DISSERTATION

A SYSTEMATIC REVIEW OF INTERVENTIONS IN SECONDARY MATHEMATICS WITH AT-RISK STUDENTS: MAPPING THE LITERATURE

Submitted by

James Joseph Dugan

School of Education

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2007

UMI Number: 3266389

COLORADO STATE UNIVERSITY

March 28, 2007

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY JAMES JOSEPH DUGAN ENTITLED "A SYSTEMATIC REVIEW OF INTERVENTIONS IN SECONDARY MATHEMATICS WITH AT-RISK STUDENTS: MAPPING THE LITERATURE" BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Paul A. Kennedy

Paul A Kennedy, Ph. D.

Chance W. Lewis

Chance W. Lewis, Ph. D.

Gene W. Gloeckner

Advisor: Gene Gloeckner, Ph. D.

R. Brian Cobb

Co-Advisor: R. Brian Cobb, Ph. D.

Timothy S. Davies

Interim Director: Tim Davies, Ph. D.

ABSTRACT OF DISSERTATION

A SYSTEMATIC REVIEW OF INTERVENTIONS IN SECONDARY MATHEMATICS WITH AT-RISK STUDENTS: MAPPING THE LITERATURE

A systematic review of 3,814 published and unpublished citations identified in a literature search on interventions in secondary and early post-secondary mathematics with at-risk students between 1990 and 2004 was conducted. The goal of the systematic review was to identify those interventions which are most successful at improving the mathematics achievement of at-risk students while reducing the mathematics achievement gap.

The utilization of a multi-phased coding process resulted in 100 empirical studies meeting the inclusionary criteria while yielding effect size estimates for a number of interventions and for multiple at-risk factors. This report summarizes these studies through a two-level mapping of the literature base. It identifies the interventions, at-risk factors, sample characteristics, study settings, and methodological research designs utilized in this body of empirical studies.

The results of this approach suggest that those interventions identified as effective in increasing mathematics achievement with secondary and early post-secondary, at-risk students can be grouped into two primary intervention categories (a) pedagogical or instructional interventions, including cognitive and metacognitive approaches, cooperative learning, and peer tutoring, and (b) technology-based interventions. Two other types of intervention provided inconclusive results. No conclusions could be reached regarding curriculum interventions due to the small number of studies meeting the inclusionary criteria and the wide variety of curricular approaches included in those studies.

The final results of the systematic review will be a series of research articles describing effective interventions for student populations with specific at-risk factors. In addition, the results of this systematic review can be used to direct future research efforts by defining investigations into specific interventions with specific populations while informing the knowledge base.

James Joseph Dugan
School of Education
Colorado State University
Fort Collins, CO 80523
Spring 2007

ACKNOWLEDGEMENTS

With the completion of both this dissertation and of my doctoral program, I have reached the end of another chapter in my life: one that I have thoroughly enjoyed. As with most experiences in our lives, the people we meet and share our experiences with play a large role in defining the memories we have of those experiences. I can say that I have been able to share this experience, and this chapter, with some of the most interesting people I have ever met in my life. Any acknowledgement of individuals I could make would surely exclude a sizable number of fine people I have grown to appreciate. However, I do want to acknowledge some of the individuals whose effort and support (and patience with me) are greatly appreciated.

First and foremost on my list of acknowledgements are the fine professors I have had the chance to learn from. These include those four incredible individuals who assisted me as my dissertation committee: Dr. Gene Gloeckner, Dr. Brian Cobb, Dr. Paul Kennedy, and Dr. Chance Lewis. I was quite proud to tell other graduate students that I had the best dissertation committee imaginable. I could never thank Drs. Cobb and Kennedy enough for the opportunity to be involved with the CLT-West and to participate in the research effort reported in this dissertation. I would also like to acknowledge the assistance and my appreciation of the other doctorial fellows who helped in this research: Shelly Parsons, Laurie Hillman, Dora Gonzales, and Jerry Overmyer.

Those professors from the School of Education who added so much to my education though their insights and their challenging classes were Dr. Len Albright, Dr. Dean Brown, Dr. Ann Foster, Dr. Gene Gloeckner, Dr. Jean Lehmann, and Dr. Barbara Nelson. Of special note is Dr. Carole Makela, one of the most incredible and knowledgeable people I have had the opportunity to know.

The R&D Center, being a joint effort with the Colorado State University and the Poudre School District, was my "home away from home" for much of the last four years. Dr. Ellyn Dickmann, Dr. Cobb, Dr. Tom Cavanagh, and Dr. Michael DeMiranda provided me with opportunities to learn that greatly enhanced my knowledge and my experience level. That experience was invaluable to me. In addition, the interaction with other graduate students and R&D Center staff was invaluable.

A final group of individuals who have been instrumental in adding to the value of my doctoral program has been my fellow graduate students. This group includes Rose Quinn, Jenni Wolgemuth, Tim Pearson, Kellie Enns, Dr. Dorothy Garrison-Wade, Anthony Wade, Bruce Douglas, Alina Waite, and many others. These individuals have made the effort worthwhile.

Thank you to all who have supported me.

DEDICATION

The hardest part of writing this document has been this short section. As John Lennon of the Beatles wrote in a song entitled *Beautiful Boy*, "Life is what happens to you while you're busy making other plans." That is exactly what happened during my doctoral program.

I dedicate this work to my family, including my wife Colleen; my mother Elva; and my three sons, Joe, Tom and Jeff. Thank you, Colleen, for putting up with me during this ordeal. This was something I always wanted to do and your support and patience made it happen. Thank you, Mom, for all your support and love. Joe, Tom, and Jeff: you will never know how proud I am on the three of you. I take more pride in the three of you than I do in anything else I've done or been associated with in my life.

Finally, thank you God for two things. First, thank you for assisting me in completing this experience and surviving the process, literally. Secondly, thank you for sparing Jeff.

I had told people that when Jeff decided to enter the U.S. Marines out of high school, it allowed me to pursue my dream of getting a Ph.D. Little did any of us know how close we would come to losing Jeff in Afghanistan. Jeff, you have always been my hero. And, now as we know, you are an American hero.

TABLE OF CONTENTS

LIST OF TABLES

TABLE OF FIGURES

CHAPTER 1

INTRODUCTION

The need to increase the level of mathematics achievement demonstrated by U.S. students has been the focus of numerous school reform movements over the last 100 years. As the need to apply mathematics in the workplace has increased, so has the need for higher levels of competence in mathematics among the labor force (National Science Board, 2004). As our society's economic base has changed from agriculture to industry to information, the call for additional research on how to improve mathematic achievement has increased. This push has manifested itself in both an increase in educational accountability and a call for more empirical evidence of what works to improve mathematics achievement with school-age children.

The intent of this study was to perform a systematic review of interventions aimed at improving secondary and early post-secondary student mathematics achievement. Stated differently, this study is a systematic review of "what works" in improving mathematics achievement. Of primary interest is the at-risk student population: defined as students at risk of educational failure because of limited English proficiency, poverty, race, geographic location, or economic disadvantage (Office of Educational Research and Improvement (OERI), 2002). This population has been often overlooked but is now, with the reporting of Annual Yearly Progress (AYP) mandated by the No Child Left Behind Act of 2001, an increasingly visible population. It is the goal of this study to identify those interventions which are most successful at improving the mathematics achievement of secondary and early post-secondary, at-risk students and most effective in reducing the mathematics achievement gap.

Study Context

In the early 1990s, the National Education Goals Panel (NEGP) set eight national goals for education in the United States, including academic achievement goals in mathematics (National Education Goals Panel, 2002). As part of their charge, the NEGP called for U.S. students to be first in the world in mathematics and science achievement. In one of the last reports published by NEGP, prior to its dissolution, Barton (2002) reported on the progress made in mathematics by 4^{th} graders from 1992 to 2000 and 8^{th} graders from 1990 to 2000 using National Assessment of Educational Progress (NAEP) data. Overall, the mathematics results showed statistically significant improvements for the majority of the 36 states that participated in the NAEP assessment program. However, Barton indicated that this improvement did not include a reduction in the achievement gaps between selected groups of students, including minority students, students eligible for free or reduced lunch, and other at-risk students. This failure to reduce the achievement gap for at-risk students has been noted also by the National Science Board (2004).

President George W. Bush noted recently (Domestic Policy Council, 2006) that "The bedrock of America's competitiveness is a well-educated and skilled workforce." Nations today require skilled, knowledgeable workforces and citizens with a sound understanding of mathematics and science (National Science Board, 2004). To compete in the international or "world" market, the U.S. economy is dependent on individuals with higher levels of mathematics, science, engineering, and technical skills. It is also noted that, too maintain our competitive edge in the world economy, the U.S. must focus on scientific and technological development and real world applications, which are fueled by scientific research, a strong education system, and an innovative environment (Domestic Policy Council, 2006).

Additionally, a greater understanding of mathematics is necessary for today's student. To gain success in the job market tomorrow, students will need more than math computation competence. They will also need to apply mathematical knowledge to solve problems. The responsibility for the development of a more highly-qualified work force is entrusted to the U.S. public education system (National Science Board, 2004).

The level of mathematics achievement demonstrated by students in the United States continues to fall below the expectations set by NEGP, education policy analysts and the general public. Since the publication of "*A Nation At Risk*" in 1983 (National Commission on Excellence in Education, 1983), the desire to increase children's science and mathematics achievement levels has driven many educational reform efforts in our public schools (Tyack & Cuban, 1995). The recent release of the mathematics assessment data from the Program for International Student Assessment (or PISA 2003), shows students from the

United States ranked 24th of 29 participating nations on mathematics problem-solving and mathematics literacy assessment (Organization for Economic Co-operation and Development, 2004). The release of this study was followed by a somewhat more positive report from the Trends in International Mathematics and Science Study (TIMSS). The TIMSS study showed that 4th and 8th grade students in the U.S. scored above the international average on mathematics and science (International Association for the Evaluation of Educational Achievement, 2003).

During the early 1990s, progress was made in improving student performance in mathematics and science (National Education Goals Panel, 2002; National Science Board, 2004). The achievement gap between males and females in mathematics largely disappeared during this time period (National Science Board, 2004). However, overall U.S. student performance in mathematics and science remained at or below the average level of students in other developed countries (National Science Board, 2004). Additionally, it was noted that U.S. student performance becomes weaker at higher grade levels as seen on the TIMSS (National Science Board, 2004).

With the No Child Left Behind (NCLB) Act of 2001, the latest reauthorization of the Elementary and Secondary Education Act (ESEA) of 1965, the federal government has increased efforts to improve student achievement in all academic areas including mathematics. NCLB requires that states not only show significant improvements in overall student achievement, but are also required to demonstrate achievement gains for all student subgroups (National Governors' Association, 2004). Educational policymakers are concerned that not all students, especially those students considered to be at-risk of failure in the U.S. public school system, will benefit from these new requirements. The impact of at-risk students has been seen nationally with the implementation of Adequate Yearly Progress (AYP) measures, a part of the NCLB legislation. Many school districts nationally have failed to meet NCLB guidelines due to missing the AYP requirements for one or more subgroups of at-risk students. The need to ensure that all subgroups of students benefit from the NCLB legislation has become a measuring stick for all public K-12 educational organizations in the country (National Governors' Association, 2004).

Statement of Problem

The Center for Teaching and Learning in the West (CLT-West), a collaboration of five higher educational institutions in the western U.S., including Colorado State University, was established by the National Science Foundation (NSF) to foster science and mathematics teaching development and research. In developing the effort reported here, members of the CLT-W organization at Colorado State University chose to conduct a systematic review of mathematics interventions with at-risk students. The problem statement for this study, as originally proposed by Cobb and Kennedy (personal communications, August 12, 2004) to Elizabeth Swanson, Director of CLT-West, was as follows: "What are the characteristics of curricular and/or instructional interventions that are most effective in reducing (or reversing) the mathematics achievement gap for secondary students who are at greatest risk of school failure?"

Research Questions

Two key research questions have been developed to focus this systematic review. These questions are:

1. What are the characteristics of curricular interventions that are most effective in reducing the mathematics achievement gap for at-risk secondary students?
2. What are the characteristics of instructional interventions that are most effective in reducing the mathematics achievement gap for at-risk secondary students?

Definition of Terms

The definitions of terms to be used in this report are as follows. A systematic review is considered a quantitative summary of primary research studies which uses an explicit search strategy to locate potential studies, involves a rigorous inclusionary appraisal of those studies, and is evidence-based (Cook, Mulrow, & Haynes, 1997). A systematic review which includes a statistical synthesis of data is called a meta-analysis.

At-risk students are defined as students for whom traditional approaches to organizing and delivering instruction have worked least well – those students most "at risk" of not completing or benefiting from their elementary and secondary school experiences (Hixson & Tinzmann, 1990). These students are primarily minorities, the poor, or immigrants who are considered culturally or socially disadvantaged. The U.S. Department of Education defines "at-risk" students as students at risk of educational failure because of limited English proficiency, poverty, race, geographic location, or economic disadvantage (Office of Educational Research and Improvement (OERI), 2002). For our purposes, we will also consider gender to be an at-risk factor as it has been noted that females may be considered a disadvantaged group in science or mathematics achievement (Snyder, Tan, & Hoffman, 2004). For clarification, those studies in which the at-risk group was defined as "high minority" were studies in which the author identified the percentage of minority students as being greater than 50% of the total students.

Curricular interventions are interventions associated with the structure of the content materials in an educational context. Instructional interventions are interventions that are related to the method of content delivery in an educational context.

Review Delimitations

The delimitations of a systematic review set the scope of the review under consideration. Typically, the delimitations are described as the inclusionary criteria for the review and are provided in the review's protocol. The coding documents used in a systematic review process define the delimitation of the study characteristics. Copies of the three coding documents used in this review, the initial abstract coding document, the keyword coding document, and the data extraction coding document, are found in Appendixes A, B, and C, respectively.

Delimitations of this review include the years of publication for original studies included in the review, sources of studies, publication status, types of research designs, types of interventions, outcomes, educational setting, grades of students, mathematics topics studied, and characteristics of the at-risk students in the study. Table 1 provides a description of delimitation categories with included or excluded category elements.

Changes were made to the inclusionary criteria between the initial review's inception and the completion of the review. These changes were made to limit the scope of the review to empirical studies with cognitive outcomes and included the removal of selected research designs, such as qualitative and single-subject designs; selected interventions, such as professional development for teachers; and selected outcomes, such as affective, process, and enrollment in higher mathematics courses.

Assumptions

Many assumptions are made in the process of conducting a systematic review. These assumptions are associated with the systematic review process itself, the types of studies included in such a review, attempts made to avoid bias, and in the calculation and comparison of effect size estimates.

It should be noted that a systematic review is a review of previously conducted research studies. As such, a systematic review is considered secondary research. No primary research is conducted during a systematic review. The systematic review is then limited by the quality of the studies

Table 1

Delimitations of Review (final version) by Group

Delimitation Category	Category Elements (Inclusionary Criteria)	
Time period of studies included	Dated from January, 1990 through December 2004	
Sources of studies	All possible sources	
Status of study article	Published, in press, or unpublished	
Grade of learners	Grades 6th through 14th(secondary & early post-secondary)	
Characteristic	Included	Excluded
Types of research designs	Multi-group quantitative Single group pre/post test Multiple regression Regression discontinuity Interrupted time series	One-group correlational Qualitative Single subject All others
Types of interventions	Curriculum Pedagogical/Instructional Grouping structure/Tracking Technology School restructuring	Personal/Affective Professional development All others
Outcomes	Cognitive	Affective Process Enrollment in higher math
Educational settings	Middle/junior high school High school Community/junior college Four-year college Home schooled Residential school Correctional setting	All others
Mathematics content area	Numbers, fractions, decimals Algebra Geometry/Measurement Probability/Statistics Pre-Calculus Calculus	All others
At-risk factors	English language learners Gender High minority Low-achievers Low socioeconomic status Special needs	None

included in the review. The primary intent of the process of a systematic review is to identify for inclusion those studies of the highest quality.

