and implement these solutions. By careful attention to the information presented by experts in the field of staff development in regard to planning and adult learning needs, schools are more likely to experience positive benefits from their staff development rather than investing in a passing fad.

Literature on Assessment

> Accountability emphasizes looking back in order to assign praise or blame; evaluation is better used to understand events and processes for the sake of guiding future activities. (Cronbach et al., 1980, p. 4)

When examining the power of assessment, Sirotnik (1994) noted the strength of achievement tests. He observed that politicians make judgments that effect the entire nation based on a slight change in average test scores. States pass legislation to address deficiencies when the achievement of students in one state is compared to the achievement of students in other states. Parents buy and sell homes based on student achievement. All of this is done based on the answers given to a few multiple-choice questions that were asked on one day during the school year.

How do schools effectively evaluate achievement in mathematics education? How does student performance influence selection of mathematics curriculum? Are there ways to compare learning to identify more successful teaching strategies? To answer these questions, a review of the assessment literature base was conducted (Cooper, 1989). This literature resulted from four literature review search strategies: (a) computer search of abstract data bases, (b) manual search of abstract data bases, (c) computer search of citations index, and (d) review of papers written by others. Seven references were identified as containing assessment information. They will be discussed in detail in the following pages.

Salvia and Ysseldyke (2001) group concerns relating to assessment into two categories: concerns of the general public and concerns of assessors. The general public asks two broad questions: Is the test fair? and Does the test ask valid questions? Of course, fairness is a relative term. What is fair to one person might not appear fair to someone else. (Remember sharing a cookie with your sister?) According to Salvia and Ysseldyke, when the public questions the fairness of a test there are five areas of concern: (a) opportunity to learn, (b) bias, (c) subjective scoring, (d) unequal treatment, and (e) unfair group comparisons.

First, was there a lack of opportunity to learn? For example, a child living in North Dakota would have almost no opportunity to learn about ocean tides. For that reason, an assessment including that information would appear unfair to the general public in North Dakota. Just because a child has not had the opportunity to learn something, such as ocean tides, it cannot be assumed that he or she cannot learn it (Hilliard, 2000). Opportunity to learn was listed by Newman and Beck (2000) as an important factor to be considered when examining assessment material.

The next major concern of fairness is bias. Salvia and Ysseldyke (2001) identify three areas of potential unfairness related to bias: (a) representation of individuals from diverse backgrounds, (b) representation of experiential background diversity, and (c) attention to language and concepts. Gender or race bias could exist in an assessment. The general public wants tests that give all the participants an equal opportunity to know the answers. For example, children living in North Dakota would have a difficult time answering questions with references to Jewish or Muslim traditions because that is not typically a part of their culture.

The third point of potential unfairness involves questions that are scored subjectively. This is why some people in the general public prefer multiple-choice questions. A well-designed multiple-choice question should have only one answer. The answer is either correct or incorrect based on behaviorist assumptions and theory. It is easy to defend a grade determined from that type of assessment. The general public looks for assessments that have the criteria explicitly stated. If a student is drawing a picture of the water cycle, there could be some subjectivity in scoring the drawing and interpreting what the child meant by certain parts of it. The general public might then question the grade assigned to that assessment.

Next, Salvia and Ysseldyke (2001) state that the general public asks if there has been unequal treatment. If one group of students has had accessibility to materials that another group of students has not had an opportunity to experience, it would seem unfair to expect both groups to be equal in the assessment. For example, because North Dakota is a rural state with significant differences in resources, some schools offer three foreign language options for students and other schools offer none. If the vocabulary section of a standardized assessment included several words of Latin origin, that might unfairly favor students from large schools who had the opportunity to take Latin.

The final area that Salvia and Ysseldyke (2001) list as a concern of the general public is unfair comparisons between groups of students. One group of students might be made up of students with many disadvantages while another group might have many talents. To then give both groups the same assessment and compare their scores might be seen as unfair by the general public. For example, some schools in North Dakota have a higher concentration of students of poverty. One school in this study had 60% of the students who qualified for free and reduced lunch, the criteria used to identify students as being in poverty in this study. Another school in the study had only 5% of its students who qualified for free and reduced lunch. The general public might view comparing their achievement scores as unfair.

In addition, Salvia and Ysseldyke (2001) also see the general public concerned with the face validity of the assessment. Does the test really test what it is supposed to? Does a test created to be taken by fourth graders actually test what fourth grade students were taught? Hilliard (2000) states that aligning the curriculum goals that were supposed to be taught, the textbook and materials used, and the standardized test administered is an almost impossible task. Add to that the curriculum that was actually taught (teachers do not always teach what is supposed to be taught) and it becomes even more difficult. As states work to bring alignment to their standards, curriculum, and assessment there is some evidence that this is resulting in higher achievement scores. Fuhrman (2001) identified two states, Texas and North Carolina, whose National Assessment of Educational Progress (NAEP) scores have shown significant and sustained gains between 1990-1997. She points out that those same two states are also leaders in alignment of standards, curriculum, and assessment.

Hilliard (2000) saw standardized assessment in schools separated into two categories, IQ tests and achievement tests. Although he questioned the validity of both types of tests, he particularly objected to the IQ tests. Achievement tests, especially criterion referenced ones, usually indicate skill achievement. The goal should be to monitor student progress and plan accordingly (D'Agostino, 2000). For example, an achievement test might indicate that the child has mastered subtraction of basic facts but has not mastered regrouping. That information could be used to develop learning objectives for the students. When achievement tests are used properly they can screen for areas that need additional attention and assist teachers in planning lessons. Intelligence tests, particularly group intelligence tests, are used unfairly to sort students into "intelligent" and "not so intelligent" groups (Hilliard, 2000). He questioned the face validity of a test containing multiple-choice questions being able to predict a person's future intellectual functioning.

The general public knows that data, especially objective sources of data like standardized tests, carry a heavy weight (Badal, 2002). Decisions that have a long-lasting impact are sometimes made on the basis of one test. These decisions may effect future options in life. The public rightly questions the fairness and validity of these tests.

The professionals who work with assessment on a daily basis have additional concerns. According to Salvia and Ysseldyke (2001) certification boards are concerned with establishing standards and assuring qualified persons are given licenses. One would hope that assessors administrating standardized tests would be adequately trained in the administration and interpretation of them.

Concerns of these assessors are more technical. In addition to the concerns of the public, they have a greater awareness of the construction of an assessment instrument and more carefully examine its construction. When assessors inspect a test, according to Salvia and Ysseldyke (2001), they have four major concerns: (a) accuracy, (b) generalizability, (c) meaning, and (d) utility. Assessors ask the same questions of fairness that the general public asks when examining the accuracy of an instrument in regard to areas such as bias. Assessors want assessments that accurately represent reality. Within standard measurements of error, the assessment should produce the same results if given at a different time.

Assessors examine an assessment instrument for generalizability (Salvia & Ysseldyke, 2001). The assessor looks for generalizability in (a) larger domains, (b) other times, and (c) other settings. Obviously a test cannot ask all of the possible questions related to a skill; therefore, it will never assess the entire domain. For example, if a test were assessing a student's knowledge of addition facts, it would contain a sampling of possible addition problems. The assessor would examine the instrument and ask if those questions were representative of the range of questions related to that skill. If there were only 2 questions involving regrouping, out of a group of 20 questions, the generalizability of that instrument would be in question.

Generalizability to other times is another goal of good assessment. Assessors want to have the instrument accurately indicate within standard measurements of error similar results if the test were administered again within a few days, provided no additional instruction occurred (Salvia & Ysseldyke, 2001). For example, an instrument might indicate that a student did not understand the concept of addition of fractions with unlike denominators. If no additional teaching took place, a similar assessment should yield similar results, within a short time span.

Salvia and Ysseldyke (2001) point out that assessors also look for generalizability to other settings. A well-administered assessment should indicate a student's ability to consistently do similar problems in a similar manner. If a student can demonstrate mastery of a skill such as division using two-digit divisors in the testing setting, assessors would expect the child to also be able to complete that task in the classroom. If the test were given under circumstances where the child felt very intimidated and unsafe (e.g., high-stakes testing situations), he or she might not answer many questions correctly. Under more relaxed situations, the student might have no difficulty with similar questions. Assessments that are given individually by competent assessors include time to establish rapport with the student. Group administered tests allow for very little interaction of that type.

The final area of concern for assessors in examining assessment instruments is the utility of the instrument (Salvia & Ysseldyke, 2001). Efficiency contributes to the utility of an instrument. Efficiency refers to the speed and economy of collecting data. Most standardized achievement tests are given in a group setting, an efficient process. The administrators are often classroom teachers who follow the standardized script and administer the test following time limits and other concerns set forth by the manufacturer. This may be an efficient system of testing but it introduces a risk of test validity. The teachers administrating the test are guided only by their own personal set of ethics. There could be teachers who, either intentionally or unintentionally, do not follow the administration guidelines. In these times of high-stakes testing where financial incentives might be attached to student scores, there might be an increased possibility of unethical behavior (Kohn, 2000b).

The other area relating to the utility of an instrument, according to Salvia and Ysseldyke (2001), is sensitivity. Tests must be designed to detect small differences across groups. Tests include items that help distinguish between students. When Kilpatrick (2001) reviewed standardized tests he determined that their function was to rank order students, schools, and districts. Items were chosen from a wide enough range to sort students into above average and below average. For example, if a test had items that all the students answered correctly, or incorrectly, it would not yield any way to separate the students because everyone would have the same score.

As mathematics instruction was going through evolution, so too was testing, or assessment. The first uniform testing in schools was in 1845 in Boston (D'Agostino, 2000). The test was to be used as an instrument of reform. This instrument of reform gradually evolved into both professional and political reform. Professional reform advocates used testing to monitor student progress. Political reform advocates used testing to monitor and sanction or reward schools based on achievement comparison among schools (D'Agostino, 2000).

In 1974, Lyndon Johnson's War on Poverty created Title I programs and the requirement of testing for program evaluation (D'Agostino, 2000). This required evaluation was implemented to make schools accountable and created a three-tiered system of reporting. Local districts reported to a state agency, which then reported to a federal agency. These evaluations were summative, although schools were encouraged to use results for program improvement.

The 1994 reauthorization of Title I required standards-driven assessments. D'Agostino (2000) found that states were charged with setting standards that were (a) rigorous and coherent, (b) defined what students were expected to know and be able to do, and (c) focused on advanced skills. States were responsible for determining what constituted "adequate yearly progress" on the standards-based assessments and identifying three performance levels: (a) partially proficient, (b) proficient, and (c) advanced. Beginning in 1996, states were required to assess and report on the progress of all students (Salvia & Ysseldyke, 2001).

In recent years, there has been a movement toward authentic assessment of students. Authentic assessment could involve classroom activities, teacher observations, portfolios of student work, as well as teacher-made or standardized tests (Badal, 2002; Salvia & Ysseldyke, 2001). Good (2002) advocates the use of multiple forms of assessment. By having several forms of assessment, the magnitude of an individual item is diminished. The more data collected, the more accurate the picture of the student. For example, if a student's acceptance into a special program, such as gifted and talented, were based on only one piece of work, a drawing, it might not demonstrate his or her ability in other areas, like mathematics. A collection of student artifacts would be more representative of students' work, showcasing both "good" days and "bad" days and areas of strength and weakness. Sirotnik (1994) sees this same issue in comparing schools. He believes that both quantitative, standardized test scores and qualitative, learning experiences should be used to describe schools. Relying on only one standardized test, possibly given to only one grade level, does not produce an accurate image of the total school.

The review of literature on assessment (Badal, 2002; Cooper, 1989; D'Agostino, 2000; Fuhrman, 2001; Good, 2002; Hilliard, 2000; Kilpatrick, 2001; Kohn, 2000b; Newman & Beck, 2000; Salvia & Ysseldyke, 2001; Sirotnik, 1994) found that formal forms of assessment have existed in American schools for over 150 years. Questions of fairness and validity arise when examining tests because of the potentially enormous impact some assessments can have on the lives of students. The comparison of student achievement scores has resulted in political ramifications for schools. Assessments can, however, yield useful data for the teachers, administrators, and other school leaders.

Literature on Poverty

> The largest barrier to educational achievement is poverty. It is no exaggeration to argue that if poverty in this country had been eliminated during the past decade, even if no attempts had been made to improve schools, American education would have improved more than it has as a result of all the educational task forces and reform movements. (Drew, 1996, p. 121)

In his study of characteristics of genius, Armstrong (1998) identified a student's home life as the most powerful learning experience. He believed that even if schools could operate 24 hours a day, they could not match the impact of the all-important emotional bond between parent and child. Some students come to affluent schools from affluent families. They have never known hunger and have never worried about where they would sleep that night. Other students come to school with everything they consider important in their backpack because they have no idea where they will be in three days, or even three hours.

Do students living in poverty have unique learning needs? Can schools effectively evaluate achievement of students living in poverty? Are there ways to identify qualities of successful schools serving a high proportion of students living in poverty? To answer these questions, a review of the poverty, or low socioeconomic status, literature base was conducted (Cooper, 1989). This literature resulted from six literature review search strategies: (a) computer search of abstract data bases, (b) manual search of abstract data bases, (c) computer search of citation indexes, (d) references in books written by others, (e) communication with professionals who typically share information with me, and (f) topical bibliographies complied by others. Twenty-five references addressing these questions were identified. The data from this research are discussed in detail in the following pages.

This review of literature examined poverty in three dimensions: (a) characteristics of poverty, (b) the effects of poverty on assessment, and (c) traits of successful schools that serve a high proportion of students living in poverty. What is poverty? Poverty most commonly is defined as an economic term showing a comparison to the contemporary standard of living (Hernandez, 1997). Every year it is increased proportionally to the Consumer Price Index from the previous year (U.S. Department of Health and Human Services, 1996). The size of the family and the number of children under the age of 18 are used to calculate the poverty threshold for that family. In the year 2001, a parent and two children with a total family income of less than $14,269 was considered living in poverty (U.S. Bureau of Census, 2002). Eligibility in the free and reduced lunch program was the basis for identification as living in poverty for this study.

Although most commonly thought of as an amount of income, poverty can also be described in terms of occupational prestige, social class, socioeconomic background, and economic disadvantage (Land & Legters, 2002). For example, many garbage collectors earn an income that would be above the poverty level but would still be considered negatively in terms of status. Likewise, there are examples of individuals from backgrounds of poverty who become well educated, but continue to be associated with the low socioeconomic group because of language or habits. The writer's personal experience brought this concept home when a homeless boy brought his baptismal certificate to school and asked his teacher to care for it because he did not know where he would be sleeping or what would be happening, but he was confident he would know where to go retrieve the certificate if the teacher had it.

Established in 1935, Aid to Families with Dependent Children, AFDC, was the first government program created to care for children in poverty. In addition to an economic concern, AFDC also classifies children's well-being. Well-being of children includes healthy development across cognitive, emotional, social, and health domains in addition to the economic status. Overall, children on AFDC rose 68% from 1970-1995 (U.S. Department of Health and Human Services, 1996). The most recent census data (U.S. Bureau of Census, 2002, 2003) for the year 2001 identifies 16.3% of all people in the United States as living in poverty. During that same year, 15.8% of all children were living in poverty. This rate fluctuates but has been gradually declining since 1990. It is down from a high in 1959 of 26.9%. Within the classification of persons living in poverty is a subgroup of homeless people. Stronge (2000) reported 630,000 homeless students in 1997. Not all homeless students live in urban areas. He found one third of the homeless students in rural areas.

When Smith, Brooks-Gunn, and Klebanov (1997) looked at poverty they saw it defined by three characteristics: (a) time in life, (b) persistence, and (c) family structure. At certain times in a child's development, poverty might have a more significant effect on the child than other times. They believed that early poverty was especially detrimental because the early years form the experiential background for later learning. Armstrong (1998) and Salend (2001) also agreed with Smith et al. that poverty in early life, even prenatal, could affect nutrition, which in turn could effect brain development and subsequent academic achievement.

Persistent poverty continues for many years, perhaps an entire lifetime. Teachman, Paasch, Day, and Carver (1997) found that poverty that is long-term is more likely to create negative cumulative effects (e.g., chronic health problems) than poverty that occurs for a short period of time. The saying "Adversity builds character" is more accurate for temporary adversity. For example, if a person is temporarily unemployed for a few months there will be some difficulties. But, a person who has been unemployed for many years will have experienced many adversities. Popham (2001) developed a statistical formula for predicting the cost of raising a child. He calculated that the family cost for supporting a child (born in 1999) until age 17 would be $117,390 for a low-income family, compared to $233,850 for an upper-income family. He specifies that these expenditures are for necessities, not the additional frills such as trips the upper-income families might provide. Thus, a child living in long-term poverty experiences the cumulative effects of living with less.

A third characteristic of poverty according to Smith et al. (1997) is family structure. Some families in poverty have a supportive extended family who will contribute stability. Supportive parents or grandparents often supply additional financial support and child care for the children. In contrast, Newman and Beck (2000) describe a family without extended family support. The family they describe is probably more typical of a family living in poverty. This family included several siblings from several fathers, most of whom were not involved in the lives of the children. The mother had moved from one urban area to another and had no relatives in the area. The seven-member family was living in a couple of rooms in a homeless shelter. The children were having both academic and behavioral difficulties in school. Had this family remained close to relatives there might have been more support for the children.

Homeless children living in poverty present unique problems in the educational setting. Some people would like to believe that there are no homeless children in the United States but statistics would prove otherwise, as stated by Stronge (2000) when he identified 630,000 homeless students. The McKinney Homeless Assistance Act of 1987 and subsequent reauthorization was written to address some of these concerns (Salend, 2001; Stronge, 2000). Stronge (2000) identified three barriers to their receiving an education: (a) residency, (b) guardianship, and (c) student records. First, most schools require a residency to attend a school. "What is your address?" is the first question most parents are asked when registering their child for school. The next barrier is guardianship. Many homeless children are shifted from relative to relative without the legal paperwork to clearly identify guardianship. Obtaining school records is probably the most difficult barrier to overcome. Schools must require proof of immunization. Homeless children may not have current immunizations or the paperwork showing which immunizations they have had. In addition, the paperwork establishing special services is often difficult to locate. Stronge (2000) identified the McKinney Homeless Assistance Act of 1987 as helpful in alleviating some of these difficulties. This act removed the residency requirement for school admissions and provided grant funding for services. Nevertheless, homeless children continue to experience complications entering school.

Poverty and schools tend to collide in assessment. Much literature has been written questioning the validity and reliability of testing children living in poverty. The overriding sentiment in the literature is that low socioeconomic status "has an impact" on standardized test scores. The following pages will review the leading research related to this concept.

According to Popham (2001), "the kind of item on standardized achievement tests is the kind more likely to be answered correctly by children from affluent and middle class families than from low-income families" (p. 55). He lists several examples of items that would favor affluent and middle-class students, such as asking a question that requires the students to select, from a list of four choices, the plant that is not a fruit. Low socioeconomic families often do not purchase fresh fruits; therefore, this question would favor children from affluent and middle-class families. He then provides a rationale for why items of this sort might be on standardized tests. According to Popham (2001), achievement tests are designed to provide score-spread or distribution of students that will then be used to classify students. Because socioeconomic status is already established and relatively stable, it provides a framework for devising tests that will distribute students.

Kohn (2000a) likewise makes similar statements questioning the structure of achievement tests. Because large companies that publish standardized tests also sell teaching materials designed to improve scores, there is an economic incentive to continue to publish tests of the same design. In addition, Kohn (2000a) states that in every standardized test, 10% of the students will score in the top 10%; likewise, half will score above 50% and half below 50%. The design of the test creates this distribution. Students below the median might have an acceptable knowledge of the skill but miss only one more question than students above the median. He views standardized tests as "contrived exercises that measure how much students have managed to cram into short-term memory" (p. 316). Authentic assessment, on the other hand, would look for deep understanding. For example, rather than selecting the correct vocabulary word to complete a sentence the student would include the word in a sentence or story that he or she would write.