Conceptual Framework

The conceptual framework used to guide the philosophical orientation to this systematic review is grounded in the ecological model of social functioning. This framework was used to answer "what works" questions regarding interventions for increasing mathematics achievement with at-risk students. The framework focuses on the transactional relationship among persons, environments and behaviors and was first introduced by Karl Lewin (1936). Additionally, the framework provided the conceptual structure to guide the scope, the methodology, and the development of this research synthesis. The question of "what works" can be translated by the classical ecological question posed by Wachs (1987): "Under what environments (situations, programs, and settings) have what kinds of persons (secondary or early-postsecondary at-risk students) changed in what kinds of behaviors (mathematics achievement)?"

We are currently in an era where accountability has taken center stage as the primary driving force behind much of the federal government's efforts and funding related to education. With the implementation of the No Child Left Behind Act of 2001, the federal government has shifted its education policy direction from being a source of compensatory funding for low-income students to one where the federal government is defining the outcomes of education (Fusarelli, 2004). The reshaping of the Office of Educational Research and

Improvement (OERI) to become the Institute of Education Sciences (IES) within the U.S. Department of Education, the shift in focus by IES to fund primarily scientifically-based research, and the formation and funding of operations such as the What Works Clearinghouse are indicators of this increased emphasis on accountability.

All of this activity reflects the current shift of theoretical and philosophical thinking in educational policy and research whereas the primary philosophy is now considered to be in line with a postpositivist approach. This postpositive approach is one that includes the beliefs that educational research can be based on conclusions which include standards of truth, are subject to hypotheses testing, and are able to withstand criticism (Phillips & Burbules, 2000).

Context for Learning

In a monumental body of work synthesized to define the "science of learning", the National Academies of Sciences has recently published two major works summarizing the existing knowledge base on how people learn (National Research Council, 2000, 2005). In the book *How People Learn: Brain, Mind, Experience, and School* (National Research Council, 2000), it is acknowledged that three fundamental and well-established principles of learning are of importance for teachers to understand and incorporate into their approach to teaching. These key principles are (a) students bring into the classroom previous life experiences and preconceptions about the world (world view) which must be engaged for new learning to occur, (b) to become competent in a study area, students must possess a strong base of factual knowledge within a conceptual framework that allows the student to retrieve and apply that information in novel situations, and (c) students taught using a "metacognitive" approach to instruction can learn to monitor and evaluate their own learning and learning goals. Based on these principles, our approach to this systematic review, and ultimately our understanding of how to improve the teaching of mathematics to all students rests primarily on the critical nature of the teacher's capabilities in the classroom.

Study Significance

Comparisons made between U.S. secondary students and comparable students from other industrialized countries in mathematics achievement have not been favorable for the United States over the last couple of decades. Determining what strategies are effective in increasing mathematics achievement would be of significant interest to U.S. educators. This decision process addresses the specific question of "what works" in increasing mathematics achievement. In addition, determining effective strategies for increasing the achievement levels of at-risk student populations could significantly decrease the achievement gap while improving the overall ranking of U.S. students in international comparisons.

What curricular or instructional techniques can the high school teacher with 30 low-achieving students or 15 students with learning disabilities utilize to increase the students' understanding of mathematics? What has worked for other teachers in similar situations? There are no clear answers to these questions for the mathematics educator currently. With the utilization of a systematic review and meta-analysis methodology, we may find a means of identifying some of the most critical characteristics of effective interventions for increasing mathematics achievement with at-risk students.

This review adds to the existing knowledge base by specifically addressing the question of what curricular and instructional interventions are effective for increasing mathematics achievement within at-risk, secondary and early-postsecondary student populations. Though numerous research studies aimed at describing the effectiveness of selected curricular and instructional interventions in secondary mathematics have been published, few have addressed specific at-risk student subgroups.

The approach used in this review is unique in two ways. First, the results of the review are presented through both a descriptive mapping of entire body of citations identified in the literature selection phase of the review and in a systematic mapping of the characteristics of those studies meeting the inclusionary criteria for the review. Secondly, the qualified studies are summarized by specific interventions and include estimates of the main effect size for the intervention and for the specific at-risk factor.

Mapping of the Literature Base

Two literature mapping approaches will be used to summarize the characteristics of the literature base reviewed. The first mapping approach is referred to as a descriptive mapping (EPPI-Centre, 2006) and is described as the classification by key characteristics of the range of studies encountered in the literature search as a whole. This mapping approach can be used to describe the scope and coverage of the research activity in the topic area, to show relevance of research efforts in the topic area, and to provide context for interpreting the results of the systematic review synthesis. The second mapping approach is termed a systematic mapping (Bennett, Lubben, Hogarth, Campbell, & Robinson, 2005) and generally refers to the mapping of the specific characteristics of the studies meeting the inclusion criteria in a systematic review synthesis. This mapping approach is typically more detailed and includes the primary characteristics used in the studies, such as specific sample characteristics, intervention types, and outcomes.

CHAPTER 2

REVIEW OF LITERATURE

The intent of this chapter is to present a review of the relevant literature related to interventions in mathematics with at-risk secondary students. It includes a discussion on the role of mathematics in today's society, including implications related to the mathematics achievement gap. A review of previously completed systematic reviews addressing similar topics is presented, along with an in-depth review of five key systematic reviews. This is followed by a discussion on the origins and implications of both curricular and instructional interventions. Finally, an overview of the current literature on the mathematics achievement gap is presented.

The Importance of Mathematics in Today's Society

Historically, mathematics has proven to be a critical measure of academic achievement. Today, many education researchers and policy analysts use measures of mathematics and reading achievement as proxies for overall academic achievement. More than 25 years ago, Sells (1978) called mathematics the *critical filter*. Campbell (1991) has stated that mathematics has served as the critical filter in education within our society with the payoff for outstanding mathematics students being both significant scholarships opportunities when entering postsecondary institutions and the prospect of higher paying occupations after college.

Mathematics has become the *gate keeper* for qualification to enter most higher-level academic courses in secondary public schools (Singham, 2003). A student's course work in mathematics has been the gateway to technological usage and to higher education (Schoenfeld, 2002). Moses has gone so far as to argue that mathematics literacy is the key to racial equality, stating that mathematics education is a civil rights issue (Moses, 1994; Moses & Cobb Jr., 2001) As the study of the hard sciences, including the physical sciences of chemistry and physics, has become more quantitative over the years, mathematics prerequisites have become more stringent. Parallel to this, mathematics and science achievement gaps are now the most pronounced among secondary school students (College Board, 1999).

Whereas mathematics may be considered the gate keeper for most students, its role as a gate keeper for higher level education opportunities is especially true for low-achieving students, at-risk students and students with special needs. Historically, Badian (1983) presented evidence that arithmetic-related learning disabilities are prevalent in 6% of the general population. Students with learning disabilities often exhibit specific problems with mathematics, with at least one-fourth of the students identified as having learning disabilities actually receiving special education services specifically for mathematics (Brian, Bay, Lopez-Reyna, & Donahue, 1991). Independently, teachers have reported that 26% of their students with learning disabilities receive special education services for mathematics difficulties encountered in the general curriculum (McLeod & Armstrong, 1982). Miller, Butler, & Lee (1998) noted that a number of researchers have reported that students with disabilities exhibit extensive problems with mathematics computation and problem-solving.

So why are mathematics skills so important to our way of life? The answer may lie in the fact that we are immersed in a social environment that is more technical in nature today than ever before. The sheer magnitude of the technical advances we will see in our lives may be an indication of how technical our society has become. The acquisition of adequate mathematics skills may be directly related to this increasingly technical society in which we live.

The Mathematics Achievement Gap

One of the areas of interest for policymakers and researchers in education is the existence of differences among student subpopulations in achievement: commonly referred to as an *achievement gap*. Dating back to the late 1960s when the first National Assessment of Educational Progress (NAEP) was administered by the U.S. Department of Education, the existence of these differences among demographic subpopulations was noted (Wenglinsky, 2004). Today, with the No Child Left Behind Act of 2001 and the Adequate Yearly Progress measure included in the Act, the visibility of achievement gaps among these subpopulations has grown.

Adequate Yearly Progress (AYP) is a measure of the improvement from year to year of students' scores on standardized tests. The reporting of AYP data is required annually by each state to the U.S. Department of Education for all student subgroups. This disaggregating of student scores by demographic subpopulation has drawn attention to the progress, or more often the lack of progress, being made by students of poverty, by minority student populations, and by students with special needs. As mandated in the No Child Left Behind Act, by the year 2013, all subpopulations of students are expected to show proficiency, with the anticipation that these achievement gaps should have disappeared (Education Commission of the States, 2003).

In measuring the existence of achievement gaps between student subpopulations, the majority of the literature focuses on the mathematics and reading achievement gaps as indicators of academic achievement, from grades 4 through 12. This focus is also true of assessment studies on an international basis.

The majority of the states are now addressing the achievement gap by adopting different strategies aimed at reducing the gap. These strategies vary from the extensive, eleven-step approach taken by the state of North Carolina to the inventive focus on professional development being employed by the state of Kentucky (Education Commission of the States, 2003; Wenglinsky, 2004).

As the states individually begin to address the reduction of the achievement gap, questions related to the effectiveness of different strategies addressing different aspects of the education process are raised. A review of the available literature on strategies to close the achievement gap shows very little in the way of solid, empirical evidence of what works. Consequently, the need for a systematic review addressing the closing of the mathematics achievement gap is indicated.

Prior Systematic Reviews

As part of the initial research plan, the author located and reviewed other previously published systematic reviews of mathematics achievement with at-risk students. Twenty-five systematic reviews were identified and the important factors were extracted, including purpose, sample, intervention and other factors using a matrix. This allowed for the summarization of study characteristics, including interventions and outcomes considered important from these prior studies. This approach also provided a means of identifying what factors were not covered adequately in prior reviews. The complete item summary of the 25 systematic reviews is presented in Appendix D with the respective studies.

The majority of the interventions examined were instructional interventions, accounting for 19 of the 25 studies. Only one review on curricular interventions was included in the group, this one being the What Works Clearinghouse mathematics curriculum review (2004a). Five reviews studied other interventions, including class size (school restructuring), mathematics anxiety and attitude towards mathematics, and gender differences in attitude towards mathematics.

Of the studies reviewing instructional interventions, 13 identified the participants as students with disabilities, primarily learning disabilities (LD). Six of the studies were of students without disabilities, of which two of these specifically included low-achieving students. A low-achieving student was defined by Baker, Gersten and Lee (2002, p. 55) as "low achieving in mathematics on the basis of their performance on standardized or informal tests or by their placement in remedial mathematics classes."

Of interest was that, as shown in Table 2 below, the majority of these reviews presented findings which indicated, in general, that the results of the reviews were mostly positive: the effect sizes presented showed on the whole that the interventions were positive and meaningful. The only reviews that did not produce overall significant, positive effects sizes were the What Works Clearinghouse review (2004a) and the Etsey and Snetler (1998) review of gender differences related to mathematics attitude.

Table 2

Summary of 25 Systematic Reviews

Intervention	Sample	# of SR	Results
Instructional	K-12 + postsec, low achievers	6	Mostly positive
Instructional	K-12 + postsec, LD	13	Mostly positive
Curricular	All students	1	Limited
Other	All students	5	Mostly positive
Total	--	25	Mostly positive

Comparable Systematic Reviews

In addition to summarizing the 25 systematic reviews, a more detailed examination of five of these systematic reviews was undertaken to provide an overview of the state of the literature regarding the effectiveness of interventions in mathematics achievement. The five studies presented were selected for different but specific reasons. Table 3 provides a summary of these five studies which will be examined in more detail.

In the first review, published in 1991, historical precedence was set with one of the pioneering reviews in mathematics (Mastropieri, Scruggs, & Shiah, 1991). A number of the other systematic reviews identified earlier cite this review as being the starting point of reviews in this field.

Mastropieri et al. (1991) noted that mathematics was an area of difficulty for most students with learning disabilities (LD). LD students spend as much as one-third of their time in the resource room working on mathematics (Carpenter, 1985). Based on this information, Mastropieri et al. choose to do a review of the research to determine what was known about mathematics instruction effectiveness. Findings from the review, which included a literature search resulting in 30 studies, included the statement that a review of all the included studies showed positive student gains.

Mastropieri et al. (1991) subdivided the studies reviewed into three types of instructional interventions and commented on each. The three types were: behavioral interventions, cognitive interventions, and alternative delivery interventions. Each of these subdivisions consisted of studies describing different instructional strategies. For instance, the behavioral interventions included reinforcement and use of manipulatives and pictures, cognitive interventions included word problem solving strategies, and the alternative delivery interventions included computer-aided instruction. In their final analysis, Mastropieri et al. reported that virtually all the interventions reviewed did produce positive effects on learning of mathematics. However, as no meta-analysis was completed, no indication was made as to the most effective interventions included in the review.

Two of the reviews of interest, Baker el al. (2002) and Barley et al. (2002), reviewed studies addressing the low-achieving student: a critical group in the "at-risk" student population. As discussed previously, low-achieving students are generally identified by low performance on standardized or informal tests, or by being placed in remedial mathematics classes. Baker et al. realized that there were no other systematic reviews at the time of their work that had investigated effective interventions in mathematics achievement with low-achieving, at-risk students. Barley et al., representing the Mid-continent Research for Education and Learning (McREL) organization, approached their synthesis review based on the need to support teachers who asked what instructional practices would work with their low-performing or at-risk students.

Table 3

Summary of Five Systematic Reviews Examined in More Detail

#	Author (Publication Year)	Study Population	Study Purpose	Specific Intervention(s)	Major Findings
1	Baker, Gersten, & Lee (2002)	Low-achieving students, K-12	To analyze findings from experimental research conducted in schools to improve the mathematics achievement of students struggling to learn math.	Instructional Intervention: Provide data teacher Peer tutoring Feedback to parents Explicit instruction	Effective interventions included: (a) Provide data teacher (ES = 0.57 wt), Peer tutoring (ES = 0.62 wt), Feedback to parents (ES = 0.43 wt), (d) Explicit instruction (ES = 0.65 wt)
2	Barley, Lauer, Arens, Apthorp, Englert, Snow, & Akiba (2002)	K-12 students with low-achievement	Synthesis on the effectiveness of strategies designed to assist low-achieving students during the school day.	(a) General Instruction (approach), (b) Cognitively-oriented Instruction (approach), (c) Grouping structures (intervention), (d)Tutoring (intervention), (e) Peer Tutoring (intervention), (f) Computer-Assisted Instruction (intervention),	(a) General Instruction (approach), Inconclusive, (b) Cognitively-oriented Instruction (approach), positive effects, (c) Grouping structures (Intervention), positive effects, (d)Tutoring (intervention), positive effects, (e) Peer Tutoring (Intervention), positive effects, (f) Computer-Assisted Instruction (Intervention), Conclusive positive effects.
3	Forness, Kavale, Blum, & Lloyd (1997)	Special education in general	Reports on some of the broad studies (meta-analyses) of research in special education that seem to indicate effective and noneffective teaching strategies.	All	Those teaching strategies that work are: Mnemonic strategies (ES = 1.63), Enhanced reading comprehension (ES = 1.13), Behavior modification (ES = 0.93), Direct instruction (ES = 0.84), Cognitive behavior modification (ES = 0.71), Formative evaluation (ES = 0.70), Early intervention (ES = 0.68)
4	Mastropieri, Scruggs, & Shiah (1991)	K-12 w/LD	Studies that evaluated the effectiveness of mathematics teaching strategies to learning disabled students were reviewed.	Instruction interventions were subdivided into behavioral interventions, cognitive interventions, and alternative delivery systems.	Virtually all instructional procedures were said to be effective by the authors.
5	Swanson, Hoskyn, & Lee (1999)	Students with LD	Was conducted to provide a comprehensive synthesis of instructional intervention studies that address students with learning disabilities.	Instruction interventions: Academic, Cognitive, Behavioral domains	Overall, studies that include components related to strategy instruction or direct instruction yield larger effect sizes that studies that do not use these instruction modes.