Popham (2001) has reviewed numerous editions of standardized tests and presented several sample questions to defend his assertion of bias toward socioeconomic status. One such item is from a sixth grade science multiple-choice question that asks what would be used to find out if another planet had mountains or rivers. The answers included (a) a pair of binoculars, (b) a telescope, (c) a microscope, and (d) a camera. Obviously, especially to middle- or upper-class students, the answer is a telescope. Those students might have a telescope, might have gone on a visit to an observatory, or might have watched a show on the Discovery Channel about planets. Children from low socioeconomic families might not have had any of those experiences. There might have been a discussion about telescopes in science class resulting in some of the disadvantaged students answering it correctly. However, Popham (2001) claims that, on average, more students from middle- to upper-class groups would get this question correct than students from low socioeconomic groups. Popham (2001) judged the highest percentage of items linked to socioeconomic status was in the language arts area at 65%. In his analysis, mathematics had the lowest percentage of items linked to socioeconomic status, 5%. Science and social studies both had 45% of their items linked to socioeconomic status.

Several other researchers found similar problems with standardized tests. Smith et al. (1997) found evidences of bias on standardized tests. They found that children living in poverty scored 6 to 9 points lower on various assessments including IQ tests, verbal ability tests, and achievement tests. In addition, Hodgkinson (2002) stated that knowing the household income and educational level of the parents of a child taking the NAEP could be used to predict over one half of the variance in the test scores. Finally, Hilliard (2000) observed that standardized tests measure a very narrow range of behaviors. He did not feel that they represented the real world, particularly for children living in poverty.

Popham (2001) goes on to identify specific factors present in affluent families in the United States that provide additional assistance for children from these families: (a) English language, (b) literacy materials, and (c) cable television. Most families in affluent homes speak English. Children in these homes hear words being used in the form that is used on standardized tests. They hear words with multiple meanings being used in various situations. For example, the word field could mean a farmer's field, a baseball field, or the field of work a person does. The third meaning is used less frequently and might be unfamiliar to children living in poverty.

The abundance of literacy materials in affluent homes might include magazines, books, journals, and newspapers (Popham, 2001). In addition to the availability of these materials, the students from affluent families observe their parents using these resources. Dining room discussions might center on a recent article in the newspaper. An easy access to transportation can provide the affluent child with the means to visit the library. Without a permanent address, access to libraries is more limited for children of low socioeconomic families.

While some people might question the educational value of television, Popham (2001) identifies that as a possible strength for affluent families. Access to cable television will include exposure to the Discovery Channel and the History Channel, both possible sources for answers to questions asked on standardized tests. In addition to television, other forms of technology might be added as having positive value for children from affluent families. Computers and Internet access can expose children to a wide range of learning opportunities. Even the computer games that children play have educational components to them that could provide added background when taking standardized tests.

Similar to Popham, Newman and Beck (2000) found disadvantages in homes of children living in poverty. They found a shortage of school supplies. Access to resources such as dictionaries or encyclopedias did not exist. There were limited reading materials available for the children. In addition, they listed what they considered the greatest handicap for children living in poverty, access to adult time. They described a typical home environment of a disadvantage child as being chaotic. The kitchen table, a very small space, was the only "conference" area available for homework help and was frequently occupied by several family members. Children competed with each other and the demands of household duties for limited adult time. Some people have the impression that people living in poverty do not work. Land and Legters (2002) stated that because of welfare reform the number of working parents has increased. In 1997, there was at least one working parent in 65% of the homes with children living in poverty. The parent in Newman and Beck's study arrived home after 6:00 p.m., making it difficult to provide homework help for the children. Because she had such limited time, she abdicated the responsibility for education to the teachers. In addition to limited time to work with their children, Bruchey (1997) found low socioeconomic parents also had limited time for parent conferences or time to participate in school activities. They also had difficulty with transportation. These factors further separated the family from the school.

In addition to difficulties with standardized tests, children living in poverty have other obstacles in school. The teachers in schools serving children of poverty are more likely to be uncertified (Lewis, 2001). Similarly, Dunn (2000) found that the schools with high levels of poverty had poorly trained teachers and limited money. Affluent schools had money to recruit "good" teachers. Because of high-stakes testing, intense pressure to improve achievement scores resulted in experienced teachers leaving low socioeconomic schools, the very schools that needed experienced teachers (Kohn, 2000b). The teachers knew, just like Land and Legters (2002), that high concentrations of students living in poverty negatively effected academic achievement. The teachers left to be associated with high performing schools.

Even though schools with high levels of children living in poverty have difficulties, there are successes. Research identifies several key factors in creating successful educational environments for students living in poverty. Hodges' (2001) and Reed-Victor and Stronge's (2001) information regarding effective teaching strategies will be used for the framework of this discussion. Eleven specific strategies relating to working with children in poverty will be presented.

Reed-Victor and Stronge's (2001) first strategy of effective teaching is to activate prior knowledge. Classrooms that group students into small working units foster discussions that can link new information to prior knowledge. As Vygotsky (1962) reminds us, this is critical for learning. During this time the teacher can assess for deficient areas. This is particularly critical with background vocabulary. Newman and Beck (2000) illustrate how a child's lack of vocabulary can interfere with school performance. The boy in their study did not know that Europe was a continent and therefore did not complete the assignment correctly. In a classroom activating prior learning, children who are working as a group might list all the facts they know about Europe before additional new information was discussed.

The next strategy that Reed-Victor and Stronge (2001) focused on was the use of the constructivist approach. Children create knowledge for themselves rather than passively receiving it from an expert. This exposes children to inquiry learning situations where they investigate and discuss their findings with other students (Hodges, 2001). Based on Piaget's research this encourages children to create learning. For example, students might manipulate coins to discover regrouping.

Another strategy advocated by Reed-Victor and Stronge (2001) was effective classroom environment. This strategy would involve most of the "management" of the students and organization of the classroom. Teachers who have procedures in place to manage the classroom make more efficient use of time. For example, when students know the agenda for the day, they can organize their materials and be prepared. Thus, if a teacher listed the agenda on the board, it would be available for students' use. Another reference to the need for an effective classroom environment was Newman and Beck's (2000) observation concerning the importance of structuring homework through use of assignment notebooks. In addition, Reed-Victor and Stronge (2001) emphasized the need for teachers to build sensitivity to poverty and homelessness. Building sensitivity might employ some of Salend's (2001) ideas for assisting homeless and poor students. He suggests helping families complete school forms, collaborating with personnel from shelters, breaking assignments into smaller chunks, and giving students some control over their learning.

Next, Reed-Victor and Stronge (2001) advocate quickly assessing student skills and needs for service and then implementing instructional supports. Necessary paperwork for special education service might not be available (Salend 2001) but that should not serve as an impediment to meeting the child's needs. Following an informal assessment of skills the school needs to begin assisting the student with learning tasks. These might involve specific special education services but could also involve the use of a volunteer or peer helper. For example, a child in the fourth grade who cannot read obviously needs special assistance. Individualized support could commence with follow-up paperwork as needed on a case-by-case basis.

In addition to educational needs, schools need to address basic physical needs, according to Reed-Victor and Stronge (2001). Health issues might prevent regular attendance. A common problem is lack of necessary eyeglasses. If a child cannot comfortably read, it is difficult to expect him or her to do so willingly. Poor nutritional habits contribute to many health and learning concerns of children. For that reason, free and reduced breakfast and lunch are supplied in schools to provide children the food needed to give them energy for learning.

Schools fostering psychological well-being was the next strategy listed by Reed-Victor and Stronge (2001). Kohn (2000b) was concerned that schools with high levels of poverty spent so much time preparing for standardized tests that important aspects such as building a sense of community were omitted. It is through those community-building activities that a shared vision for improvement can be formed. Students could become empowered and believe that learning is possible.

Reed-Victor and Stronge (2001) advocate fostering high expectations within a supportive climate. Material presented to the Council of Chief School Officers (2002) would also support that strategy. The Council of Chief School Officers found that successful schools in Texas shared several common traits: (a) The staff embraced the belief that all students can be academically successful, (b) administrators put talents and experiences of teachers to their best use, (c) faculty and staff regularly communicated and learned from each other, (d) staff used student assessment to identify areas needing improvement, (e) the culture of student-centered learning was predominate, (f) educators persisted in addressing academic barriers to learning, (g) parents were viewed as critical partners, and (h) referrals to special education were seen as a last resort.

Hilliard (2000) highlighted another way of describing educators' persistence when he said that hard work matters. Just trying one approach might not be successful. To be successful, teachers have to believe that students can learn and continue trying different methods. Kannapel, Aagaard, Coe, and Reeves (2001) found similar findings (i.e., as in Hilliard, 2000) in successful schools in Kentucky. They listed the characteristic of staff believing that all students could succeed and encouraging and assisting students as the key difference between successful and unsuccessful schools. They documented one school in their study with the highest percentage of poor students, 70% on free and reduced lunch, as attaining CTBS achievement scores well above the state average. Kannapel et al. credited those gains to the staff belief in the potential of the students.

The next strategy listed by Reed-Victor and Stronge (2001) for successfully working with children in poverty was fostering constructive peer relationships. Children in poverty often attend several different schools. They might not know how to make friends because they have not been in a school long enough to develop those bonds. Each new setting has different rules that require time to understand. For example, each classroom has its own procedures for using the bathroom or going to lunch. Setting up study buddies could foster friendships and improve academic achievement.

Both Hodges (2001) and Reed-Victor and Stronge (2001) encourage the use of authentic tasks and meaningful problem solving. For example, rather than using textbook examples, like buying supplies to make a kite, it would be more effective for children living in poverty to use an example of buying food. In addition to using authentic tasks for learning, Dunn (2000) also advocates using authentic tasks for assessment. Authentic assessment could include journals, portfolios, discussions, and interviews.

The literature review of poverty (Armstrong, 1998; Bruchey, 1997; Cooper, 1989; Drew, 1996; Dunn, 2000; Hernandez, 1997; Hilliard, 2000; Hodges, 2001; Hodgkinson, 2002; Kannapel et al., 2001; Kohn, 2000a, 2000b; Land & Legters, 2002; Lewis, 2001; Newman & Beck, 2000; Popham, 2001; Reed-Victor & Stronge, 2001; Seland, 2001; Smith et al., 1997; Stronge, 2000; Teachman et al., 1997; U.S. Bureau of Census, 2002, 2003; U.S. Department of Health and Human Services, 1996; Vygotsky, 1962) has shown that poverty is a multifaceted problem. Schools are obligated to implement strategies that will improve the academic success of children living in poverty. Resources are available through government programs to help schools with this task. Schools can help provide a stimulating learning environment that will enable all students to achieve their potential.