Baker et al. (2002) reviewed 17 studies involving low-achievement students, deliberately excluding any study in which students with disabilities were indicated. Similar to the approach taken by Mastropieri et al. (1991), Baker et al. grouped the studies in their review by intervention approach. This resulted in five intervention categories being considered: (a) providing of data and feedback on student performance to teachers and students; (b) peer tutoring; (c) involvement of parents to support instruction; (d) use of explicit or teacher-directed instruction; and (e) computer-assisted instruction. With the utilization of a meta-analysis approach, Baker et al. were able to determine a measure of effectiveness for each of the categories through the usage of both an average, unweighted effect size and a weighted effect size for each group of interventions.

For the category of studies defined as those providing data and feedback on student performance to teachers and students, Baker et al. (2002) found the weighted effect size to be 0.57 with the confidence interval indicating that the effect was significantly different from zero. Peer tutoring, a popular instructional intervention in mathematics, showed a weighted effect size of 0.66 with a confidence interval indicating that the effect was significantly different from zero. The category identified as involving parents to support instruction through feedback to the parents showed a weighted effect size of 0.42, however it was not significant, based on the confidence interval range containing zero.

The category identified as including the use of explicit or teacher-directed instruction showed that explicit instruction usage yielded a significant, weighted effect size of 0.58, while teacher-facilitated instruction and practice was not significant with a weighted effect size of 0.01. The final category, computer-assisted instruction, was not summarized as there were only two studies in this category.

Baker et al. (2002) summarize the results of their review of instructional interventions which are most effective for low-achieving students as: (a) providing teacher and students with information regarding student performance enhances mathematics achievement; (b) using peers as tutors enhances achievement; (c) providing specific feedback to parents may enhance achievement, and (d) direct, explicit instruction may be useful for enhancing student achievement.

The third review requiring further examination was that involving low-achieving, at-risk students (Barley et al., 2002). This is a comprehensive review which only in part addresses mathematics achievement. The sample for this review was K-12 students. It focused on low-achieving, at-risk students and examined instructional interventions in all academic areas. As seen with the prior studies, Barley et al. categorize the instructional interventions into categories as follows: (a) general instruction, which includes behavioral and constructivist strategies; (b) cognitive-oriented instructions in reading, writing and oral, and mathematics; (c) grouping structures, both across abilities and by ability; (d) tutoring structures, including professional tutors, volunteers and older students as tutors; (e) peer tutoring, and (f) computer-assisted instruction. Two of the categories were considered to be instructional approaches: general instruction and cognitive-oriented instruction. The other four categories were considered to be interventions. Each category was assigned to a different member of the research team. A total of 118 studies were synthesized by the researchers.

In summarizing the Barley et al. (2002) review, only those findings directly related to the academic area of mathematics were considered. The overall summary of each of the six categories (two approaches and four interventions) was reviewed however. Table 4 presents a summary of those studies included in the Barley et al. review dealing with mathematics, including the specific approach or intervention, the number of studies and their conclusions and the overall category conclusion.

Overall, the Barley et al. (2002) review results showed that instructional approaches such as those using cognitive-oriented instruction and interventions involving mixed ability grouping, tutoring and computer-assisted instruction, especially in mathematics, are effective strategies. These strategies, when used as prescribed by the classroom teacher, can be effective in increasing the mathematics achievement of low-achieving, at-risk students.

Table 4

Summary of Barley et al. (2002) by Category for Mathematics Studies Only

Category	Intervention Type	Number of Studies	Effectiveness Measure
General Instruction (approaches)	Behaviorist	3[a]	Even, Even, Mostly Positive
	Constructivist	1[a]	Mostly Positive
	Overall	15[b]	Inconclusive
Cognitive-Oriented Instruction (approaches)	--	2	Mostly Positive
	Overall	15	Mostly Positive
Grouping Structures (interventions)	Across Ability (cooperative learning)	4	Mostly Positive
	By Ability	2	Even, Mostly Positive
	Overall	18	Mostly Positive
Tutoring (interventions)	Professional tutoring	0	NA
	Volunteer tutoring	1	Even
	Student tutoring	0	NA
	Overall	23	Positive
Peer Tutoring (interventions)	Classwide Peer Tutoring (CWPT)	3	NA, All Positive, All Positive
	Peer-Assisted Learn Strategies (PALS)	3	Mostly Positive, All Positive, Mostly Positive
	Reciprocal Peer Tutoring	6	All Positive
	Other Forms	3	Even, Mostly Positive, NA
	Overall	30	Mostly Positive
Computer-Assisted Instruction (Interventions)	Math	12	All Positive, Mostly Positive
	Overall	21	Significantly Positive ($d = 0.37$)
Total	Overall	118	Mostly Positive

Notes: [a] Number of studies in mathematics area.
[b] Number of studies overall

The fourth review needing further examination was by Forness, Kavale, Blum and Lloyd (1997). This review is of interest due to both its purpose of reporting effective and non-effective teaching strategies in special education and in its simplicity. In this review, the research team summarized 18 meta-analyses on special education. In doing so, the researchers provide a relative power summary of the teaching strategies by comparing the effect sizes reported by the authors of the reviews. The seven most effective teaching strategies are shown, with their respective effect sizes, grouping notes and sources, in Table 5. It should be noted that some source meta-analyses may be somewhat outdated; however the comparison Forness et al. make regarding the teaching strategies is of value.

The final review to be examined more closely is the most comprehensive systematic review discovered during this effort. This review (Swanson, Hoskyn, & Lee, 1999) is a comprehensive review of over 30 years of intervention research focused on students with disabilities. The review was conducted in response to the many teachers and parents requesting H. Lee Swanson's assistance in finding something that would work with children who had learning disabilities. Like the Barley et al. (2002) review, this work covers all academic areas: not just mathematics. It included 275 studies conducted between the years of 1969 and 1999.

Table 5

Summary of Effective Teaching Strategies (Forness et al., 1997)

Teaching Strategy	Grouping Notes	Effect Size (*d*)	Authors, Date
Early Intervention	Overall If 30+ months	0.68 > 0.68	(Casto & Mastropieri, 1986)
Formative Evaluation	Overall	0.70	(Fuchs & Fuchs, 1986)
Cognitive-Based Modification	Overall	0.71	(Robinson, Smith, Miller, & Brownell, 1999)
Direct Instruction	Reading Math Overall	0.85 0.50 0.84	(White, 1986)
Behavior Modification	Behavioral Improvement Academic Improvement Overall	0.61 1.57 0.93	(Skiba & Casey, 1985)
Enhanced Reading Comprehension	Overall	1.13	(Talbot, Lloyd, & Tankersley, 1994)
Mnemonic Strategies	Overall	1.62	(Mastropieri & Scruggs, 1989)

Swanson et al. (1999) found that, overall and across all content domains, studies that included components related to instruction of strategy or the use of direct instruction provided larger effect size estimates than did those that did not involved the teaching of strategy or did not include direction instruction. Baker et al. (2002), referring to the Swanson et al. review, noted that only 10% of the intervention studies with students with learning disabilities involved mathematics. In these studies, the mean effect size estimate was 0.40, which is considered a moderate effect (Cohen, 1988). However, these effect sizes were among the lowest observed in the review, with the effect sizes associated with reading comprehension being equal to 0.72.

In analyzing the 275 studies in the review, Swanson et al. (1999) devised a comprehensive structure of defining instructional interventions, which may be of interest to all researchers looking at instructional interventions in the K-12 environment. Initially, the Swanson research team had devised a coding sheet which included 45 codes. As they proceeded to code the studies, they realized that the interventions could be considered to be different combinations of 20 components. Based on this, the team identified two general models of instruction: direct instruction and strategy instruction. Each model was composed of different combinations of the 19 components. Table 6 lists the components that are make up the two different models.

Curricular Interventions

It is evident from the number of systematic reviews and meta-analyses produced in the past that few studies have been done on the effectiveness of mathematics curriculum interventions. Based on one review identified, namely the What Works Clearinghouse (2004a) review of middle school curriculum materials, more research is necessary to evaluate the effectiveness of curricular interventions. It is hoped that the pending completion of the WWC review on elementary and high school mathematics curricula will be of value.

Table 6

Intervention Models and Their Components (Swanson et al., 1999)

Direct Instruction Model [a]	Strategy Instruction Model [b]
Breaking down a task into small steps	Elaborate explanations
Administering probes	Modeling from teachers
Administering feedback repeatedly	Reminders to use certain strategies or procedures
Providing pictorial or diagram presentations	Step-by-step prompts or multiprocess instructions
Allowing for independent practice and individually paced instruction	Dialogues -- students and teachers talking back and forth
Breaking the instruction down into simpler phases	Teacher asks questions
Instructing in small groups	Teacher provides only necessary assistance
Modeling a skill or behavior	
Providing set materials at a rapid pace	
Providing individual instruction	
Asking questions	
Presenting new materials	

Notes: [a] Any procedure that includes at least four of these components is considered direct instruction.
[b] Any procedure that includes at least three of these components is considered strategy instruction.

Instructional Interventions

As was evident from the discussion above, the categorization of instructional interventions is subject to varied interpretations. To support our identification and placement of a variety of studies addressing the effectiveness of different instructional intervention approaches, it may be of value to define the theoretical frameworks of instructional interventions and establish a framework to use.

Three frameworks for categorizing instructional interventions were discovered in a search of the literature. The first is one where selected instructional strategies are recommended for usage in the inclusive classroom (Lawrence, 1988). In this framework, the instructional interventions identified by the author are considered to be of value to not only at-risk students, but also for students with learning disabilities. The recommended instructional interventions are: (a) direct instruction for the acquisition of new skills, (b) peer tutoring, (c) cooperative learning, and (d) self-instructional strategies.

Another approach refers to Bloom's taxonomy and relates to the domains and systems approach was advanced by Marzano, Gaddy, & Dean at McREL (2000). This approach was developed as a means of defining instructional interventions for usage in a meta-analysis on classroom instruction where the primary concern was to avoid the mixing of "apples and oranges". Marzano et al. divided the interventions into four groups: (a) knowledge domain: including information, mental processes, and psychomotor skills; (b) cognitive system:

including processes that address storage and retrieval, basic information processing, communications and knowledge utilization; (c) metacognitive system: including processes that address goal specification, process specification, process monitoring, and disposition monitoring; and (d) self-system: including self-attributes, self and others, nature of the world, efficacy, and purpose – along with the internal monitoring mechanism that estimates the discrepancy between desired and perceived status.

The third and final approach is that defined above by Swanson et al. (1999) for usage in the development of their sizable systematic review addressing interventions for students with disabilities. Of primary interest in the Swanson approach is the identification and listing of the components parts for direct and strategy instruction, as described in Table 6.

Addressing the Mathematics Achievement Gap

One of the major considerations addressed in this systematic review is the goal of identifying the curricular and instructional interventions that are effective in reducing the mathematics achievement gap. To that end, a discussion to foster our understanding of what the mathematics achievement gap entails and how it can be addressed is warranted.

Defining the Mathematics Achievement Gap

In education, the term *achievement gap* generally refers to a persistent difference over an extended time in standardized assessment scores between two or more student subpopulations of interest. For example, the achievement gap in mathematics between White, non-Hispanic 13-year-old students and Black, non-Hispanic 13-year-old students was 38.9 points as measured by NAEP data in 2000, favoring the White student population (Snyder et al., 2004).

This achievement gap did decrease from the 1970s to the mid-1980s. However, since 1986, this achievement gap has been slowly widening. The same holds true for the achievement gap between White, non-Hispanic 13-year-old students and Hispanic 13-year-old students. The mathematics achievement gap between these three student subpopulations exists for all three levels of students who have participated in the NAEP testing program since the mid-1970s. Figure 1 is a graphical representation of the mathematics achievement gap fluctuations seen for 13-year-old U.S. students from 1973 to 2005, as measured using NAEP data (Snyder et al., 2004; U.S. Department of Education, 2006).

Though the average NAEP score for 2005 shown in Figure 1 is approximately 280 points, it should be noted that this average mathematics achievement scores is considered unacceptably low. A recently released summary of NAEP mathematics achievement scores for 2005 show that only 23% of 12-grade students perform at proficient or above levels while 77% perform below the proficient mark (Schneider, 2007).

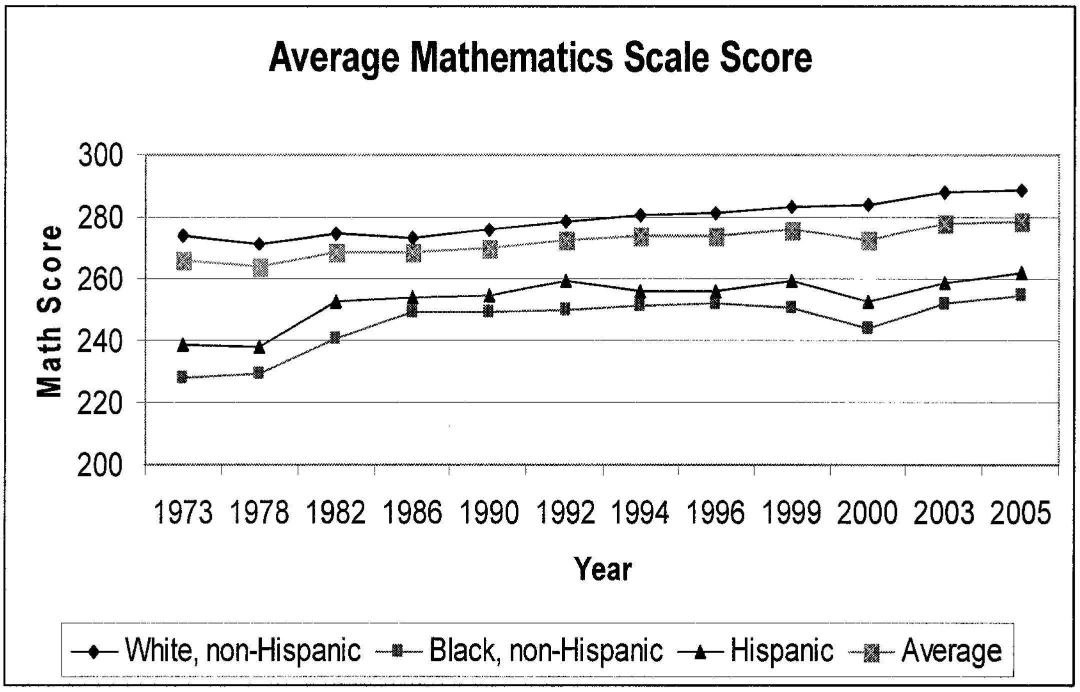

Figure 1. Average mathematics score for 13-year-olds by race/ethnicity.[1]

The mathematics achievement gap by gender, historically the gap in assessment scores between male and female students, has virtually disappeared over the last 30 years. Figure 2 shows the average mathematics scores by gender for 13-year-olds from 1973 to 2005.