Summary

This chapter has reviewed the literature related to this study. It has examined factors contributing to successful staff development, the role of assessment in education, and poverty and its affect on learning. Through the staff development process, schools can identify problems, design solutions, and implement change. Assessment is an effective tool schools can use to help students be successful in schools, as long as faculty and administrators keep in mind the need for fairness and validity. Poverty affects many children in schools and presents special learning problems that can be overcome.

CHAPTER III

METHODS

This quantitative study explored the teaching of mathematics and the training of elementary teachers to teach mathematics. Specifically, it examined the effects of the staff development Developing Mathematical Ideas (DMI) on fourth grade student performance. Student performance was measured using the mathematics subtest of a standardized assessment, Comprehensive Tests of Basic Skills (CTBS) (McGraw-Hill, 2001). It examined the mathematics scores of the general population as well as within subgroups of the student population.

The study answered the following eight questions:

First, was there a significant difference on the composite mathematics scores of the CTBS test given to fourth grade students over a four-year period, 1999-2002, during which time the district focused on mathematics staff development activities for the teachers?

Second, was there a significant difference on the composite mathematics scores of the CTBS test given to fourth grade students in low socioeconomic schools over a four-year period, 1999-2002, during which time the district focused on mathematics staff development activities for the teachers?

Third, were the composite mathematics scores of the CTBS test given to fourth grade students significantly different in schools with 50% or more of the teachers being trained in the district sponsored mathematics staff development activities over a four-year period, 1999-2002?

Fourth, were there differences between schools with trained and untrained teachers on the 2002 composite mathematics scores of the CTBS test given to fourth grade students?

Fifth, were there differences between schools with trained and untrained teachers on the 2002 composite mathematics scores of the CTBS test given to fourth grade students adjusting for the 1999 test?

Sixth, were there significant differences between any of the eight gain scores of the mathematics subtests on the CTBS test given to fourth grade students in 1999 and 2002?

Seventh, was there a significant difference between any of the eight gain scores of the mathematics subtests on the CTBS test given to fourth grade students in low socioeconomic schools in 1999 and 2002?

Eighth, was there a significant difference between any of the 1999 and 2002 eight gain scores of the mathematics subtests on the CTBS test given to fourth grade students in schools with 50% or more of the teachers participating in the district sponsored mathematics staff development training?

Setting

The school district in this study is the Bismarck Public Schools, Bismarck, North Dakota. There are 16 elementary schools, 3 middle schools, and 3 high schools in the district. There are approximately 5,000 K-6 students. The average class size has slowly declined from a high in 1999 of 21.0 to 20.3 in 2002 (R. Hoffman-Walker, personal communication, March 18, 2003). The student body is predominately white middle class. For the past 10 years, there has been about 16% of the total school population who qualified for free and reduced lunch. The elementary schools range from a high of 60% free and reduced lunch to a low of 5% (Joersz, 2003). There may be low-income students who do not file the necessary paperwork to qualify for this program. The percentage of students receiving free or reduced lunch was used in this study to identify low socioeconomic status for the school.

In the past, the school district as a whole has scored well on the state achievement tests. In 1999, the district fourth grade composite mathematics national percentile score was

54.2. The highest composite mathematics percentile score for an individual school involved in the study was 72.8% in 2001. The lowest composite mathematics percentile score for an individual school involved in the study was 41.9% in 1999.

Because federal legislation holds schools responsible for all students to make academic progress, it is important to examine the effect the new teaching strategies might have on schools with a large percentage of low socioeconomic students. Low socioeconomic status can create barriers for some students. Drew (1996) says, “The largest barrier to educational achievement is poverty” (p. 1). Schools with a high proportion of low socioeconomic status often have corresponding low test scores. For the schools in this study, the 1999 average composite percentile mathematics score for the five low socioeconomic schools was 48.6. During that same year, the average composite percentile mathematics score for the other 11 schools was 56.7. Although there was some growth, the same pattern occurred in 2002 when the average composite percentile mathematics score for the low socioeconomic schools was 52.4 compared to the 59.7 composite percentile score in schools that were not identified as low socioeconomic.

Schools tested all students in the same standardized manner. Classroom teachers were not aware of students who received free and reduced lunch and made no special educational adjustments based on economic status. Two of the low socioeconomic status schools were school-wide Title I schools. Each of those schools had differing improvement plans that they have implemented to address their needs. The Department of Public Instruction identified those same two schools as failing to make adequate yearly progress in 2001 and 2002.

The school district in this study had central committees that evaluated different areas of the curriculum. The committee evaluating mathematics considered various materials currently on the market. Two summer mathematics workshops stimulated an interest in searching for a less traditional approach to mathematics instruction. The district in this study examined an emerging research-supported program focusing on constructing mathematical understanding (J. Salwei, personal communication, February 12, 2003).

After selecting an inquiry-based mathematics teaching program, it was determined that implementation of it would require extensive teacher staff development training to establish these new instructional strategies. The district decided to pilot a new format for professional development or staff development. Previous staff development programs were generally one of two main types, short after-school sessions or elective evening/weekend/summer courses. Concentrated district sponsored staff development offered during the school day had never been attempted in this district.

Participants

The district in this study operates 16 elementary schools of varying sizes. The unit of analysis in this study was these schools. During the 1999-2002 school years, there were approximately 5,000 K-6 students enrolled each year. State mandated standardized tests were administered to fourth grade students in March. Approximately 800 fourth grade students from this district completed the testing each year from 1999 to 2002. Individual student data were used to compute school scores. The average scores of the fourth graders attending these schools became the school scores that were then compared to answer the research questions. The students represented the general population of the district. Students with Individual Education Plans (IEPs) were tested in accordance with their plans and within the context of the allowable accommodations and modifications for this specific standardized test. It was assumed that all other students were tested in accordance with the assessment protocol. For instance, all elementary students must complete the full assessment battery within a specified two-week time frame. All of the fourth grade students present during the days of testing in the district were participants in this study.

Any teacher in the district was permitted to take the DMI staff development. This resulted in varying percentages of teachers within each school involved in the staff development. For purposes of this study, the schools were separated into two categories, schools with trained teachers and schools with untrained teachers. Table 3 shows the frequency distribution of the schools. Five schools with 50% or more of the classroom teachers trained were identified as schools with trained teachers.

Table 3. Frequency Distribution for Number of Schools With Percentage of Trained Teachers in the Schools.

Percentage of Trained Teachers	Number of Schools
0-10%	8
11-20%	0
21-30%	1
31-40%	2
41-50%	0
51-60%	0
61-70%	1
71-80%	0
81-90%	1
91-100%	3

All of the schools in the study had segments of their population identified as low socioeconomic. In the study, 5 schools with a four-year average of greater than 25% of their students qualifying for free and reduced lunch were identified as being low socioeconomic schools. The remaining 11 schools had a four-year average of less than 25% of their students qualifying for free and reduced lunches. During 1999, when 17% of the total school population was identified as low socioeconomic, the low socioeconomic schools averaged 43% of their students receiving free and reduced lunches. Likewise, in 2002 when 16% of the total school population was low socioeconomic, the 5 identified schools had

47.5% of their population receiving free and reduced lunches (Joersz, 2003). Table 4 presents the frequency distribution for percentage of students qualifying for free and reduced lunch in the schools. For purposes of this study, it was decided that schools that had more than 25% of the students receiving free and reduced lunch would be labeled as a low socioeconomic school. Five schools were identified as low socioeconomic schools.

Table 4. Frequency Distribution for Number of Schools With Percentage of Students Qualifying for Free and Reduced Lunch in the Schools.

Percentage of Students	Number of Schools
0-5%	0
6-10%	1
11-15%	4
16-20%	5
21-25%	1
26-30%	0
31-35%	1
36-40%	0
41-45%	2
46-50%	1
51-55%	1

Procedures

During the 2000-2001 school year, 11 teachers from one elementary school voluntarily enrolled in this initial stage of new staff development training program Developing Mathematical Ideas (DMI). In the summer of 2001, 13 people received leadership training in DMI. These 13 persons represented teacher/administrator groups from six elementary schools. The following year, 2001-2002, those persons conducted training in their participating elementary schools. During the 2001-2002 year, 100 staff completed the training program, representing 12 of the 16 elementary schools. Therefore, data were examined for change in 2002. Some schools had only one teacher attending the training, and some schools had the majority of their staff involved. Teachers of regular classrooms, as well as specialists, such as teachers of gifted and talented students and teachers of special education students, participated in the training (E. Knudson, personal communication, December 11, 2002; B. Livermont, personal communication, December 11, 2002).

Part of the analysis of the data was the comparison of the improvement in achievement scores at schools where the teachers were involved in the training to the improvement in achievement scores at schools of teachers who were not involved in the training. Schools with more than 50% of the staff who completed training were identified as schools with trained teachers. Five schools in the study were identified as schools with trained teachers, and 11 were identified as schools with untrained teachers. One of the 5 schools identified as trained was also identified as low socioeconomic. The teaching responsibilities of the staff members who were trained were not identified. Because only students in the fourth grade were tested, it is unknown how many of those students were in classrooms with trained teachers. No attempt was made to identify the level of implementation of the training. Some of the teachers who received the training might not have implemented it in their classrooms.

Developing Mathematical Ideas (DMI) was conducted under the supervision of the district specialist, Ellen Knudson. There were 13 people who provided leadership for the 2000-2002 staff development classes, all of whom had completed two weeks of training during the summer of 2000 or 2001. Four-hour training sessions were held about every two weeks. Because the training was held during the school day, substitute teachers were hired. This enabled the teachers to devote large blocks of uninterrupted time to examine the theory of mathematics. This mode of staff development training represented a dramatic change from previous training. The district had never devoted large quantities of paid leave for professional development. During each session, one chapter of Developing Mathematical Ideas, Number and Operations, Parts 1 and 2 (Schifter et al., 1999a, 1999b) was discussed. In the interim between sessions, teachers read the next chapter and completed assignments related to the previous session. The teachers were encouraged to construct their own personal knowledge base of mathematical principles.

A sample chapter would present four or more case studies of children working on specific mathematics principles, such as addition and subtraction. The students represented in the case studies spanned several grade levels. The teachers examined the students' comments to identify what formed the basis for their answers. When a child arrived at an incorrect answer, the teachers focused on understanding what led to that error, and correcting the thinking process rather than just saying that it was incorrect. During the class period, the teachers discussed the case studies and clarified with each other the mathematical thinking. The teachers had to explain why one student's process was correct, creating dialog and discussion, and sometimes disagreement, among the teachers.

Most of the sessions also included a videotape of a master math teacher modeling a lesson. Guided discussion followed each segment of the class. A sample discussion question might be "What does the student understand? What is missing? What are you confused about?" In addition to reflecting on the readings, the teacher participants completed student interviews and sample lessons. There was a total of 64 class hours plus homework hours during the yearlong staff development.

The focus of the staff development training was to deepen teacher understanding of mathematics. Many teachers, especially those at the elementary level, have had a limited background in mathematics, particularly on theory of mathematics. Elementary teachers are generalists. They are responsible for teaching literacy, social studies, and science, as well as mathematics. Pre-service training in mathematics instruction is often only one or two courses. For many of the teachers, this DMI training represented the first time in many years that they had spent a concentrated amount of time on analyzing mathematical thinking.

In this study, the 1999 CTBS student mathematics test scores were used as baseline scores. These scores represented achievement prior to any district organized interventions. Following Department of Public Instruction requirements, these tests were given to all fourth grade students during the first two weeks of March. It is assumed that the tests were given under standard conditions as set forth by the company. No data were collected to indicate the mathematics teaching techniques used by the teachers in 1999, but it is assumed that the district sponsored mathematics textbook was the basis of instruction.

The 1999-2001 state mandated CTBS (McGraw-Hill, 2001) and 2002 Comprehensive Achievement Test (CAT) (McGraw-Hill, 2002) student mathematics test scores were used to assess changes in achievement after district organized interventions were implemented. The CTBS *TerraNova, Second Edition* is constructed as a comprehensive modular assessment series of student achievement. It is designed and considered to be parallel and used interchangeably with the CAT (McGraw-Hill, 2002). The norm referenced test batteries are designed to measure basic skills of reading, language, and mathematics. The mathematics section complies with the NCTM standards (Educational Testing Service, 1986; McGraw-Hill, 2001, 2002; Murphy, Impara, & Plake, 1999).

One of the major focuses of inquiry-based learning is communication. Communication, both written and oral, is an integral component of the new mathematics methods taught in DMI. This communication focuses on problem solving and explaining answers to problems. The students work in small groups and discuss strategies to solve problems. Students individually write explanations to accompany their solution. Whole group discussions present solutions and pose further questions. Because this emphasis on communication might result in greater success on the communications subtest of the mathematics test, that subtest was analyzed. In addition to communication, the subtests (a) number and number relations; (b) computations and estimation; (c) operation concepts; (d) measurement; (e) geometry and spatial sense; (f) data, statistics, and probability; and (g) problem solving and reasoning were also analyzed to identify possible changes. The Objective Performance Index (OPI) school score was subtracted from the national scores in each of the eight subtests. The resulting difference was labeled gain score and was used to compare the subtests. The study compared each 1999 subtest with its 2002 counterpart for all schools, for low socioeconomic schools, and for schools with trained teachers.

Because the composite mathematics scores in the low socioeconomic schools were lower than the other schools, this study also examined the effect the new methods of mathematics instruction had on achievement. The scores of schools identified as belonging to the low socioeconomic group were compared for those four years to determine if their scores had improved.

Assessments

Standardized achievement tests, the CTBS *TerraNova* (McGraw-Hill, 2001) and the CAT *TerraNova* (McGraw-Hill, 2002), were used to assess student achievement. The CTBS *TerraNova* was administered in 1999-2001. It provided norm referenced and standard referenced information on the students. The mathematics section complies with the NCTM standards (Educational Testing Service, 1986; Murphy et al., 1999). Extensive validity and reliability tests were conducted on this test.

> TerraNova was designed and developed to provide achievement scores that are valid for most types of educational decision making. The primary inferences from the test results include measurement of the achievement of individual students relative to a current nationwide normative group and relative program effectiveness based on the results of groups of students. Progress can be tracked over years. (McGraw-Hill, 2001, p. 31)

This statement of validity is based on analysis of content, technical, and construct criteria.

> The TerraNova mathematics test was aligned with the NCTM Standards. More emphasis was placed on a balance of skills, concepts, knowledge, and problem solving than on procedural and computational processes . . . constructed-response items allow students to generate their own ideas. In addition, students use manipulatives to investigate and model concepts and situations. They explain their answers in writing and with diagrams, displaying their understanding of mathematical concepts and problem-solving strategies. (McGraw-Hill, 2001, p. 23)

This standardized test also provided OPI scores on the students. These scores were used to identify the gain scores of the students. The OPI indicated mastery of specific skills or objectives, such as measurement (McGraw-Hill, 2001). The difference between the national average and the school average on each objective provided the gain score for that objective.

In 2002, the statewide assessment changed to a similar standardized test, the CAT *TerraNova*, distributed by the same company. The CAT *TerraNova* provided similar statements of validity and reliability. It also stated that CAT *TerraNova* was designed and considered to be parallel and used interchangeably with other *TerraNova* tests. The tests contain the same format and measure the same objectives (Educational Testing Service, 1986; McGraw-Hill, 2002; Murphy et al., 1999).

CHAPTER IV

RESULTS

The study was conducted in the Bismarck Public Schools district with approximately 5,000 K-6 students. The district contains 16 elementary schools, all of which were included in the study. The 16 elementary schools varied in size from less than 125 to more than 400. The average class size has slowly declined from a high in 1999 of 21.0 to 20.3 in 2002. State sponsored, nationally normed achievement tests were given to fourth grade students, approximately 800 students each year. During the four years of the study, about 16% of the total school population qualified for free and reduced lunch (Joersz, 2003). The elementary schools ranged from a high of 60% free and reduced lunch to a low of 5%. Based on a four-year average of greater than 25% of the student body receiving free and reduced lunch, five schools were identified as low socioeconomic schools for this study.

The unit of analysis in this study was the 16 elementary schools, not individual students. Student achievement scores were combined to provide school scores. Because disaggregated student data were available only for 2002, the study was unable to identify individual student progress. Individual students also could not be tracked from year to year because only fourth grade students were tested each year. District data provided percentages of students within each school belonging to the low socioeconomic group but not individual names. For those reasons, it was decided to use school-wide scores as the unit of analysis in the conduct of this research.

Although teachers from all grade levels participated in the district sponsored staff development training, only fourth grade scores were available for this study because the district only tests fourth grade students. Some schools had only one teacher attending the training while other schools had the majority of their staff involved. Schools with more than 50% of the classroom teachers involved in the district sponsored training were classified as schools with trained teachers. See Table 3 for the frequency distribution for the percentage of trained teachers in the schools. There were 5 schools identified as schools with trained teachers and 11 schools identified as schools with untrained teachers. The data from the mathematics scores on the CTBS tests of 1999, 2000, and 2001 and the CAT test of 2002 were examined to answer the questions identified in this study. The rest of this chapter provides the analysis for each of the research questions posed.

The first research question. Was there a significant difference on the composite mathematics scores of the CTBS test given to fourth grade students over a four-year period, 1999-2002, during which time the district focused on mathematics staff development activities for the teachers? The composite mathematics scores were examined with a general linear model and repeated measures over time using MANOVA. Wilks' lambda was the multivariate statistic used to determine significance at the .05 level. The mean scores varied from 54.1 to 57.4. Wilks' lambda indicated a significant difference at p=.036. This result would indicate that there was a significant difference over a four-year period in the composite mathematics scores of the CTBS test given to fourth grade students. When the composite mathematics scores of all 16 elementary schools were compared for 1999, 2000, 2001, and 2002, there was an increase in these scores during this time when the district focused on mathematics staff development activities. Because the null hypothesis was rejected and there were more than three treatments, a post hoc test was run to identify where the significant difference existed. Tukey's Honestly Significant Difference (HSD) compared the individual treatments two at a time in a pairwise comparison. Difference of greater than 2.72 would be significant. Only the 1999 and 2002 scores were found to be significantly different. The results are provided in Table 5.

Table 5. Means, Standard Deviations, and MANOVA Results for Mathematics Composite Scores (1999-2002) for All Schools.

	Year			
Mathematics Composite	1999	2000	2001	2002
Mean NCE	54.2	56.6	55.3	57.4
SD	6.3	6.2	6.7	6.2

Wilks' lambda=.530; *df*=3,13; *p*=.036

The second research question. Was there a significant difference on the composite mathematics scores of the CTBS test given to fourth grade students in the low socioeconomic schools over a four-year period, 1999-2002, during which time the district focused on mathematics staff development activities for the teachers? The composite mathematics scores were examined using repeated measures tests, MANOVA and Wilks' lambda. Five schools in the district were identified as low socioeconomic schools based on a four-year average of more than 25% of students receiving free and reduced meals during the month of March. The mean scores varied from 48.6 to 52.3 over time. Wilks' lambda indicated no significant difference at the .05 level. There was not a significant difference in the composite mathematics scores of the five low socioeconomic schools during the four years that the district focused on mathematics staff development activities. These results are provided in Table 6.

The third research question. Were the composite mathematics scores of the CTBS test given to fourth grade students significantly different in schools with 50% or more of the teachers being trained in the district sponsored mathematics staff development activities over a four-year period, 1999-2002? The data for the schools with trained teachers only were analyzed with repeated measures tests, MANOVA and Wilks' lambda. There were five schools that had 50% or more of their teachers involved in the training. The mean scores

Table 6. Means, Standard Deviations, and MANOVA Results for Composite Mathematics Scores (1999-2002) for Low Socioeconomic Schools.

	Year			
Mathematics Composite	1999	2000	2001	2002
Mean NCE	48.6	50.7	49.4	52.4
SD	5.0	2.6	5.4	2.5

Wilks' lambda=.124; *df*=3,2; *p*=.180

varied from 55.6 to 58.9 over the four years. Wilks' lambda indicated no significant differences at the .05 level. There was not a significant difference over the four-year period. These results are provided in Table 7.