As a means of setting a perspective on the meaning of the mathematics achievement gap, consider the following. At the end of high school, Haycock (2001) noted that 1 in 30 (3.3%) Latino students and 1 in 100 (1.0%) Black students can comfortably do multi-step problem solving and elementary algebra,

---

[1] Source: (Snyder et al., 2004; U.S. Department of Education, 2006), National Assessment of Educational Progress (NAEP) data.

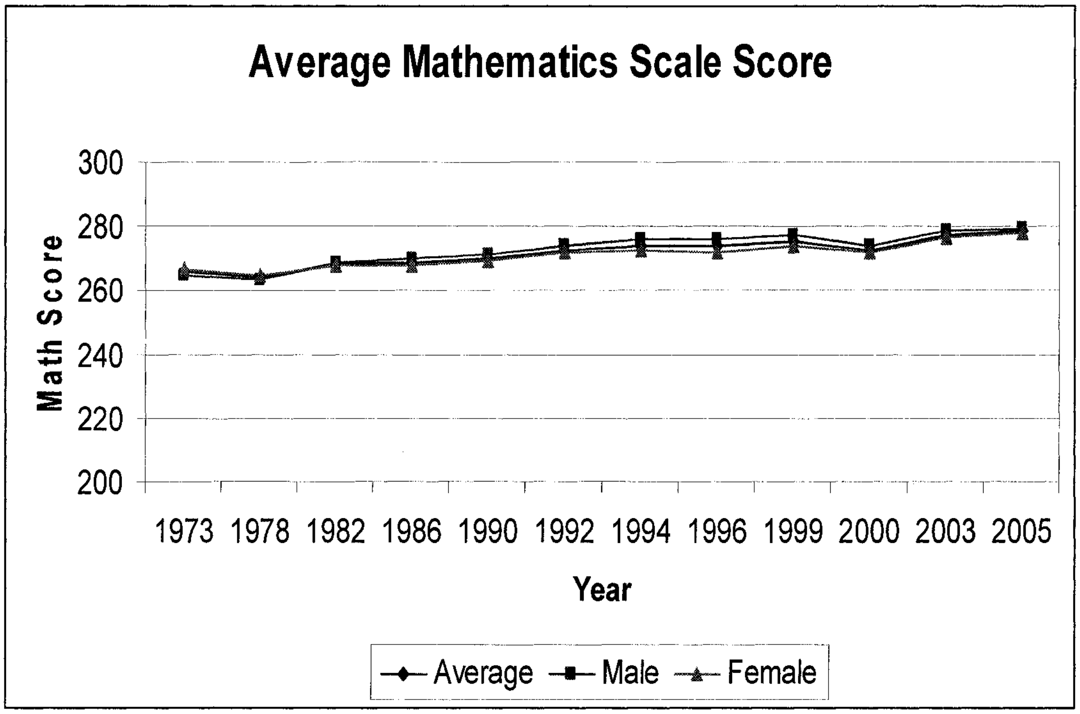

Figure 2. Average mathematics score for 13-year-olds by gender.[2]

compared to 1 in 10 (10%) White students. Additionally, the average Black or Hispanic graduating high school senior's mathematics skill and comprehension level is equivalent to the average 8th grade White student.

Causes of the Achievement Gap

Attempts to identify specific causes of the mathematics achievement gap have been many and varied. Unfortunately, no single causal factor has been identified to date. Jencks and Phillips (1998), among others, have convincingly laid to rest the notion that the gap is a result of biological or genetic factors

---

[2] Source: (Snyder et al., 2004; U.S. Department of Education, 2006), National Assessment of Educational Progress (NAEP) data.

related to race. Socioeconomic and cultural correlations, specifically related to poverty, have been suggested in the literature. Other factors associated with poverty, including low parental-education levels, inadequate pre-natal care leading to low birth-weight, and inadequate child-rearing practices may be contributing factors (D'Amico, 2001). Haskins and Rouse (2005), basing their comments on data from the Early Childhood Education Longitudinal Study, Kindergarten Cohort (ECLS-K), have noted the existence of a substantial mathematics and reading achievement gap in children prior to entry into kindergarten.

Causes for the achievement gap may also be related to the school and educational process itself. Lower teacher perceptions and expectations of minority students have been documented (Ferguson, 1998). However, Bol and Berry (2005), based on a survey of National Council of Teachers of Mathematics (NCTM) members, found that teachers attributed the mathematics achievement gap primarily to characteristics of the minority student population, including low motivation levels, low-levels of work ethics and poor family support structures. School-based factors which have been studied and may contribute to the issue include less rigorous course taking by minority students, lower funding levels of high-minority urban schools, classroom instructional practices, school policies, and teacher quality (Bol & Berry III, 2005; D'Amico, 2001; Darling-Hammonds & Sykes, 2003; Haycock, Jerald, & Huang, 2001; Lubienski, 2002)

Closing the Achievement Gap

Hanushek (2001), in an analysis of federal government intervention and the Black-White achievement gap, noted that school desegregation had a noticeable impact on reducing the gap between the 1970s and 1990s. In addition, legal action to more equitably fund schools across the U.S. had minimal impact. Family structure changes, including a decrease in family size and an increase in educational levels of Blacks families during this period may have been contributing factors. Recently, Berends, Lucas, Sullivan and Briggs (2005), in a significant analysis of multiple federal databases including NAEP and the High School and Beyond Study database have confirmed the reduction in the mathematics achievement gap between the years of 1972 and 1992 based on multiple socio-economic, cultural, family-related and state policy changes.

Haycock (2001), in examining methods of closing the achievement gap in education, states that what schools do matters greatly. Kati Haycock, the Director of The Education Trust, a private organization involved with educational research, has outlined the key lessons learned through research and experience. Her first observation is that standards are the key to increasing achievement. Haycock (2001) relates the experiences of the staff of The Education Trust in observing high poverty and high minority schools when she states that they were stunned to see how little was expected of the students and how low-level the few homework assignments were. She refers specifically to 11th grade students being given an assignment to color a poster summarizing their reading of *To Kill a Mockingbird*.

Haycock (2001) identifies the second lesson learned with the statement that all students must be involved in a challenging curriculum. "Current patterns are clearest in high schools, where students who take more-rigorous coursework learn more and perform better on tests. Indeed, the more-rigorous courses they take, the better they do" (2001, p. 9). To support this lesson, Haycock points out that in mathematics, students who complete a college prep curriculum in high school perform much higher on the National Assessment of Educational Progress (NAEP) than do students who complete only a couple of courses.

A third lesson learned and advanced by Haycock (2001) and The Education Trust is that students do need extra help. In fact, anything that a school or parent can do to assist a student is worthwhile. This includes before and after-school programs, summer programs, and doubling or tripling the amount of time students spend in the mathematics classroom during the school day.

The final observation advanced by Haycock (2001) is that teachers matter. Recent research shows that, "What schools do matters enormously. And what matters most is good teaching." (2001, p. 10). Contributions to our understanding of teacher impacts and quality by William L. Sanders and his colleagues from the University of Tennessee, based on their work with value-added assessments, have increased our understanding of the impact teachers have on student achievement (Sanders & Rivers, 1996).

Yet, as Haycock (2001) points out, large numbers of students from poor, minority backgrounds are taught by teachers who are not qualified in the subjects they teach. Research has shown that Black students are taught by a disproportionate percentage of unqualified teachers (Darling-Hammonds & Sykes, 2003; Haycock et al., 2001).

Haskins and Rouse (2005) have identified two programmatic approaches that hold promise for closing the achievement gap with pre-school children: (1) those that teach parents how to promote good child development and school readiness skills, and (2) those that directly teach poor and low-income children school readiness and social skills.

Can the achievement gap be closed for all subgroups? Singham (2003) suggests that we need to reset our focus on the difference between where all students are now and where we think they should be. He argues that there is little point to the attempt to eliminate the achievement gap by increasing Black students scores to the level of White students when only 27% of White students perform at or above the proficient level in mathematics.

Implications for the Achievement Gap

The implications of attempts to identify the means of closing the achievement gap in mathematics are numerous and of varying degrees of importance. One can argue, as Singham (2003) does above, that the need to reduce or remove achievement gaps is of little value when entire generations of students are not mastering mathematics. On the other hand, the reduction or removal of the mathematics achievement gap may have a large impact on the quality of life experienced by minority students or by those who have experienced generations of poverty. The task at hand, which is the identification of curricular and instructional interventions which may be effective in reducing the achievement gap in mathematics, is significant. Whether the achievement gap can be reduced is an unanswered question at this point. We may find that the most effective means of reducing the gap may be related to the improvement of teacher quality and the equitable access to high quality teachers and instruction for all students.

CHAPTER 3

METHODOLOGY

The selection of a methodological approach for a research project is typically made based on the nature of the research question and the researcher's level of expertise with the methodology. In this case, the selection of a systematic review with a meta-analysis methodology was made based on the research questions and on the author's desire to further his expertise with the selected methodology.

Research Design

The research design selected for this review is a systematic review with a meta-analysis methodology. The systematic review with meta-analysis methodology approach was selected as it is the primary approach used today to synthesize research results from multiple studies which address similar constructs (Hunter & Schmidt, 2004). A systematic review is more rigorous than a traditional synthesis review and attempts to reduce the influence of bias by extensively documenting the review process. Statistical methods (meta-analysis) may or may not be used to analyze and summarize the results of included studies which are considered similar enough to combine (Medical Research Council of South Africa, 2004).

The Systematic Review Process

There are generally two reasons to do a systematic review or synthesis of empirical research studies: (a) to summarize the studies accurately and succinctly, and (b) to search for important differences among studies that share a common focus (Baker et al., 2002). This review follows a systematic review approach adapted from the Campbell Collaboration and from the What Works Clearinghouse. The review includes the usage of a Study Design and Implementation Assessment Device (DIAD) adapted from the What Works Clearinghouse. The DIAD describes the systematic process for determining the degree to which a study allows for interpretation of conclusions for determining causal effects of an intervention (Valentine & Cooper, 2004).

Shadish and Myers (2004) released a policy brief for the Campbell Collaboration in which they addressed acceptable methodologies used in primary studies which could then in turn be used in a systematic review of intervention effectiveness. Their conclusion was that a wide variety of research designs could be used in one systematic review. In expressing this conclusion, Shadish and Myers related that clear patterns of intervention effectiveness can emerge across a variety of research designs. Additionally, the policy brief recommended that extensive coding of study characteristics be utilized in all systematic reviews. Most importantly, Shadish and Myers (2004) stated that the kind of research design, the sample size of the treatment group and the sample size of the comparison group should be coded so that weighted effect sizes can be calculated. These recommendations will be followed in this systematic review.

The *Study Review Standards* report released by the What Works Clearinghouse (2004b), specifies standards required of What Works

Clearinghouse (WWC) systematic reviews. These standards include the review of studies in three stages: screening for inclusion; determining the degree by which a study provides evidence of causal validity; and describing other important characteristics including outcome measures, intervention fidelity and statistical analysis. The primary elements of a systematic review are as follows (EPPI-Centre, 2004b): (a) A set search strategy for locating existing research reports using several pre-defined databases that are searched in a systematic manner; (b) The usage of a defined protocol or method of systematically reviewing all the existing research available to answer a research question; (c) Includes the potential for the review to be replicated by another researcher as explicit reporting of how the review process was conducted is recorded; (d) Includes a synthesis of the results of the studies reviewed; and (e) The conclusions are based on clear and pre-defined outcome measures.

Process Flow

The process flow involved with conducting a systematic review can be described as having eight steps, as defined by the Campbell Collaboration (Abrami, Bernard, Cobb, Nye, Turner, & Wade, 2005). These steps are as follows:

1. Formulate review questions.
2. Define inclusion and exclusion criteria.
3. Locate studies.
4. Select studies.

5. Assess study quality.
6. Extract data.
7. Analyze and present results.
8. Interpret results.

Another way of describing this process flow is through the use of a flow chart as presented in Figure 3. This figure demonstrates the adapted process defined by both the What Works Clearinghouse and the Campbell Collaboration and

A brief description of the process flow, as described in Figure 3, is as follows. In the definition phase of the systematic review, the problem statement and research questions are defined. This is followed by the development of three coding documents which are used to define the inclusionary criteria for studies to be considered as part of the review. An abstract review document was created to assist the research team in identifying studies for inclusion based on an abstract review only. A keywording document was created to identify the key characteristics of each study. The third document, the data extraction coding document, was developed for use in determining critical measures such as research design characteristics and study reliability and validity. These three documents make up the protocol for this review.

The citation selection phase includes a literature search for potential studies to be included in the review using a variety of resources and search

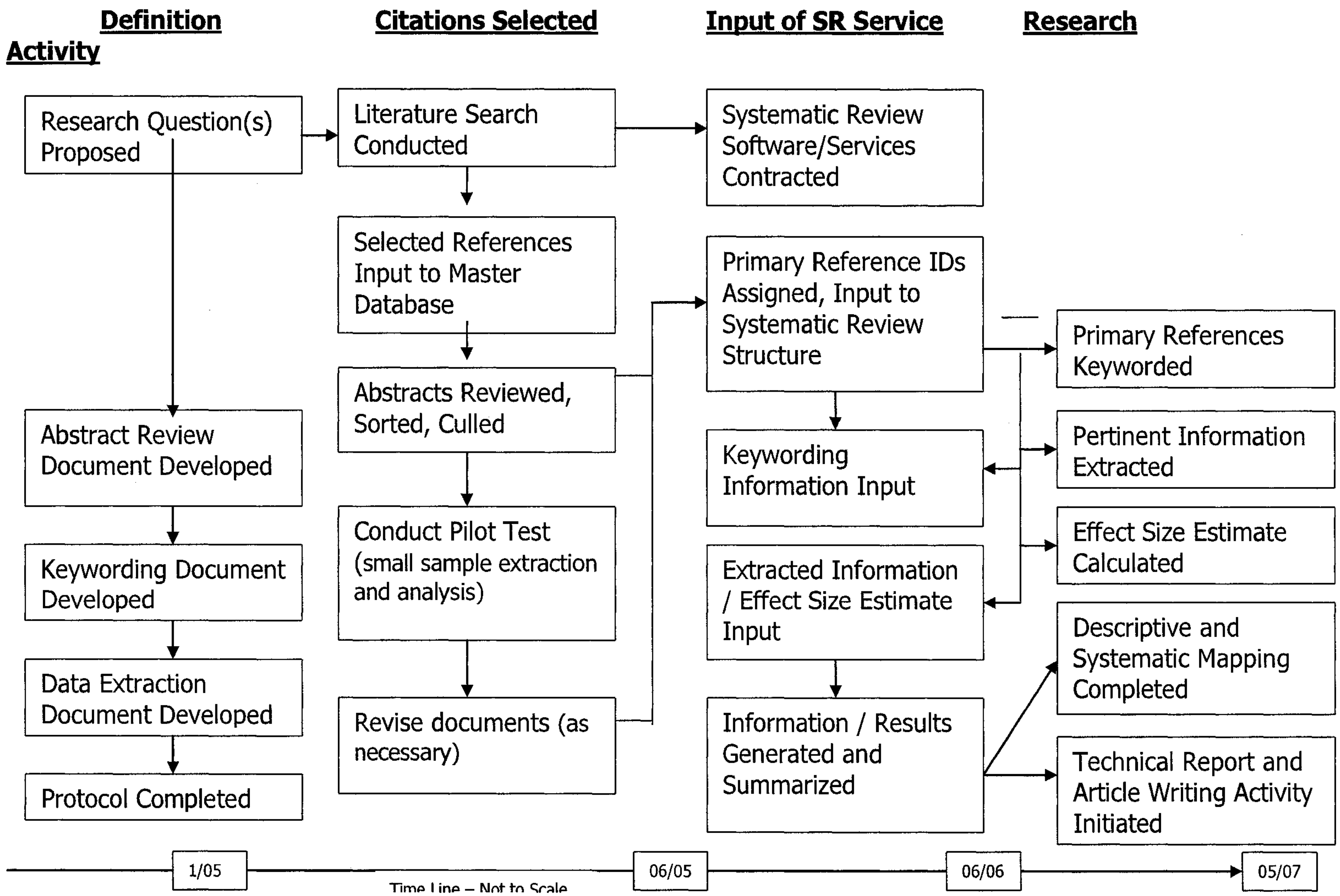

Figure 3. Systematic review process steps and flow

mechanisms. The resulting citations are combined into a master database of citations for review. The abstract review document was used to identify those citations with abstracts which may qualify for inclusion in the review. The testing of the coding documents through the usage of a small number of studies was done to test the usability of the documents. The documents were then revised as necessary.

The third phase of the process shown in Figure 3 includes the coding of each study using the keywording and data extraction documents, the calculation of effect size estimates, and the recording of the results in a computerized database. The review included the usage of EPPIReviewer (EPPI-Centre, Social Science Research Unit, Institute of Education, University of London): an online, document coding software system with meta-analysis capability (EPPI-Centre, 2004a).

The final phase of the process flow includes the summarization of results, comparison of effect sizes, and the development of the descriptive and systematic mappings of the results. The production of practice briefs, summary articles and a technical report of the systematic review are included in this phase.

Document Development

To assist the researcher in defining the coding to be used in this systematic review, three documents were developed, in accordance to the Campbell Collaboration and the What Works Clearinghouse (Shadish & Myers, 2004; Valentine & Cooper, 2004). These documents are: (a) The abstract review document; (b) A keywording document; and (c) The data extraction document.

The abstract review document, consisting of a single page of instructions for the review team, defines four criteria for usage in qualifying each study based on a review of the study abstract. The four criteria specified are used sequentially to determine if a study qualifies for further review. A copy of the abstract review document is provided in Appendix A.

The keywording document was developed to assist the research team in the initial screening of studies. This document takes the form of a study DIAD and is used to specify the key criteria of a study under consideration for inclusion. Characteristics identified and recorded using the keywording document are the study source; a description of the study including research design, intervention type, outcome, educational setting of the study; and a description of the study sample. A copy of the keywording document is provided in Appendix B.

The data extraction document was the final review document used and was the most comprehensive. This document also takes the form of a study DIAD and defines the specific criteria required of the studies for inclusion in the review. Criteria defined include: study identification, research design and study purpose; study sample characteristics; data collection and analysis, including reliability and validity measures; intervention description and characteristics; and the specific outcome description. A copy of the data extraction is provided in Appendix C.

A unique feature included in this review was the usage of a dichotomous approach (Yes or No) for response options to many of the questions included in the data extraction document. A common characteristic of many coding protocols reviewed prior to the development of the data extraction document for this review is the inclusion of multiple response options, including "Yes", "Maybe yes", "Maybe no", and "No". The researcher coding the study was provided a rubric for each question and used it to evaluate the information provided in the study. In this review, response options were often restricted to "Yes" or "No". Consequently, a question such as: "D.2 Was the reliability of the outcome measure selected for this study assessed and considered to be acceptable?" would be answered with a "Yes" or a "No", based on the information presented in the study. This procedure limited the ambiguity of the responses required of the researcher.

The Meta-Analysis Methodology

A meta-analysis methodology is a quantitative research methodology designed to describe a pattern of findings from a selected set of studies on a specific topic (Lipsey & Wilson, 2001). Smith and Glass (1977) first defined the methodology through the statistical standardization and averaging of treatment-control differences in psychotherapy studies. Completing a meta-analysis allows the researcher to report the efficacy of interventions using a standardized effect size value that can be compared among treatments or interventions.

Combining a systematic review with a meta-analysis methodology increases the value of the review by providing a higher-level of accountability for the conclusions made in the review (Hunter & Schmidt, 2004; Lipsey & Wilson, 2001). The utilization of this methodology goes beyond the qualitative summary or "vote-counting" exercise presented with other synthesis reviews. In addition, the application of a meta-analysis methodology can sometimes illuminate findings and relationships "that are obscured in other approaches to summarizing data" (2001, p. 6). Finally, the utilization of a meta-analysis methodology allows for the comparison of effect size magnitudes of different treatments across studies resulting in a better overview of the effectiveness of specific interventions (Xin & Jitendra, 1999).

Methodology Implementation

The implementation of the research methodology for this review is described in detail to ensure both transparency in the research process and to provide a description for others interested in replicating this approach. The extensive usage of the processes described by the Campbell Collaboration and the What Works Clearinghouse is noted.

Literature Search Strategy for Identification of Relevant Studies

The strategy defined for the search to identify relevant studies was based the need to canvas published and unpublished education studies. Consequently, four separate electronic databases were selected for the search. These four electronic databases were the Educational Resources Information Center (ERIC) database, using both the ERIC Cambridge Abstract Services (CAS) search engine and the ERIC OCLS (First Search) search engine, the PsychINFO database, the Dissertation Abstracts International (DAI) database, and the UMI ProQuest Digital Dissertations electronic database.

Database Search

The literature search was conducted according to *The Campbell Collaboration: Information Retrieval Policy Brief* (Rothstein, Turner III, & Lavenberg, 2004). The database search was limited to studies published between 1990 and 2004, for a total of fifteen years maximum. Additional studies published after 2004 may have been added if discovered during the article collection and review process. No restriction was placed on publication type, with journals, books, conference proceedings, monographs and dissertations (published or unpublished) being considered acceptable for inclusion.

The electronic database search was conducted in December 2004 using identical keyword searches. The list of keywords and Boolean search terms used in the search were: *mathematics achievement* AND *secondary education* OR *postsecondary education* OR *grade 6.*

Once the four database searches were conducted, the references obtained were transferred into Reference Manager, Version 11.0 (Thomson/ISI Researchsoft). The references were then merged into one database and the removal of duplicate references was completed.

Prior Reviews and Reference Lists

A number of other synthesis reviews and meta-analyses addressing secondary mathematics interventions were located during the search. These meta-analyses were reviewed in Chapter 2. Using the reference lists from these reviews, an ancestral search was conducted and the lists were compared to those in the master database. Those references identified as missing were added to the master database and then screened in the same manner as the original reference list.

Internet Search and Manual Searches

An Internet search were also conducted using the same keywords as used in the electronic database searches. Two different Internet searches were conducted: one using Google Scholar (Google Corporation) and one using AltaVista (Overture Services, Inc.) on May 14, 2005. The top 100 references identified in each search were compared to the master database. Only a couple of additional citations were identified from this search as most had previously been located through the electronic database search.

A manual search was also conducted of those journals most closely related to the research topic. This search was done through a review of the table of contents of each issue of the journal between 1990 and the end of 2004. If a potentially promising article was identified by title during the manual search, the abstract of the article was reviewed using the same criteria employed in the abstract review. Those citations identified as missing were added to the master database and screened in the same manner as the original citation list. Manual searches were done of the following journals: *Educational Studies in Mathematics*, *Focus on Learning Problems in Mathematics*, *Journal for Research in Mathematics Education*, *Mathematics Teacher*, and *School Science and Mathematics.*

Coding of Studies

The coding of studies for inclusion in the systematic review was completed in three separate phases. The first phase involved reviewing the 3,911 citation abstracts compiled in the database with the intent of identifying those articles which met four basic inclusionary criteria. The second phase was the keywording of study characteristics. The third and final coding phase was the complete data extraction of pertinent information including study purpose, research design specifications, characteristics of intervention and outcome measures, and effect size determination. Figure 4 provides a graphical representation of the three phases of coding employed.

Inclusionary Criteria

The basic criteria required of a study for inclusion in this review were: (a) the participants were secondary or postsecondary students involved in the study of mathematics; (b) must have included an intervention or action to change a student's experience or include a systematic change in the schooling (i.e., block scheduling); (c) must include an outcome resulting from the intervention, such as a score on a mathematics assessment test; and (d) must be quantitative in

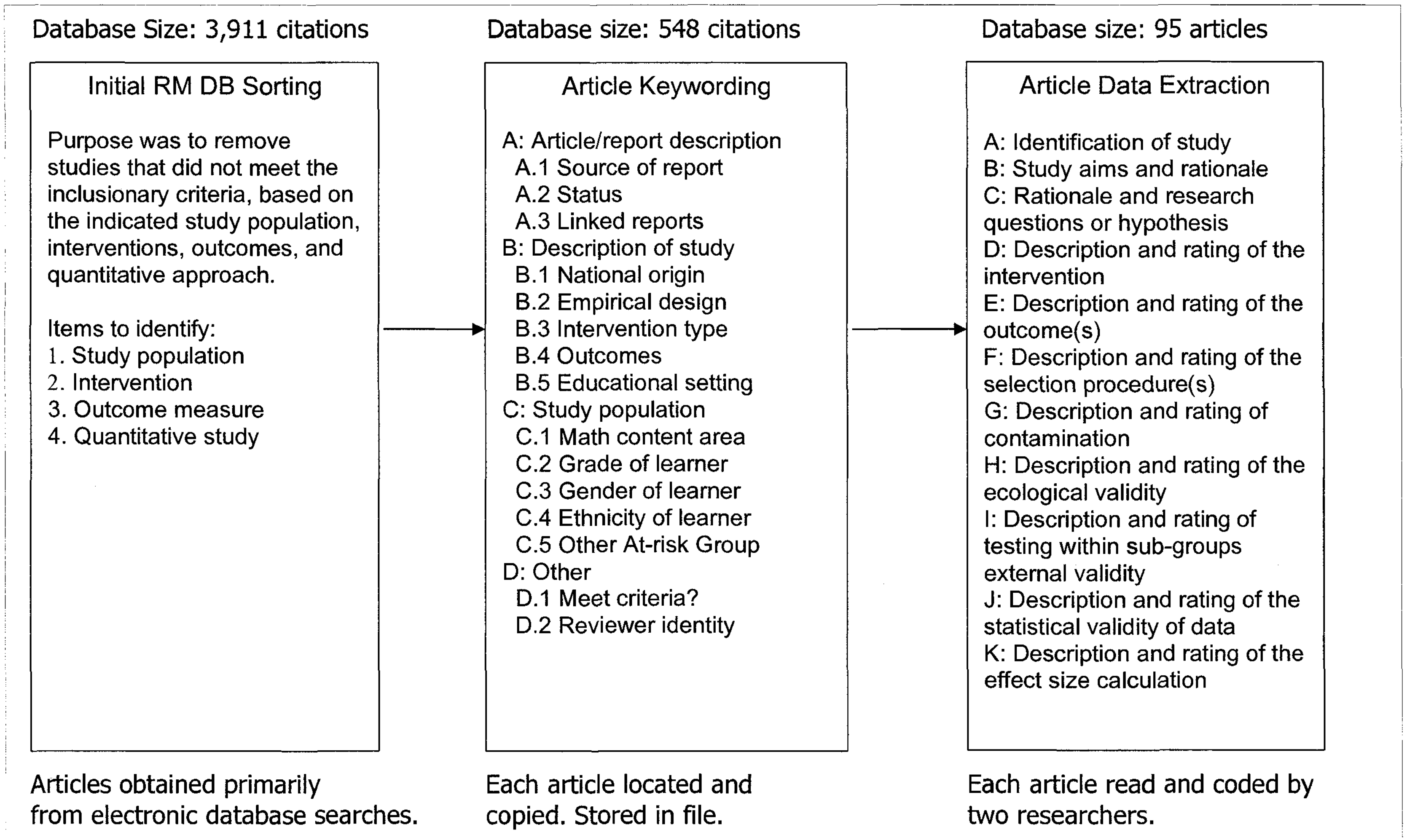

Figure 4. Mathematics achievement database coding procedure.

nature and use numerical values to measure an outcome variable. Due to the complexity in nature of educational interventions, more specific criteria for inclusion were developed to determine inclusion.

Population, Setting, and Content Area

For inclusion in the systematic review, studies must have indicated that the interventions were delivered to secondary or early postsecondary students (grades 6-14, with grades 13 and 14 being considered early postsecondary) middle school through early college, ages 12-20+ years, in regular school settings during school hours. Special education classrooms, private, parochial, alternative, residential schools, and correctional institutions would be eligible school settings. Home-schooled students, grades 6 – 14, would be included. Any qualifying school in any region or country would be eligible for inclusion. In addition, the study must have specified information on at least one of the following at-risk characteristics: gender, race/ethnicity, socioeconomic status, prior low-academic achievement, disability or other at-risk characteristic.

The students identified had to be involved in the study of some area of mathematics. Content areas included were the study of general mathematics, typically at grades 6 or 7, through college-level calculus.

Research Designs Included

Only experimental or quasi-experimental designs are eligible for inclusion. The intervention and control groups could be randomly or nonrandomly assigned but, if nonrandom, must have been matched or based on evidence of initial equivalence on key demographic variables and/or pretests. Control groups could represent placebo, wait-list, no treatment, or "treatment as usual" conditions. Research designs considered acceptable included multi-group quantitative and one-group pretest posttest. Studies using regression, regression discontinuity and interrupted time-series designs were also included. Initially, single-group correlational studies, single-subject designs, and qualitative designs were identified and included if, in the reviewers' consideration, a quantitative measure was also involved. In the final data extraction, studies that were identified as primarily single-group correlations, single-subject studies, and qualitative designs were removed.

Types of Interventions

Due to the tremendous number of single interventions and combinations of interventions which could be delivered simultaneously within the classroom, the school building or the school district as a whole, a discussion of how the interventions were identified and coded in the studies is warranted. Potential interventions were categorized in three general groups: instructional interventions, curricular interventions and all other interventions.

The curricular and instructional interventions have been defined. The third type of intervention was identified which included a number of different interventions, including: (a) technology-based interventions; (b) grouping or tracking interventions which include the grouping of students by ability, gender or other means; and (c) school restructuring interventions involving the whole school such as reducing class size, school-within-a-school, and alternative schools.

In defining the coding procedures, a specific definition of each type of intervention with more concise characteristics was created to assist the research team to correctly identify the interventions. Consequently, the interventions identified are described as: (a) curricular; pedagogical/instructional; grouping structures/tracking; personal/affective; professional development/teacher in-service training; school restructuring; technology-based; and tutoring strategies.