Table 7. Means, Standard Deviations, and MANOVA Results for Composite Mathematics Scores (1999-2002) for Schools With Trained Teachers.

	Year			
Mathematics Composite	1999	2000	2001	2002
Mean NCE	55.7	58.5	55.7	58.9
SD	2.4	5.6	3.7	6.4

Wilks' lambda=.184; *df*=3,2; *p*=.263

The fourth research question. Were there differences between schools with trained and untrained teachers on the 2002 composite mathematic scores of the CTBS test given to fourth grade students? Because the change might occur in only the fourth year after the teachers had been in training for one or more years, the data were also examined with ANOVA using the mathematics 2002 scores as the dependent variable looking for school level of training by time. The results are presented in Table 8. There was no significant difference in the composite math scores indicated on the tests that were conducted.

Table 8. Means, Standard Deviations, and ANOVA Results on Composite Mathematics Scores 2002 for Schools With Trained and Untrained Teachers.

School	Mean	SD
Untrained	56.7	6.2
Trained	58.9	6.4

F=.434; *df*=1,14; *p*=.521

The fifth research question. Were there differences between schools with trained and untrained teachers on the 2002 composite mathematic scores of the CTBS test given to fourth grade students adjusting for the 1999 test? In order to adjust for possible 1999 pretest differences, the fourth year mathematics scores were adjusted for the first year scores using ANCOVA (Analysis of Covariance). These results are provided in Table 9. There was no significant difference in the composite mathematics scores indicated on the tests that were conducted.

Table 9. Adjusted Mean and Standard Deviations and ANCOVA Results for 2002 Mathematics Scores.

School	Mean	SD
Untrained	57.2	1.3
Trained	57.8	1.9

F=.070; *df*=1,13; *p*=.796

The sixth research question. Were there significant differences between any of the eight gain scores of the mathematics subtests on the CTBS test given to fourth grade students in 1999 and 2002? To obtain the gain scores, the Objective Performance Index (OPI) school score was subtracted from the national scores in each of the eight subtests. Because the staff development stressed mathematics communication and problem solving, the subscores of the CTBS were examined to determine if these, or any other of the subscores, were significantly different.

The paired *t*-tests examined the eight subtests: (a) number and number relations; (b) computations and estimation; (c) operation concepts; (d) measurement; (e) geometry and spatial sense; (f) data, statistics, and probability; (g) problem solving and reasoning; and (h) communication. The differences between the 1999 subscores and the 2002 subscores of the fourth grade students were compared. See Table 10 for the results.

Table 10. Means, Standard Deviations, and Paired *t*-Test Results for Mathematics Subtest Gain Scores for Years 1999 and 2002.

	Year 1999		Year 2002			
Math Subtest	Mean	SD	Mean	SD	*t* value	*p*
Number and number relations	3.50	7.10	3.94	6.01	0.33	.748
Computations and estimation	2.38	5.38	5.63	6.57	2.67	.018*
Operation concepts	2.81	5.62	7.50	7.26	3.90	.001 **
Measurement	1.69	6.25	7.50	6.50	5.46	<.001 **
Geometry and spatial sense	4.25	6.50	6.31	5.24	1.67	.115
Data, statistics, and probability	2.94	5.62	3.75	5.25	0.69	.501
Problem solving and reasoning	2.94	6.55	6.63	8.07	2.60	.020*
Communication	1.94	5.78	5.94	6.66	3.06	.008 **

*Significant at the ≤.05 level; **significant at the ≤.01 level

There were significant differences on five subtest scores: (a) computations and estimation, (b) operation concepts, (c) measurement, (d) problem solving and reasoning, and

(e) communication in the year 2002. This would indicate that these subtests increased during the time that the district conducted math staff development training.

The seventh research question. Was there a significant difference between any of the eight gain scores of the mathematics subtests on the CTBS test given to fourth grade students in low socioeconomic schools in 1999 and 2002? Paired *t*-tests again examined the same eight subtests. The 1999 and 2002 subscores were used for the comparison. There was a significant difference on one subtest: (a) measurement. The results are provided in Table 11.

Table 11. Means, Standard Deviations, and Paired *t*-Test Results for Mathematics Subtest Gain Scores for Years 1999 and 2002 for Students in Low Socioeconomic Schools.

	Year 1999		Year 2002			
Math Subtest	Mean	SD	Mean	SD	*t* value	*p*
Number and number relations	2.40	7.30	0.80	1.92	0.49	.653
Computations and estimation	2.40	5.37	0.00	3.08	1.02	.366
Operation concepts	2.00	4.36	1.20	2.95	2.09	.105
Measurement	3.60	1.91	2.40	1.29	3.59	.023*
Geometry and spatial sense	0.80	4.97	1.80	2.39	1.02	.364
Data, statistics, and probability	2.20	5.45	1.00	4.00	0.39	.715
Problem solving and reasoning	2.00	4.69	0.00	3.79	1.32	.258
Communication	4.00	4.95	0.40	2.30	1.94	.125

*Significant at the .05 level

The eighth research question. Was there a significant difference between any of the 1999 and 2002 eight gain scores of the mathematics subtests on the CTBS test given to fourth grade students in schools with 50% or more of the teachers participating in the district sponsored mathematics staff development training? On the eight subtest scores, *t*-tests indicated there were no significant differences on any of the subtest scores of students in schools with trained teachers. The results are provided in Table 12.

Table 12. Means, Standard Deviations, and Paired *t*-Test Results of Mathematics Gain Scores for Years 1999 and 2002 for Schools With Trained Teachers.

	Year 1999		Year 2002			
Math Subtest	Mean	SD	Mean	SD	*t* value	*p*
Number and number relations	5.80	3.11	4.40	6.43	0.47	.663
Computations and estimation	3.80	1.48	7.80	6.94	-1.45	.220
Operation concepts	4.80	2.39	8.60	7.89	-1.22	.290
Measurement	3.60	1.82	9.60	7.16	-2.27	.086
Geometry and spatial sense	6.00	1.41	7.60	5.77	-0.76	.491
Data, statistics, and probability	5.00	1.41	5.20	4.15	-0.12	.908
Problem solving and reasoning	4.20	3.42	9.00	8.72	-1.32	.256
Communication	4.80	2.17	6.60	6.54	-0.79	.472

Summary of Results

The results of the study indicated that the composite mathematics scores of all the fourth grade students did increase over the four-year period that the district focused on mathematics staff development activities. This broad district-wide increase was not seen in the subgroups of low socioeconomic schools or the schools with more than 50% of the staff involved in the training, as those groups did not demonstrate a significant change in composite mathematics scores.

When the composite scores are separated into skill areas, the district-wide subtest scores of (a) computations and estimation, (b) operation concepts, (c) measurement, (d) geometry and spatial sense, (e) problem solving and reasoning, and (f) communications increased from 1999 to 2002. Several of these skills, such as communications and problem solving and reasoning, are emphasized in DMI staff development training. One subtest score, measurement, decreased in low socioeconomic schools. There was no significant difference in the gain scores of schools with 50% or more of the staff trained.

CHAPTER V

DISCUSSION, LIMITATIONS, AND RECOMMENDATIONS

Discussion

Statistical Results

This quantitative study examined the effects the staff development training program, Developing Mathematical Ideas (DMI) (Schifter et al., 1999a, 1999b), had on the Comprehensive Tests of Basic Skills (CTBS) mathematics scores of the fourth grade students in that district. The data collected showed that there was a significant increase, $p=.036$, in the composite mathematics scores of the students during the four-year time frame of the study. According to post hoc testing, the significant difference occurred between the composite percentile scores of 1999 and 2002. Thus, the district, as a whole, improved the student achievement on the composite mathematics scores. Comparison of the composite mathematics scores from the schools with trained and untrained teachers showed no significant difference. Further, an examination of the 2002 data showed no significant difference in CTBS scores between schools with trained and untrained teachers using mathematics scores as the dependent variable and looking for school level of training by time. In addition, the data were examined comparing fourth year, 2002, data and adjusting for the first year, 1999, differences. Consequently, this study was not able to claim that this improvement was related to the DMI staff development training. The results did not show a significant difference in the composite mathematics scores of schools with trained and untrained teachers while adjusting for 1999 differences. This study also investigated the effect of DMI training on the mathematics performance in low socioeconomic schools. There were five schools identified as having more than 25% of their students receiving free and reduced lunch, the criterion this study used to identify low socioeconomic status. When the composite scores of the low socioeconomic schools were compared to the other schools, the data did not show a significant difference at the .05 level.

Limitations of the Study

Obviously, as shown by comparisons of district-wide scores from 1999 to 2002, the district showed improvement in mathematics performance. However, this improvement does not appear to be directly tied to the DMI staff development curriculum. Thus, the study data must be interpreted with caution. Only 5 out of 16 schools were identified schools with trained teachers (i.e., having more than 50% of the teachers trained). This study did not identify which 50% of the staff completed the training. Since the CTBS assessment was administered to only fourth grade students in this district, it is possible that only a few fourth grade teachers from schools with trained teachers participated in the staff development. Further, in some of the buildings there could be sections of fourth grade classes using DMI strategies and other sections not using those strategies. In addition, since the staff development was available to any teachers, it is possible that fourth grade teachers from schools with untrained teachers might have participated in the training. In addition, the DMI strategies implemented in their classrooms could have affected the mathematics scores in the buildings with untrained teachers. Moreover, one building had been involved with DMI activities for two years so the students in that building could have been exposed to DMI strategies for both third and fourth grade; whereas, students in other buildings could have been exposed to DMI strategies for only six months. Because the CTBS data were reported as a composite score for each building, the study could not separate out the scores and link classrooms with their respective scores. Moreover, the level of DMI implementation across classrooms of teachers participating in the staff development may have been inconsistent. Training participation may not have led to classroom implementation. Thus, it was not possible for this study to evaluate the impact of DMI training on student performance. Finally, only one of the low socioeconomic status schools was also a school with trained teachers. Two of the other low socioeconomic status schools had several teachers in the training but not enough to qualify as a trained school. Thus, the effects of DMI staff development on the performance of students in low socioeconomic schools could not be effectively evaluated.