Curricular interventions are those associated with the structure of the content materials in an educational context. A change in these interventions involves changes in content, order of presentation of content, or theoretical basis of the content.

Pedagogical or instructional interventions are those in which the method or approach of content delivery can be altered by the teacher. In doing so, the teacher selects the pedagogical or instructional intervention. Examples are direct instruction, constructivist approach, or teaching of problem-solving skills.

Grouping structures or tracking interventions include strategies and techniques that involve grouping students in a classroom or school environment for delivery of instruction. This grouping could be done by level of prior academic success or by gender.

Personal or affective interventions are those designed to enhance self-esteem, such as retreats, classroom-based discussions, and counseling. These interventions typically are aimed at improving student attitudes towards mathematics.

Professional development or teacher in-service training interventions are those in which the teacher is the recipient of the training. The measure of a student outcome for this type of intervention would be through the teacher's delivery of the intervention's impact to the student: a two-step process.

School restructuring interventions are non-instructional or non-curricular interventions in which the whole school or school environment is altered. These interventions would include the implementation of block scheduling, the reduction of class size, or the implementation of a school-within-a-school approach.

Technology-based interventions involve the use of computers or graphical calculators directly as instructional aids. These interventions could include computer-aided instruction (CAI), integrated learning systems (ILS), and other computer hardware and software approaches.

Tutoring interventions include strategies such as peer tutoring, mentoring, or one-on-one instructional assistance by senior citizens or older students. Tutoring interventions are typically introduced into the classroom by the teacher, so can be considered teacher-centered interventions.

Outcomes Included

Quantitative data must have been reported in the study for at least one outcome variable for which a numeric mean difference (i.e., treatment versus control) effect size could be calculated. These outcomes were required to be cognitive outcomes such as student academic achievement as measured by standardized, teacher-made, or researcher-made assessments, or by performance measures of cognitive skills such as number of correct responses. At least one outcome variable had to represent mathematics problem solving. In general, student grades received in a class or letter grades on classroom assignments were not considered valid measures of assessment. Process outcomes, such as frequency of requests for assistance, and affective outcomes, such as attitude towards mathematics, were not consider as valid assessments.

Statistical Procedures and Conventions

Standard procedures for analysis per the conventions held for systematic reviews and a meta-analysis methodology were followed in this review. Intervention effects were measured using the standardized mean difference effect size statistic (Cohen, 1988; Lipsey & Wilson, 2001).

Conceptually, the standardized mean difference effect size statistic is based on the existence of a two-variable relationship in which a pre-post contrast is made on the same construct prior to exposure to the intervention and after the exposure. The difference between the pre-intervention mean and the post-intervention mean is assumed to be the result of the intervention only. This pre-post contrast is compared to a pre-post contrast measured with a control or non-treatment group which does not receive the intervention.

The sign of the effect size is meaningful. The standardized mean difference is derived by subtracting the control group pre-post difference from the treatment group pre-post difference and dividing the result by the pooled standard deviation of the measure. This results in an effect size estimate in which a positive value indicates that the treatment resulted in an increase in the outcome measure (primarily identified as student mathematics achievement in this review) over that seen in the non-treatment or control group. In turn, a negative effect size estimate would indicate that the outcome variable decreased in value due to the effect of the experimental intervention.

The effect size index used was Cohen's *d*, defined as the difference between the experimental and the control group means divided by the pooled standard deviation (Cooper & Hedges, 1994). Mathematically, the effect size calculation is represented as:

$$ES = \frac{X_e - X_c}{s_p}$$

where ES is effect size, X_e is the mean score of the experimental or treatment group, X_c is the mean score of the control or non-treatment group and s_p is the pooled standard deviation of the two groups.

The majority of the research designs employed by the studies reviewed were experimental and quasi-experimental designs, due to the study selection criteria used. Most were also between-group designs, though some within-subjects designs are included. Most of the studies involved pre-post assessments while others were post-assessment only. Most of the studies also utilized a treatment and control group model.

Interpretation of the effect sizes reported for each study included in the review was done using guidelines from Lipsey and Wilson (2001) and the What Works Clearinghouse (Valentine & Cooper, 2003; What Works Clearinghouse, 2006b). Additional technical resources were used to determine the best approach to define, validate and summarize the results of the reviewed studies, based on how the studies reported summary data (Hedges, 1981). Appendix E provides information on how the effect size estimates were determined for each research design and statistical analysis procedure encountered during the review.

Differences in sample sizes were encountered both within the same study and across multiple studies. As effect size estimates for larger sample sizes are more exact than those based on smaller sample sizes, weighing of effect size estimates based on the sample size involved is necessary in conducting a meta-analysis. To do this, an inverse variance weight of the effect size approach, as described by Hedges (Hedges, 1982; Hedges & Olkin, 1985) was used.

Confidence intervals for each effect size estimate were calculated using the process described by Lipsey & Wilson (2001). The confidence interval end points indicate the high and low values (range) in which a population mean is likely to reside based on the observed data. Statistical significance of an effect size estimate would then be indicated by the range of the confidence interval not including the value of zero.

Reliability Monitoring

Interrater reliability is a concern with most systematic reviews where multiple raters are used. To ensure interrater scoring consistency, reliability was monitored throughout this review. During both the initial article abstract review and the keyword coding, multiple interrater reliability checks, consisting of independent coding of a set of multiple, relevant studies by each of four reviewers (the author and three other CLT-West doctoral fellows), were conducted. For each of the reliability checks, an interrater reliability agreement of 90% was established as the benchmark by the author prior to the start of the review. A percentage agreement method (Morgan, Gliner, & Harmon, 2006) was used to evaluate the level of agreement between raters. In a similar manner, multiple interrater reliability checks were conducted for the data extraction coding involving the author and one other CLT-West doctoral fellow.

Prior to initiation of each phase of the coding process, the reviewers completed a training session conduced by the author on how to identify and code the studies. Each reviewer then coded ten studies as an initial interrater reliability check. A similar interrater reliability check was done with the coding of multiple studies during each of the three different coding phases. A comparison between the author's coding and the coding of each of the other reviewers resulted in an overall interrater reliability of 92% agreement for the six interrater reliability checks done through out the review. The range of percentage agreement for the six interrater reliability checks conduced was 96% to 90%.

Limitations

Two limitations for this review need to be discussed. The first limitation is related to the quality of the search conducted to obtain potential studies for the review. Though the search strategy included searching four electronic databases, a manual search of five key publications and an Internet search, the chance of overlooking studies does exist. Additionally, no contact was made with the authors of previous studies for the solicitation of their input of other potential studies. A second area of concern is the usage of multiple individuals for the reviewing and scoring of the studies obtained. Regardless of the level of training provided and the expertise of the individuals involved, the potential for errors in the sorting, coding and extraction of key data from complex research studies still exists. Each of these limitations would have the impact of reducing the reliability of the conclusion derived from this review. The use of caution in the interpretation of the results of this review is warranted.

CHAPTER 4

RESULTS AND FINDINGS

Two literature mapping approaches are used to summarize the characteristics of the literature base reviewed. The first mapping approach is referred to as a descriptive mapping (EPPI-Centre, 2006) and is the classification of the range of studies encountered in the literature search as a whole. This mapping approach can be used to describe the scope of existing research found in the topic area, to show relevance of research efforts in the topic area, and to provide context for interpreting the results of the systematic review synthesis. The second mapping approach is termed a systematic mapping (Bennett, Lubben, Hogarth, Campbell, & Robinson, 2005) and generally refers to the mapping of the specific characteristics of the studies meeting inclusion criteria in a systematic review synthesis.

The results of this review are presented in a limited descriptive mapping of the entire corpus of 3,814 citations initially located for the review. This group of citations with abstracts was then subjected to the first of three coding procedures involving a review of the abstracts. From that initial group, 592 citations met the inclusionary criteria and were then subjected to the second coding phase involving keyword coding. A systematic mapping of this intermediate group of qualified studies is presented. The final coding procedure, the full data extraction coding, resulted in a group of 137 studies which met the complete inclusionary criteria for the review. A systematic mapping of this group of studies is also included.

Descriptive Mapping of the Literature Base

Approximately 3,911 citations were identified as potential studies for this systematic review based on the literature search process described previously. Ninety-seven of these citations were immediately removed as they either fell outside the dates for consideration (1990 through 2004) or were duplicates of other citations in the master database. This left a total of 3,814 citations for consideration. These citations with abstracts were imported into a master database and subjected to a three-phase coding process.

Abstract Coding

The abstract coding process was applied to the initial master database of 3,814 citations with abstracts. The coding process involved the use of four questions to query the full citation, including the abstract. Those four questions were focused on the key inclusionary criteria of the review: the sample; the intervention; the outcome and the quantitative measure. The questions were: (1) "Does the study sample appear to address the review questions?"; (2) "Does the study appear to examine the effects of an intervention?"; (3) "Does the study appear to measure an outcome?"; and (4) "Does the study appear to be a quantitative study?".

As shown in Table 7, an analysis of the results from the abstract coding process indicates that the vast majority of the initial 3,814 citations with abstracts did not meet the inclusionary criteria as posed by the four screening questions. From the initial 3,814 citations, only 592 citations, or 15.5% of the total, met the four criteria required of studies for inclusion in the systematic review. This screening process resulted in a removal of 3,222 citations, or 84.5% of the total citations.

Table 7

Abstract coding results: Sorting of full literature base by inclusionary criteria

Inclusionary Criteria *	N	%
Study Population	1,087	28.5%
Use of Intervention	681	17.9%
Outcome Measured	603	15.8%
Quantitative Study	592	15.5%

* Citations required to meet all four criteria for inclusion. If study did not meet first criteria, it was removed from consideration. Of the total, only 17.9% met the first and second criteria requirements. Only 15.5% met all four criteria.

The impact on the reduction of citations for each individual question was tracked during the abstract coding. By not meeting the criteria established by the first question, which was the identification of appropriate sample population, 2,727 citations (71.5%) were eliminated. The majority of these citations were eliminated due to the sample being primarily elementary-level students instead of secondary or early-postsecondary students. The second screening question, focused on interventions, was responsible for eliminating another 406 citations (10.6% of the total). It appears that the majority of these citations was eliminated as the citations involved were primarily descriptive in nature and did not include an intervention or did not appear to report on empirical research.

The first two criteria questions accounted for the elimination of 82.1% of the citations.

The third screening question, "Does the study appear to measure an outcome?" was responsible for eliminating another 78 citations (2.0% of the total). Once a citation was screened for the first two questions and advanced, few were removed for lack of outcome measure. Fewer still were removed due to the fourth screening question: "Does the study appear to be a quantitative study?" after advancing through the first three questions. Only 10 citations (0.3% of the total) were removed due to the fourth question.

Characteristics of the Literature Base

A simple description of the characteristics of the body of citations identified makes up the descriptive mapping of the literature base. These characteristics include a summary of the publication dates of the citations, the publication status of citations identified, and the types of publications located in the literature search. This information is presented in Table 8.

An analysis of the publication dates shows a roughly equivalent distribution of the citations over three groupings of dates by 5-year interval from 1990 to 2004. A few additional citations were included from 2005 and were added after the literature search as conducted. Individually, the year 1993 provided the largest number of citations identified, with 327 citations, or 8.6% of the total, being published that year. The year 2004 provided the fewest citations in the group, with 83 citations or 2.2% of the total. The median publication date was 1997.

Table 8

Descriptive characteristics of the full literature base

Characteristic	N	%	Characteristic	N	%
Publication Year			Type of Publication		
1990 - 1994	1,327	34.8%	Journal	2,574	67.5%
1995 - 1999	1,386	36.3%	Book or Book Section	39	1.0%
2000 - 2004	1,098	28.8%	Dissertation	705	18.5%
2005	3	0.1%	Report	480	12.6%
Total	3,814	100.0%	Other	16	0.4%
			Total	3,814	100.0%
Form of Publication					
Published (peer review)	2,645	69.3%			
Unpublished	1,169	30.7%			
Total	3,814	100.0%			

The vast majority of the citations identified were published, with 2,645 citations or 69.3% of the total falling into that category. The remaining 1,169 citations, or 30.7% of the total, were unpublished documents.

Of the types of publications making up the literature base, the largest percentage of publication types was journal articles, with 2,574 citations, or 67.5% fitting this description. Dissertations made up the second largest publication type with 705 studies, or 18.5%, belonging to this type. Books or sections of books (chapters) accounted for 39 studies, or 1.0% of the total. Reports, particularly reports made available on the Internet, accounted for 480 citations, or 12.6% of the total. Other publication types, including conference presentations, accounted for 16 citations, or 0.4% of the total studies.

Systematic Mapping of Qualified Studies

The systematic mapping of qualified studies meeting the inclusionary criteria, as described by Bennett et al. (2005), represents a mapping of 592 studies. Data for this mapping analysis was gathered during the keyword coding process.

Keyword Coding

After the abstract coding process was completed, the collection of those 592 citations advancing through the abstract review screening process was done. The author was tasked to acquired full-text copies, in either electronic or hardcopy format, of each study. The author was able to locate and obtain 548 of the 592 citations, or 92.6% of the citations. The majority of the 44 citations not located proved to include inadequate or incorrect citations. The collection of full-text copies was done primarily through online access to full-text electronic publications and through inter-library loan. A small number of studies were only available as microfiche.

The research team opted to collect and store the full-text copies in electronic format, whenever possible. The electronic versions of the studies were then handled in two different manners. Those studies that were available in an electronic format and were 100 pages or less in length were printed and stored in a hardcopy format. These hardcopy studies were distributed among the reviewers for coding. The electronic format of these articles were also stored on a compact disk (CD) and distributed to the reviewers for review purposes.

Studies that were available electronically but were larger than 100 pages in length (primarily dissertations) were only stored electronically on a CD and distributed in that format to the reviewers. By having access to each of the studies in this review in an electronic format, the research team was able to reduce the paper handling requirements and the amount of file storage needed for the hardcopies, while making the transport the studies more efficient. Of the 548 studies obtained, 436 (79.6%) were collected in an electronic format while 120 (21.9%) were only available as hardcopy dissertation copies, microfiche, or as books.

These 548 studies collected were subjected to the keyword coding process whereby one reviewer coded the initial, relevant information from each study. A second reviewer completed a semi-independent coding of the study in a fashion identical to that used by the primary reviewer. The second review should be considered semi-independent as the second reviewer did have potential access to the first reviewer's comments via the online coding system. After both reviewers completed their reviews, a comparison of the results was made. If a disagreement in coding outcomes existed, a consensus process was employed to reconcile the differences. As described prior, interrater reliability checks were done throughout the review with an overall interrater reliability of 92% agreement reached across the six checks.

Characteristics of the Qualified Studies

A detailed description of the characteristics of the body of qualified studies for this review makes up the first level of systematic mapping of the literature base. These characteristics are grouped according to the sections making up the keywording document as shown in Appendix B. Section A of the keywording document includes a summary of the study source information. Additional information, including publication date and type of publication, are noted. This information is presented in Table 9. Section B of the keywording document presents a detailed description of the study, including the country of origin, type of empirical design, classification of intervention type, type of outcome, and the educational setting of the study. The third section, Section C, provides a detailed description of the study population, including the mathematics content area, grade of learners, and whether the study included information on gender, race/ethnicity or other at-risk factors being involved. Sections B and C are summarized in Table 10 and Table 11, respectively.

As was seen in the descriptive mapping of the full body of literature, an analysis of the publication dates again shows a roughly equivalent distribution of the studies over three groupings of dates by 5-year interval from 1990 to 2004, with one study from 2005. The year with the largest number of studies included is 1998, with 51 studies. The year 2004 had the fewest number of studies with 13 total. This may be again related to the dates of data collection and the

Table 9

Keyword coding results, Section A: Source of reports

Characteristic	N	%	Characteristic	N	%
Publication Year			Linked Reports		
1990 - 1994	212	35.8%	Not linked	540	99.3%
1995 - 1999	224	37.8%	Linked	4	0.7%
2000 - 2004	155	26.2%	Don't know	0	0.0%
2005	1	0.2%	Total	544	100.0%
Total	592	100.0%			
Publication Status			Type of Publication		
Published (peer review)	212	39.0%	Journal Article	338	57.1%
In Press	0	0.0%	Book or Book Section	0	0.0%
Unpublished	332	61.0%	Dissertation	222	37.5%
Total	544	100.0%	Report	32	5.4%
			Other	0	0.0%
Source of Report			Total	592	100.0%
Citation	8	1.5%			
Contact	0	0.0%			
Hand search	8	1.5%			
Unknown	4	0.7%			
Electronic database	524	96.3%			
Total	544	100.0%			

time needed to catalogue each study at the end of the year. The median publication date was 1997. It should be noted that a total of 592 citations are included in the analysis of publication date though only 548 of these studies were located for the keyword coding procedure.

Unpublished studies outnumbered published studies by 332 to 212. Based on this data, for those studies meeting the initial inclusion criteria, over 60% of the studies were unpublished while in the full body of literature, close to 70% were published. This shift can be attributed to the high number of dissertations included in this group of studies.

Over 96% of the included studies were identified through the use of electronic databases in the search process. Only 20 studies were from other sources. It should be noted that, as the initial search for studies was done via electronic database searches, if a duplicate study was later located by an ancestral or manual search, the duplicate study would not be added to the master database, biasing this characteristic to the initial electronic database source.

Four of the studies (less than 1%) were identified as being linked to other studies in this group. This linkage could be due to the inclusion of a dissertation and the subsequent publishing of the results of the dissertation as a journal article. It was determined later that, under some conditions, a linkage between two studies could be related to a study including two or more completely independent studies with independent populations. This relationship was not accounted for in the keywording coding and so was not associated with the four studies identified as being linked to another study here.

The largest percentage of publication types was journal articles, with 338 studies, or 57.1%, being identified in this category. Dissertations, which again made up the second largest publication type, included 222 studies, or 37.5% of the total. Reports accounted for 5.4% of the studies, or 32 studies total. A total of 592 studies were included in this analysis.

Table 10 presents a detailed description of study characteristics coded using Section B of the keyword coding document. The first characteristic detailed is the country in which the study occurred. As can be seen in the table, the Unites States, with 464 mentions or 86.7% of the total studies, is the country of origin for the vast majority of studies. Other countries received 80 mentions, or 14.7% of the total. Canada, Great Britain, and the Nederlands are the next highest ranking countries identified in the studies included in the keyword coding, with ten, nine, and nine mentions respectively.

A key characteristic of the studies included in the keyword coding is the research design used in the studies. Identified as empirical design in Table 10, it is seen that multi-group quantitative studies account for 71.7% or 391 of the studies coded. Sixty-nine of the studies, or 12.7% of the total, were coded as "none of the above", indicating the inability of the reviewers to clearly identify the type of research design used. Five other design types, including one-group pretest/posttest, single subject, qualitative, multiple regression and one-group correlations designs made up another 15% of the studies. Two studies each were identified as regression discontinuity and interrupted time series designs.

A second key characteristic of studies included in the keyword coding is the identification of the intervention or treatment used in the studies. Due to the tremendous number of single interventions and combinations of interventions which could be delivered simultaneously within the classroom, the school building or the school district as a whole, identification of interventions is at times difficult to determine. Additionally, fidelity of an intervention is often compromised in its implementation due to the nature of education and the classroom and students involved. Consequently, it is difficult at times to identify and classify a single intervention based on a published or unpublished study.

With the above in mind, the research team identified the pedagogical or instructional intervention type as having the largest number of mentions, with 163 studies, or 30.0% of the total number of studies. Technology-based interventions, primarily computer-aided instruction (CAI) and calculator-based instruction, made up the second largest group of interventions with 101 mentions or 18.6% of the total. "Other" interventions, which would include interventions that could not be identified by the research team, received 101 mentions. Curricular interventions, with 48 mentions or 8.8% of the total, and school restructuring interventions, with 46 mentions or 8.5% of the total, were ranked fourth and fifth highest respectfully.

Identification and classification of study outcomes during keyword coding show that cognitive outcomes were by far the largest category of outcomes measured with 436 mentions or 63.3% of the total studies coded. Based on the question used to review the abstracts and the requirement of evidence of a quantitative measure, this is not surprising. Affective outcomes, in particular student attitudes towards mathematics, provided the second largest category of outcomes, with 154 studies identified or 22.4% of the total. The remaining outcomes identified as process, "other", or indeterminate made up the majority of the remaining outcomes identified, with 97 mentions or 14.1% of the total. Enrollment in higher math also received a few mentions. It should be noted

Table 10

Keyword coding results, Section B: Description of the study

Characteristic	N	%	Characteristic	N	%
Country of Study			Outcomes		
United States	464	85.3%	Cognitive	436	63.3%
Other	80	14.7%	Affective	154	22.4%
Total	544	100.0%	Process	31	4.5%
			Enrollment in Higher Math	2	0.3%
Empirical Design			Other	35	5.1%
Multi-Group Quantitative	391	71.9%	Indeterminate	31	4.5%
One Group Pretest Posttest	22	4.0%	Total	689	100.0%
One Group Correlational	10	1.8%			
Multiple Regression	14	2.6%	Educational Setting		
Single Subject	17	3.1%	Middle/junior high school	179	32.9%
Qualitative	17	3.1%	High school	223	41.0%
Regression Discontinuity	2	0.4%	Community/Junior college	14	2.6%
Interrupted Time Series	2	0.4%	Four-year college	49	9.0%
None of the above	69	12.7%	Home	1	0.2%
Total	544	100.0%	Residential school	2	0.4%
			Correctional institution	0	0.0%
Interventions			Other educational setting	56	10.3%
Curriculum	48	8.8%	Indeterminate	20	3.7%
Grouping Structures/Tracking	20	3.7%	Total	544	100.0%
Pedagogical/Instructional	163	30.0%			
Personal/Affective	11	2.0%			
Professional Development	2	0.4%			
Technology	101	18.6%			
Tutoring Strategies	30	5.5%			
School Restructuring	46	8.5%			
Other	101	18.6%			
Indeterminate	22	4.0%			
Total	544	100.0%			

that, due to the possibility of multiple outcomes being reported in a study, the total number of outcomes identified exceeds the number of studies coded. Approximately one of every four studies had at least two outcomes identified.

Analysis of the educational settings for the studies coded using the keyword coding document showed that secondary schools, grades 6-12, accounted for 402 studies, or 73.9% of the studies. High schools, typically grades 9-12, accounted for 223 studies, or 41.0% of the total. Middle schools, grades 6-8, accounted for the educational setting for 179 studies, or 32.9%. Community or junior colleges and four-year colleges accounted for 63 studies, or 11.6% of the total. The groupings of "other" or "indeterminate" were assigned to 76 studies, or 14.0% of the total.

Section C of the keyword coding document was used to identify and categorize the characteristics of the study populations involved in the 544 qualified studies. Table 11 presents a summary of these characteristics. The characteristics included in this coding are the math content areas involved, the

Table 11

Keyword coding results, Sections C & D: Description of study population

Characteristic	N	%	Characteristic	N	%
Math Content Area			Is Gender Specified?		
Number Sense, Operations	158	29.8%	Yes	139	27.3%
Measurement	2	0.4%	No	371	72.7%
Geometry and Spatial Sense	42	7.9%	Total	510	100.0%
Data Analysis, Stat, Prob	9	1.7%			
Algebra and Functions	123	23.2%	Is Race/Ethnicity Specified?		
Pre-Calculus	15	2.8%	Yes	64	12.4%
Calculus	17	3.2%	No	451	87.6%
Other	164	30.9%	Total	515	100.0%
Total	530	100.0%			
			Is Other At-Risk Group Specified?		
Grade of Learners			Yes	95	18.6%
Single grade or range	460	88.1%	No	417	81.4%
Unknown	62	11.9%	Total	512	100.0%
Total	522	100.0%			
			Section D: Other		
			Does Report Meet Inclusionary Criteria?		
			Yes	140	25.7%
			No	405	74.3%
			Total	545	100.0%

identification of students' grade level, and an indication as to whether the students' gender, race or ethnicity, or other at-risk factors are identified in the study. This coding was the first instance in which the studies were reviewed for at-risk factor identification.

As can be seen in Table 11, the math content areas labeled "other", "number sense and operations", and "algebra and functions" were the top three math content areas identified with 30.9%, 29.8% and 23.2% of the total mentions, respectively. "Other" would be indicated where the study's author did not identify the content area or where the coder could not identify the area.

Specific information on the students' grade level was indicated in 460 of the studies, or 88.1%. Sixty-two of the studies (11.9%) did not identify specific grade levels but may have indicated the levels by terms such as high school, secondary school or postsecondary level.

The identification of at-risk factors made using the keyword coding process shows that gender identification, with 139 mentions, or 27.3% of the total number of studies, had the largest number of mentions. The identification of other at-risk factors, such as low-academic achievement, was the second largest number of mentions with 95 studies, or 18.6%. Race or ethnicity was identified as an at-risk factor in 64 studies, or 12.4% of the total. As can be seen by the counts, multiple at-risk factors could have been identified in each study.

Overall, keyword coding resulted in elimination of 407 studies, or 74.8% of the 548 total studies. The remaining 137 studies represent only 3.6% of the

3,814 citations which made up the initial citation database at the beginning of the screening process.

Section D of the keyword coding document was used by the reviewers to indicate if the study appeared to meet the at-risk factor inclusionary criteria established for inclusion in the final groups of studies. If the reviewer found information in the study that indicated the inclusion of at least one at-risk factor, such as student gender, race or ethnicity, or other at-risk factor, the study was considered to fully meet the inclusionary criteria. This information is also included in Table 11.

Data Extraction Coding

The third and final step in the screening process was the full data extraction of relevant data from each of the 137 studies meeting the keyword coding screening process. To accomplish this task, a data extraction document was developed and used. The primary intent of the data extraction coding was to review the quality of the interventions and outcomes and the internal and external validity and reliability of the studies.

At the conclusion of the data extraction process, 95 studies were identified as viable studies for inclusion in the meta-analysis. Of these 95 studies, five studies were found to include two totally independent studies. Including these five studies in the listing raised the final, included study count to 100 studies.

A total of 42 studies were eliminated from the review during the data extraction coding. Table 12 provides analysis information. Based on the coding, the two most common reasons for removal of a study from consideration was the lack of sufficient data for calculating an effect size and inadequate construct validity of an outcome measure: specifically the usage of class letter grades as an assessment measure for student achievement. Other less frequently encountered reasons included inadequate specification of the intervention, lack of clarity on whether the subjects sampled in the study were actually at-risk students, and a host of design inadequacies such as single-subject studies or the lack of any assurances of comparability of groups in non-randomized group design studies. Note that multiple reasons for exclusion from the review at this point were possible and are reflected in the total count.

Characteristics of the Final Studies

The systematic mapping of the final group of studies was done in a manner similar to the keyword coding process wherein characteristics of the studies were detailed by topic, following the structure of the coding document. These topic details included study source and publication characteristics, research design, participant characteristics, data collection and analysis, intervention characteristics, and outcomes.

Additional study source characteristics, including citation source, publication status and linked status, are reported in Table 12. It can be seen that the electronic database accounted for 96% of the study sources. Additionally, the majority of the studies were unpublished and only the 5 studies which included two independent studies were linked.

The second section of the data extraction coding procedure addressed research design characteristics of the studies. These included the identification of the study problem and research questions, research design type, the control or comparison group characteristics, and other quality-related analyses. Table 13 presents a summary of the information.

Table 12

Data extraction coding results: General study source characteristics

Characteristic	N	%	Characteristic	N	%
Reason for Exclusion			Source of Report		
Not applicable	95	65.1%	Citation	3	3.0%
Sample not appropriate	1	0.7%	Contact	0	0.0%
Inadequate intervention validity	12	8.2%	Hand Search	1	1.0%
Inadequate outcome validity	16	11.0%	Unknown	0	0.0%
Inadequate comparison information	4	2.7%	Electronic database	96	96.0%
Inadequate statistical information	16	11.0%	Total	100	100.0%
Single subject design	2	1.4%			
Qualitative study	0	0.0%	Linked Reports		
Total	146	100.0%	Not linked	95	95.0%
			Linked	5	5.0%
Publication Status			Don't know	0	0.0%
Published	37	37.0%	Total	100	100.0%
In press	0	0.0%			
Unpublished	63	63.0%			
Total	100	100.0%			

The explicit statement of a problem and the identification of the pertinent research questions are expected of research studies. As can be seen in Table 13, less than half of the studies explicitly stated the research problem while 74% explicitly identified research questions.

The majority of the research designs employed in these empirical studies were between groups (61.0% of the studies), while mixed designs accounted for 35.0% of the design. A few studies (4 total) were within subjects designs. This is reflective of the inclusionary criteria used in the review. In addition, as most of the studies were multi-groups comparisons, only 5% of the studies included single group, pretest/posttest correlations. It was determined by the reviewers that the studies were adequately informed and utilized acceptable quality research approaches. All of the studies were judged to contain adequate information for study replication.

Table 13

Data extraction coding results, Section B: Research design characteristics

Characteristic	N	%	Characteristic	N	%
Problem Statement Stated			Single Group, Pretest/Posttest Correlation		
Yes, explicitly	41	41.0%	Yes	0	0.0%
Not clear or not relevant	29	29.0%	No	5	5.0%
Not identified	30	30.0%	Not applicable	95	95.0%
Total	100	100.0%	Not state or unclear	0	0.0%
			Total	100	100.0%
Research Questions/Hypotheses Identified					
Yes, explicitly	74	74.0%	Study sufficiently Informed?		
Not clear or not relevant	4	4.0%	Yes	99	99.0%
Not identified	22	22.0%	No	1	1.0%
Total	100	100.0%	Total	100	100.0%
Research Design Type			Research Approach Quality Acceptable?		
Between groups	61	61.0%	Yes	100	100.0%
Within subjects	4	4.0%	No	0	0.0%
Mixed design	35	35.0%	Total	100	100.0%
Multiple regression design	0	3.8%			
Cannot be determined	0	0.0%	Enough Information to Replicate Study?		
Total	100	100.0%	Yes	100	100.0%
			No	0	0.0%
Comparison Group Characteristics			Total	100	100.0%
No treatment control	65	65.0%			
Alternative treatment	31	31.0%			
Wait list control	0	0.0%			
Other	0	0.0%			
Not stated or unclear	0	0.0%			
Not applicable	4	4.0%			
Total	100	100.0%			

The third section of the data extraction coding dealt with study participant characteristics. These characteristics include information of the participants' grade level in school, the urbanicity of the school, and selected personal demographics. In this coding section, the identification of specific at-risk factors was done, along with a measure on attrition rate. Table 14 summarizes the data from this coding.