Reasons for No Measured Effect

It is unclear why the district is making progress improving the mathematics scores of the fourth students on the state sponsored nationally normed assessment. There has been growth. This might be a reflection of the extensive staff development training, although the data do not strongly support that hypothesis. Some of the possible explanations of findings in this study include (a) duration of implementation, (b) level of DMI implementation assessment, (c) evaluation issues, and (d) administrative support.

First, the duration of implementation of the DMI staff development was limited. As Speck (1996) reported, significant change may take four to seven years. A few of the teachers had been involved with DMI activities for two years, but the majority of the teachers had only been in training for part of the year. Cuban (1997) noted that implementation of new programs involves ". . . changes in the actual materials [teachers] use, changes in their beliefs, and changes in how they teach" (p. 27). This type of evolution occurs slowly; thus, the length of implementation is the greatest potential cause of the no-effect outcomes reported in this study.

Second, the level of DMI implementation across the teachers with training would have to have been limited. With the majority of the teachers receiving less than a year of training, it is highly unlikely that they were implementing DMI effectively and consistently. Again, Speck (1996) and Sparks and Hirsh (1997) noted that staff development needs to be an ongoing process, not a one-shot approach. Significant change takes anywhere from four to seven years.

Third, the assessment may have been flawed. The only form of assessment used in this study was a standardized test. The question might be asked, "Was the CTBS a fair and valid assessment of what students learned in the fourth grade classrooms using DMI mathematics strategies?" The DMI strategies might influence learning in ways that were not tested. In other words, there may be mathematics concepts affected by the new strategies that were not assessed by this standardized testing. If a standardized test is going to be used to evaluate DMI, or any other new program, a review of the questions on the assessment instrument should be performed to verify that they match what is reflected in the new strategies (Salvia & Ysseldyke, 2001). In addition, a different form of assessment might provide a different picture of the students' learning. Good (2002) advocated the use of multiple forms of assessment. By having several forms of assessment, the magnitude of an individual item was diminished. The more data collected, the more accurate the picture of the student. Consequently, reliance on one standardized assessment selected by the state department of education rather than by the school district and the use of only one measure of student performance is a second potential cause of the no-effect outcomes reported in this study.

Finally, administrative support may have influenced the no-effect outcomes reported in this study. Speck (1996), in agreement with Pinks and Hyde (1992), stressed the importance of the principal. They noted that administrative support could encourage or extinguish educational change. Principals' interest in DMI or lack thereof likely influenced teachers' motivation to use DMI or emphasized their mathematics instruction without DMI. Consequently, even teachers who had not participated in DMI staff development could well have improved their students' mathematics performance due to their increased emphasis on mathematics.

Recommendations for Future Research

This study supports the need for additional research regarding the DMI staff development program. Schools wishing to evaluate the impact of DMI staff development should design a study that clearly differentiates trained and untrained groups of teachers (e.g., 80% or more of the teachers in the school are trained for the trained group of teachers). They should compare student performance in these teachers' classrooms before DMI staff development with these teachers' student performance three to four years into DMI training. Furthermore, they should evaluate students' mathematics performance using multiple measures (e.g., standardized tests, teacher-made tests, daily learning activities). Finally, they should disaggregate their data to provide in-depth analysis of student performance. In addition, they should evaluate the DMI staff development by its effect on teachers' mathematics knowledge base, philosophy, and teaching strategies. They should evaluate principals' support of and participation in DMI staff development.

"Successful schools do not simply happen; they are successful because people make them so" (Blandford, 2000, p. 13). Schools can develop more effective learning environments through hard work, appropriate staff development, and thoughtful research.

REFERENCES

Allen, D. (1998). *Assessing student learning from grading to understanding.* New York: Teachers College Press.

Ardovino, J., Hollingsworth, J., & Ybarra, S. (2000). *Multiple measures: Accurate ways to assess student achievement.* Thousand Oaks, CA: Corwin Press.

Armstrong, T. (1998). *Awakening genius in the classroom.* Alexandria, VA: Association for Supervision and Curriculum Development.

Badal, A. (2002). *Elementary school analysis.* Lanham, MD: Department of Education. (ERIC Document Reproduction Service No. ED464400)

Bay, J., Reys, B., & Reys, R. (1999). The top 10 elements that must be in place to implement standards-based mathematics curricula. *Kappan, 80,* 503-506.

Bergeron, J. C., & Herscovics, N. (1990). Psychological aspects of learning early arithmetic. In P. Nesher & J. Kilpatrick (Eds.), *Mathematics and cognition: A research synthesis by the International Group for the Psychology of Mathematics Educator* (pp. 31-51*).* New York: Cambridge University Press.

Blandford, S. (2000). *Managing professional development in schools.* New York: Routledge.

Borman, G., Rachuba, L., Datnow, A., Alberg, M., Mac Iver, M., Stringfield, S., & Ross, S. (2000). *Four models of school improvement: Successes and challenges in reforming low-performing, high-poverty Title I schools.* Baltimore, MD: Center for Research on the Education of Students Placed at Risk. (ERIC Document Reproduction Service No. ED447238)

Brownell, W. (1928). *Development of children's number ideas in the primary grades.* Chicago: University of Chicago Press.

Bruchey, S. (1997). (Ed.). *Children of poverty: Studies on the effects of single parenthood, the feminization of poverty, and homelessness.* New York: Garland Publishing.

Bruner, J. S. (1966). *Toward a theory of instruction.* Cambridge, MA: Belknap Press of Harvard University Press.

Burns, M. (1989). Teaching for understanding: A focus on multiplication. In P. Trafton & A. Shulte (Eds.), *New directions for elementary school mathematics* (pp. 123-133). Reston, VA: National Council of Teachers of Mathematics.

Burns, M. (1992). *About teaching mathematic: A K-8 resource.* Sausalito, CA: Math Solutions Publications.

Bybee, R., & Sund, R. (1982). *Piaget for educators.* Columbus, OH: Charles Merrill Publishing Co.

Case, R. (1996). Changing views of knowledge and their impact on educational research and practice. In D. Olson & N. Torrance (Eds.), *The handbook of education and human development: New models of teaching, learning and schooling* (pp. 75-99). Cambridge, MA: Blackwell Publishing.

Cooper, H. (1989). *Integrating research: A guide for literature reviews* (2nd ed.). Newbury Park, CA: Sage Publications.

Council of Chief School Officers. (2002). *Expecting success: A study of five high performing high poverty elementary schools.* Washington, DC: Author. (ERIC Document Reproduction Service No. ED468010)

Cronbach, L., Ambron, S., Dornbusch, S., Hess, R., Hornik, R., Phillips, D., Walker, D., & Weiner, S. (1980). *Toward reform of program evaluation: Aims, methods, and institutional arrangements.* San Francisco, CA: Jossey-Bass.

Cuban, L. (1988). Constancy and change in schools (1880's to the present). In P. Jackson (Ed.), *Contributing to educational change: Perspectives on research and practice* (pp. 85-106). Berkeley, CA: McCutchan Publishing.

Cuban, L. (1997). School reform: The riddle of change and stability. In B. Kogan (Ed.), *Common schools, common futures: A working consensus for school renewal* (pp. 14-33). New York: Teachers College Press.

D'Agostino, J. V. (2000). Achievement testing in American schools. In T. L. Good & M. Early (Eds.), *American education: Yesterday, today, and tomorrow. 99th yearbook of the National Society for the Study of Education Part II* (pp. 313-337). Chicago: University of Chicago Press.

Davis, R. B. (1990). Discovery learning and constructivism. In R. Davis, C. Maher, & N. Noddings (Eds.), *Journal for research in mathematics education monograph number 4. Constructivist views on the teaching and learning of mathematics* (pp. 93-107). Reston, VA: National Council of Teachers of Mathematics.

Davison, D., Miller, K., & Metheny, D. (1999). *Integrating science and mathematics in elementary curriculum.* Bloomington, IN: Phi Delta Kappa Educational Foundation.

Dixon-Krauss, L. (1996). *Vygotsky in the classroom: Mediated literacy instruction and assessment*. White Plains, NY: Longman Publishing.

Drew, D. (1996). *Aptitude revisited.* Baltimore, MD: Johns Hopkins University Press.

Dunn, E. (2000). *Appropriate reading assessment: Are we on the right track?* (ERIC Document Reproduction Service No. ED448407)

Education Development Center. (2002). *Research into the use and impact of DMI.* Retrieved December 13, 2002, from http://www.edc.org/MLT/CDT/dmiresearch.html

Educational Testing Service. (1986). *The ETS test collection catalog. Vol. 1. Achievement tests and measurement devices.* Phoenix, AZ: Oryx Press.

Elkin, D. (1969). Conservation and concept formation. In D. Elkin & J. Flavell (Eds.), *Studies in cognitive development; Essays in honor of Jean Piaget* (pp. 171-190). New York: Oxford University Press.

Fuhrman, S. (Ed.). (2001). *From the capitol to the classroom: Standards-based reform in the states. 100th yearbook of the National Society for the Study of Education.* Chicago: University of Chicago Press.

Furth, H. (1970). *Piaget for teachers.* Englewood Cliffs, NJ: Prentice-Hall.

Good, R. (2002). *Using discriminate analysis as a method of combining multiple measures of student performance*. New Orleans, LA: American Educational Research Association. (ERIC Document Reproduction Service No. ED464951)

Grouws, D. A., & Cebulla, K. J. (2000). Elementary and middle school mathematics at the crossroads. In T. L. Good & M. Early (Eds.), *American education: Yesterday, today, and tomorrow. 99th yearbook of the National Society for the Study of Education Part II* (pp. 209-255). Chicago: University of Chicago Press.

Hernandez, D. (1997). Poverty trends. In G. Duncan & J. Brooks-Dunn (Eds.), *Consequences of growing up poor* (pp. 18-34). New York: Russell Sage Foundation.

Hiebert, J. (1990). The role of routine procedures in development of mathematics competence. In T. Cooney & C. Hirsch (Eds.), *Teaching and learning mathematics in the 1990's* (pp. 31-40). Reston, VA: National Council of Teachers of Mathematics.

Hilliard, A. (2000). Excellence in education versus high-stakes standardized testing. *Journal of Teacher Education, 51,* 293-304.

Hodges, H. (2001). Overcoming a pedagogy of poverty. In R. Cole (Ed.), *More strategies for educating everybody's children* (pp. 1-9). Alexandria, VA: Association for Supervision and Curriculum Development.

Hodgkinson, H. (2002). The demographics of diversity. *Principal, 82,*14-18.

Jacobs, S. H. (1984). *Foundations for Piagetian education.* Lanham, MD: University Press of America.

Joersz, D. (2003). *Free and reduced data sheet.* (Available from Bismarck Public Schools, Hughes Building, 400 Washington, Bismarck, ND 58501).

Kamii, C. (1990). Constructivism and beginning arithmetic. In T. Cooney & C. Hirsch (Eds.), *Teaching and learning mathematics in the 1990's (*pp. 22-30). Reston, VA: National Council of Teachers of Mathematics.

Kannapel, P., Aagaard, L., Coe, P., & Reeves, C. (2001). The impact of standards and accountability on teaching and learning in Kentucky. In S. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states* (pp. 242-262). Chicago: University of Chicago Press.

Kilpatrick, J. (2001). *State proficiency testing in mathematics ERIC digest.* Washington, DC: Office of Educational Research and Improvement. (ERIC Document Reproduction Service No. ED466353)

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics.* Washington, DC: National Academy Press.

Kline, M. (1973). *Why Johnny can't add: The failure of new mathematics.* New York: St. Martin's Press.

Kohn, A. (2000a). Burnt at the high stakes. *Journal of Teacher Education, 51*, 315-327.

Kohn, A. (2000b). *The case against standardized testing: Raising the scores, ruining the schools.* Portsmouth, NH: Heinemann.

Kroll, D. L. (1989). Connections between psychological learning theories and elementary mathematics curriculum. In P. Trafton & A. Shulte (Eds.), *New directions for elementary school mathematics* (pp. 199-211). Reston, VA: National Council of Teachers of Mathematics.

Land, D., & Legters, N. (2002). The extent and consequences of risk in US education. In S. Stringfield & D. Land (Eds.), *At risk students. 101[st] yearbook of the National Society for the Study of Education* (pp. 1-28). Chicago: University of Chicago Press.

Lewis, A. (2001). *Adding it up: Using research to improve education for low-income and minority students.* Washington, DC: Poverty and Race Research Action Council.

Lovitt, C., Stephens, M., Clarke, D., & Romberg, T. (1990). Mathematics teachers reconceptualizing their roles. In T. Cooney & C. Hirsch (Eds.), *Teaching and learning mathematics in the 1990's* (pp. 229-236). Reston, VA: National Council of Teachers of Mathematics.

McCollum, J. (1978). *Ah hah! The inquiry process of generating and testing knowledge.* Santa Monica, CA: Goodyear Publishing Co.

McGraw-Hill. (2001). *TerraNova, technical report.* Monterey, CA: CTB/McGraw-Hill.

McGraw-Hill. (2002). *TerraNova, second edition, technical bulletin 1.* Monterey, CA: CTB/McGraw-Hill.

Murphy, L., Impara, J., & Plake, B. (Eds.). (1999). *Tests in print V.* Lincoln: Buros Institute of Mental Measurements, University of Nebraska-Lincoln Press.

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

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Staff Development Council. (2001). *Standards for staff development.* Oxford, OH: Author.

Newman, R., & Beck, L. (2000). Standards, curriculum reform, and the educational experiences of one homeless youngster: Some reflections. In R. Mickelson (Ed.), *Children on the streets of the Americas* (pp. 136-145). New York: Routledge.

North Dakota Department of Public Instruction. (2002). *Child nutrition and food distribution program: Income eligibility guidelines 2000-2001.* Bismarck, ND: Author.

Nunes, T., & Bryant, P. (1996). *Children doing mathematics.* Cambridge, MA: Blackwell.

Olson, D. R., & Bruner, J. S. (1996). Folk psychology, folk pedagogy. In D. Olson & N. orrance (Eds.), *The handbook of education and human development: New models of learning, teaching and schooling* (pp. 9-27). Cambridge, MA: Blackwell Publishing.

Peltzman, B. R. (1998). *Pioneers of early childhood education.* Westport, CT: Greenwood Press.

Piaget J. (1978). *Success and understanding.* (A. Pomerans, Trans.). Cambridge, MA: Harvard University Press.

Pinks, W., & Hyde, A. (1992). *Effective staff development for school change.* Norwood, NJ: Ablex Publishing.

Popham, W. (2001). *The truth about testing: An educator's call to action.* Alexandria, VA: Association for Supervision and Curriculum Development.

Reed-Victor, E., & Stronge, J. (2001). Diverse teaching strategies for homeless children. In R. Cole (Ed.), *More strategies for educating everybody's children* (pp. 10-32). Alexandria, VA: Association for Supervision and Curriculum Development.

Reys, R. (2001). Curricular controversy in the mathematics wars: A battle without winners. *Kappan, 83*, 255-258.

Rowan, T., & Bourne, B. (1994). *Thinking like mathematicians.* Portsmouth, NH: Heinemann.

Salend, S. (2001). *Creating inclusive classrooms: Effective and reflective practices* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Salvia, J., & Ysseldyke, J. (2001). *Assessment* (8th ed.). Boston: Houghton Mifflin.

Sawyer, A. (1995). *Developments in elementary mathematics teaching*. Portsmouth, NH: Heinemann.

Schifter, D. (1996). *What's happening in math class: Reconstructing professional identities*. New York: Teachers College Press.

Schifter, D., Bastable, V., & Russell, S. (1999a). *Number and operations, Part 1: Building a system of tens*. Parsippany, NJ: Dale Seymour Publications.

Schifter, D., Bastable, V., & Russell, S. (1999b). *Number and operations, Part 2: Making meaning for operations*. Parsippany, NJ: Dale Seymour Publications.

Silver, E. (1990). Contributions of research to practice: Applying findings, methods, and perspective. In T. Cooney & C. Hirsch (Eds.), *Teaching and learning mathematics in the 1990's* (pp. 1-11). Reston, VA: National Council of Teachers of Mathematics.

Silver, E., Kilpatrick, J., & Schlesinger, B. (1995). *Thinking through mathematics: Fostering inquiry and communication in mathematics classrooms*. New York: College Entrance Examination Board.

Sirotnik, K. (1994). Equal access to quality in public schooling: Issues in the assessment of equity and excellence. In J. Goodlad & P. Keating (Eds.), *Access to knowledge: The continuing agenda for our nation's schools* (pp. 159-185). New York: College Entrance Examination Board.

Smith, J., Brooks-Gunn, J., & Klebanov, P. (1997). Consequences of living in poverty for young children's cognitive and verbal ability and early school achievement. In G. Duncan & J. Brooks-Dunn (Eds.), *Consequences of growing up poor* (pp. 132-190). New York: Russell Sage Foundation.

Sparks, D., & Hirsh, S. (1997). *A new vision for staff development.* Alexandria, VA: Association for Supervision and Curriculum Development.

Speck, M. (1996, Spring). Best practice in professional development for sustained educational change. *ERS Spectrum,* 33-41.

Steffe, L. P. (1990). Adaptive mathematics teaching. In T. Cooney & C. Hirsch (Eds.), *Teaching and learning mathematics in the 1990's* (pp. 41-51). Reston, VA: National Council of Teachers of Mathematics.

Stigler, J., & Hiebert, J. (1997). Understanding and improving classroom mathematics instruction: An overview of the TIMSS video study. *Kappan*, *79*, 14-21.

Stronge, J. (2000). The education of homeless children and youth in the United States: A progress report. In R. Mickelson (Ed.), *Children on the streets of the Americas: Globalization, homelessness and education in the United States, Brazil and Cuba* (pp. 66-78). New York: Routledge.

Sund, R. (1976). *Piaget for educators*. Columbus, OH: Charles Merrill Publishing Co.

Taber, S. (1998, April). *Learning to teach math differently.* Paper presented at the meeting of the American Educational Research Association, San Diego, CA.

Teachman, J., Paasch, K., Day, R., & Carver, K. (1997). Poverty in adolescence and subsequent educational attainment. In G. Duncan & J. Brooks-Dunn (Eds.), *Consequences of growing up poor* (pp. 382-418). New York: Russell Sage Foundation.

Thorndike, E. L. (1932). *The fundamentals of learning*. New York: Teachers College Press, Columbia University.

Thorndike, E. L. (1970). The psychology of mathematics. In J. Bidwell & R. Clason, (Eds.), *Readings in the history of mathematics education* (pp. 461-477). Washington, DC: National Council of Teachers of Mathematics.

Trafton, P. R., Reys, B. J., & Wasman, D. G. (2001). Standards-based mathematics curriculum materials: A phrase in search of a definition. *Kappan, 83*, 259-264.

Underhill, R. (1972). *Teaching elementary school mathematics*. Columbus, OH: Charles E. Merrill Publishing Co.

U.S. Bureau of Census. (2002). *Poverty in the United States: 2001*. Washington, DC: Author.

U.S. Bureau of Census. (2003). *Historical poverty levels*. Retrieved June 2, 2003, from http://www.census.gov/hhes/poverty/histpov/hstpov3.html

U.S. Department of Education. (1998). *Promising practices: New ways to improve teacher quality* (Report No. SP 038 153). Washington, DC: Office of the Secretary. (ERIC Document Reproduction Service No. ED449117)

U.S. Department of Health and Human Services. (1996). *Indicators of welfare dependence and well-being*. Washington, DC: Author. (ERIC Document Reproduction Service No. ED461676)

Vygotsky, L. V. (1962). *Thought and language*. (E. Hunfmann & G. Vakar, Eds. & Trans.). Cambridge, MA: MIT Press.

Wagner, T. (2001). Leadership for learning: An action theory of school change. *Kappan*, *82*, 378-383.

Wakefield, A. (1997). Supporting mathematics thinking. *Kappan*, *79*, 233-236.

Yackel, E., Cobb, P., Wood, T., Wheatley, G., & Merkel, G. (1990). The importance of social interaction in children's construction of mathematics knowledge. In T. Cooney & C. Hirsch (Eds.), *Teaching and learning mathematics in the 1990's* (pp. 12-21). Reston, VA: National Council of Teachers of Mathematics.