It is noted that the majority of the participants in the studies were secondary students. Those identified as being in middle school, which would include the grades of 6 through 8, accounted for 46% of the studies while high school students made up 47% of the participants. A disappointingly small percentage of studies were college-aged participants, making up only 4% of the final 100 studies. Three studies were judged as lacking clarity on learner grade.

Identifying the participants by personal demographics leads to the realization that the majority of the participants were primarily male, secondary students from families of unknown socioeconomic status and unknown racial/ethnographic backgrounds. Though 26% of the studies identified the majority of participants as being White, 52% of the studies did not indicate the race/ethnicity of the students involved. Likewise, 68% of the studies did not identify the socioeconomic status of the participants.

The specific at-risk factors identified in these studies show that gender, with 66 of 133 mentions (49.6%), was measured as a key demographic of the study population. Twenty-seven studies (20.3%) identified low-academic achievement as an at-risk factor. Twenty studies identified a high-minority percentage as the at-risk factor. Other at-risk factors were identified in the studies. As can be seen the total number of mentions of at-risk factors in Table 14, some studies identified specific, multiple at-risk factors while a few studies stated there were multiple at-risk factors. As we know, it is not unusual to see multiple at-risk factors involved with most student populations.

Table 14

Data extraction coding results, Section C: Study participants characteristics

Characteristic	N	%	Characteristic	N	%
Grade of Learners			Specific At-Risk Factor Identified		
Middle school (grades 6-8)	46	46.0%	Gender	66	49.6%
High school (grades 9-12)	47	47.0%	Low-academic achievement	27	20.3%
College (grades 13-16)	4	4.0%	Low-SES	8	6.0%
Not stated or unclear	3	3.0%	High-minority percentage	20	15.0%
Total	100	100.0%	Non-English speaker (ELL)	2	1.5%
			Special need designation	3	2.3%
Gender of Participants			Multiple at-risk factors	4	3.0%
0 - 25% Male	0	0.0%	Other	0	0.0%
25 - 50% Male	40	40.0%	Not stated or unclear	3	2.3%
51 - 75% Male	28	28.0%	Total	133	100.0%
76 - 100% Male	1	1.0%	Attrition Rate		
Not stated or unclear	31	31.0%	0 - 10%	8	8.0%
Total	100	100.0%	11 - 20%	5	5.0%
			21 - 30%	4	4.0%
Participants Socioeconomic Status of Majority			Greater than 30%	0	0.0%
Mostly low class	17	17.0%	Not stated or unclear	83	83.0%
Mostly middle class	12	12.0%	Total	100	100.0%
Mostly upper class	3	3.0%			
Not stated or unclear	68	68.0%	Urbanicity of Study		
Total	100	100.0%	Primarily rural	14	14.0%
			Primarily suburban	20	20.0%
Participants Race/Ethnicity of Majority			Primarily urban	20	20.0%
White	26	26.0%	Mixed	6	6.0%
African American/Black	5	5.0%	Not stated or unclear	40	40.0%
Hispanic/Latino	14	14.0%	Total	100	100.0%
Mixed (33% each)	3	3.0%			
Not stated or unclear	52	52.0%			
Total	100	100.0%			

The level of attrition among the studies was analyzed. Unfortunately, the majority of the studies, 83 or 83% of the total, did not indicate the attrition rates. Of the 17 studies that did indicate attrition rates, most were considered acceptable with less than 10% attrition, while nine studies did have attrition rates ranging from 11 – 30%.

Urbanicity of the schools participating in the studies is of interest as a demographic measure due to the varied challenges and issues faced by each type of school location. As is indicated, 60% of the studies did identify the school setting. Of the total, schools identified as primarily suburban or urban accounted for 20% each. Rural school accounted for 14% of the school settings.

The methods and handling of data collection and analysis in the final set of studies was analyzed using Section D of the data extraction document. Included in this analysis were the identification of student achievement assessment tools and the evaluation of the adequacy of reliability and validity measures of the assessment tools. In addition, information on interrater reliability, used in some studies, and the identification of the statistical analysis tests used in the study was collected. Table 15 provides a summary of this information. Extensive usage was made of the dichotomous approach described in Chapter 3 during this coding.

The method of assessment use to measure student achievement outcomes was identified. Commercially-developed assessment tools were identified as being used by the largest number of study authors, with 44% of the total number of studies. These assessment tools include the Iowa Test of Basic Skills (ITBS) (Riverside Publishing Company), and the ACT (ACT, Inc.), among others. As these assessment tools are sold competitively, most of the publishers of these types of assessments have done extensive validity and reliability testing of the tools. In a similar manner, state standards assessments, such as the Colorado Student Assessment Program (CSAP), were identified as the assessment measure in 26% of the studies. Like the commercially-developed assessment tools, most state assessment programs have published validity and reliability information.

Teacher/researcher-developed assessment tools accounted for 30% of the assessment tools used in the studies. To be included in the final group of studies in this review, adequate validity and reliability data must have been presented to justify the inclusion of studies employing these assessment tools. As was discussed above, the lack of adequate validity and reliability information on the outcome assessment measure was the second largest factor accounting for study removal from the review.

The levels of acceptability of reliability and validity of outcome measures are reported in Table 15, along with an indication of the adequacy of the authors' attempts to operationalize the outcome measure. Though a few of the studies did not meet an acceptable measure as indicated, upon reaching consensus on studies, the reviewers felt sufficient evidence was presented to consider the measure acceptable. In addition, the use of interrater reliability measures was made in only 5 of the final studies. In these 5 studies, adequate information was provided to deem the measure acceptable by the reviewers.

Table 15

Data extraction coding results, Section D: Data collection and analysis

Characteristic	N	%	Characteristic	N	%
Method of Assessment of Outcome Measure			Study Interrater Reliability Acceptable?		
State standards test	26	26.0%	Yes	5	5.0%
Commercially-developed	44	44.0%	No	0	0.0%
Teacher/researcher-made	30	30.0%	Not applicable	95	95.0%
Questionnaire or survey	0	0.0%	Total	100	100.0%
Observation	0	0.0%			
Multiple instruments	0	0.0%	Statistics Used to Analyze Data Stated?		
Other	0	0.0%	Stated	100	100.0%
Not stated or unclear	0	0.0%	Not stated or unclear	0	0.0%
Total	100	100.0%	Total	100	100.0%
Reliability of Outcome Measure Acceptable?			Validity of Data Analysis Methods Described?		
Yes	91	91.0%	Yes	65	65.0%
No	9	9.0%	No	35	35.0%
Total	100	100.0%	Total	100	100.0%
Validity of Outcome Measure Acceptable?			Quality and Execution Acceptable?		
Yes	93	93.0%	Yes	97	97.0%
No	7	7.0%	No	3	3.0%
Total	100	100.0%	Total	100	100.0%
Operationalized measure					
Yes	96	96.0%			
No	4	4.0%			
Total	100	100.0%			

In reviewing the studies for statistical analysis utilized, it was noted that the reviewers were able to identify the statistical analysis used in all the studies. This indicates that the authors of the studies did an admirable job of informing the readership. Though only 65% of the authors provided information on why they choose to employ selected analysis methods, the vast majority of the authors (97%) did an adequate job of execution of the analysis.

Section E of the data extraction coding document was used to describe the characteristics of the intervention, including who delivered and who received the intervention, and specific information regarding the timing and duration of treatment. Table 16 presents a summary of this data.

Overall, the teacher was identified as being the individual delivering the intervention in the majority of the studies (76%). The researcher delivered the intervention in only 7% of the studies while another individual was involved in 11% of the studies. It was unclear who delivered the intervention in 6% of the studies.

As might be expected, the individual student was the recipient of the treatment in the vast majority of the studies (91%). Without a doubt, this large percentage was due to the inclusionary criteria used in this review. An entire school was the targeted recipient in 7% of the studies while multiple classrooms or multiple schools were targeted in only one study each.

The fidelity of implementation of an intervention is critical. As part of the data extraction, reviewers noted the adequacy of intervention definition in all the study was met. In addition, a full description of the intervention was presented in 97% of the studies reviewed. The quality of intervention delivery was also deemed adequate in all of the studies, as was the consistency and appropriateness of the educational setting. However, no direct measure of fidelity of implementation was noted in any of the studies.

Table 16

Data extraction coding results, Section E: Intervention description

Characteristic	N	%
Who delivered intervention?		
Researcher	7	7.0%
Teacher	76	76.0%
Other	11	11.0%
Not stated or unclear	6	6.0%
Total	100	100.0%
Recipient of intervention		
Individual student	91	91.0%
Classroom of students	0	0.0%
Multiple classrooms	1	1.0%
Entire school	7	7.0%
Multiple schools	1	1.0%
Teacher	0	0.0%
Other	0	0.0%
Not stated or unclear	0	0.0%
Total	100	100.0%
Duration of intervention		
Less than one day	1	1.0%
One day to 1 week	6	6.0%
Eight days to 1 month	13	13.0%
More than 1 month to 3 months	13	13.0%
More than 3 months to 6 months	13	13.0%
More than 6 months to 1 year	22	22.0%
More than 1 year to 2 years	10	10.0%
More than 2 years	7	7.0%
Other	1	1.0%
Not stated or unclear	14	14.0%
Total	100	100.0%

Characteristic	N	%
Delivered on a routine basis?		
Stated	41	41.0%
Not state or unclear	59	59.0%
Total	100	100.0%
Estimate total time of intervention		
Estimated at	52	52.0%
Cannot estimate	48	48.0%
Total	100	100.0%
Time of intervention		
0 - 50 hours	26	50.0%
51-100 hours	4	7.7%
101-500 hours	14	26.9%
Over 500 hours	6	11.5%
Other	2	3.8%
Total	52	100.0%
Year(s) of intervention		
Before 1990	2	4.0%
1990 - 1994	17	34.0%
1995 - 1999	26	52.0%
2000 - 2004	5	10.0%
2005	0	0.0%
Total	50	100.0%

The duration of an intervention, comparable to the dosage measure of a treatment in medical research, is important in education. A measure of duration of the mathematics intervention was included in this section of the data extraction. This measure was done through an attempt to identify the amount of time the intervention was administered combined with an indication of the delivery of the intervention on a routine basis. These two measures then led to an estimate on the total number of hours devoted to the intervention or treatment over a school year.

Recent information from the Education Sector places the amount of time devoted to learning in U.S. public schools at about 800 hours per year (Silva, 2007). As can be seen in Table 16, an estimate of 0 – 50 hours was made for the duration of the mathematics intervention in half of the studies in which as estimate of duration could be made. In actually, 21 of these 26 studies were estimated between 0 and 25 hours total per year. Twenty studies included estimated between 101-200 hours. These studies were most likely ones including full-school interventions such as the implementation of block scheduling or double-dose mathematics courses.

It was also noted that half of the studies indicated the year or years of intervention implementation. Most of the studies in which this information was identified showed the 5-year interval between 1995 and 1999 as being most mentioned.

The final section coded under the data extraction document was outcome characteristics. In this coding, the studies were reviewed for the inclusion of outcome definitions, agreement on unit of allocation and data analysis, determination of how groups were assigned to treatment and control conditions, and an indication of the awareness of the participants and data collectors on group membership. Additionally, the effect size estimate was calculated for each of the remaining studies. Table 17 presents a summary of the outcome characteristics from the data extraction coding.

Table 17

Data extraction coding results, Section F: Outcomes

Characteristic	N	%	Characteristic	N	%
Outcome defined			Awareness during intervention?		
Stated	100	100.0%	Yes	90	90.0%
Not state or unclear	0	0.0%	No	2	2.0%
Total	100	100.0%	Not applicable	8	8.0%
			Total	100	100.0%
Unit of allocation					
Individuals	86	86.0%	Data collectors unaware of grouping?		
Classrooms	4	4.0%	Yes	4	4.0%
A school	5	5.0%	No	96	96.0%
Multiple schools	0	0.0%	Total	100	100.0%
School districts	0	0.0%			
Other	0	0.0%	Agreement on data analysis & allocation units		
Not applicable	5	5.0%	Yes	100	100.0%
Total	100	100.0%	No	0	0.0%
			Total	100	100.0%
For multi-group, how groups assigned					
Randomized	16	16.0%	Shortcoming on numerical reporting?		
Cluster randomized	11	11.0%	Yes	12	12.0%
Randomized matched pairs	3	3.0%	No	88	88.0%
Matched pairs (ex post facto)	4	4.0%	Total	100	100.0%
Convenience or other	54	54.0%			
Not stated or unclear	2	2.0%	Quality appropriate		
Not applicable	10	10.0%	Yes	99	99.0%
Total	100	100.0%	No	1	1.0%
			Total	100	100.0%
Group membership unknown?					
Yes	2	2.0%	Effect size calculation possible?		
No	91	91.0%	Yes, as reported	5	5.0%
Not applicable	7	7.0%	Yes	80	80.0%
Total	100	100.0%	Not clear	15	15.0%
			Total	100	100.0%

As seen in the data summary, all 100 of the studies included definitions of the outcome variables. It is also seen that the majority of the units of allocation were students, accounting for 86% of the studies, while classrooms and schools were also identified. Agreement between the units of data analysis and units of allocation were noted in all of the studies.

In the multi-group studies, the method used to assign group membership was coded. For the 100 studies included, more than half (54%) were assigned by convenience or some other non-equivalent technique. This is recognized as a common concern when dealing with secondary classrooms in public schools. It was also noted that 16% of the studies used randomized assignment and another 11% used a cluster randomized approach of assignment.

The awareness of group membership, both among participants and data collectors, is another area of concern in educational research. Seldom do educational research studies include double-blind experiments. As shown in Table 17, most of the participants (91%) in the studies were aware of their group membership (i.e., treatment versus control) and most were of aware of their involvement during the intervention (90%). Likewise, the majority of the data collectors (96%) were aware of group membership while the collection of data was being done.

Overall, some shortcomings with the quality of the studies were identified in 12% of the studies, such as participant awareness of group membership. However, the reviewers felt that sufficient quality was evident to warrant inclusion of the final 100 studies in the review. The overall appropriateness of the quality of the study was recognized in 99% of the studies. The one questionable study was included when the reviewers agreed that the usage of a percentage score as a classroom grade, taken as the measure of assessment, was to be considered a valid measure of assessment. Other studies in which a letter grade was used as the outcome measure were considered to have inadequate validity of outcome measures.

The last part of Table 17 shows that 85% of the studies were considered to have sufficient information to allow an effect size estimate calculation during the data extraction coding. The 15 studies that were questionable did yield sufficient data for effect size calculation when the studies were reviewed in detail for adequacy of data.

The final set of studies that made up this review of mathematics interventions with at-risk secondary students numbered 100 studies. Of these 100 studies, five studies included two completely independent study samples. Consequently, 95 studies of the original 3,814 citations collected in the literature search were considered appropriate for the review. These 95 studies represent 2.5% of the total number of citations.

Effect Size Comparisons

A somewhat unique comparison of effect size estimates between the overall study population and the specific at-risk factor is made in this review. Typically, effect size comparisons provided in systematic reviews are limited to the overall or main effect. The approach taken in this review is to include a comparison of the effect size estimate for the identified at-risk factor, such as gender or low-academic achievement.

This approach is actually a form of moderator analysis in that the comparison of effect size for the main effect between the treatment and control groups is compared to the effect size estimate of the disadvantaged component of the specific at-risk factor. Of primary interest is the effect size comparison

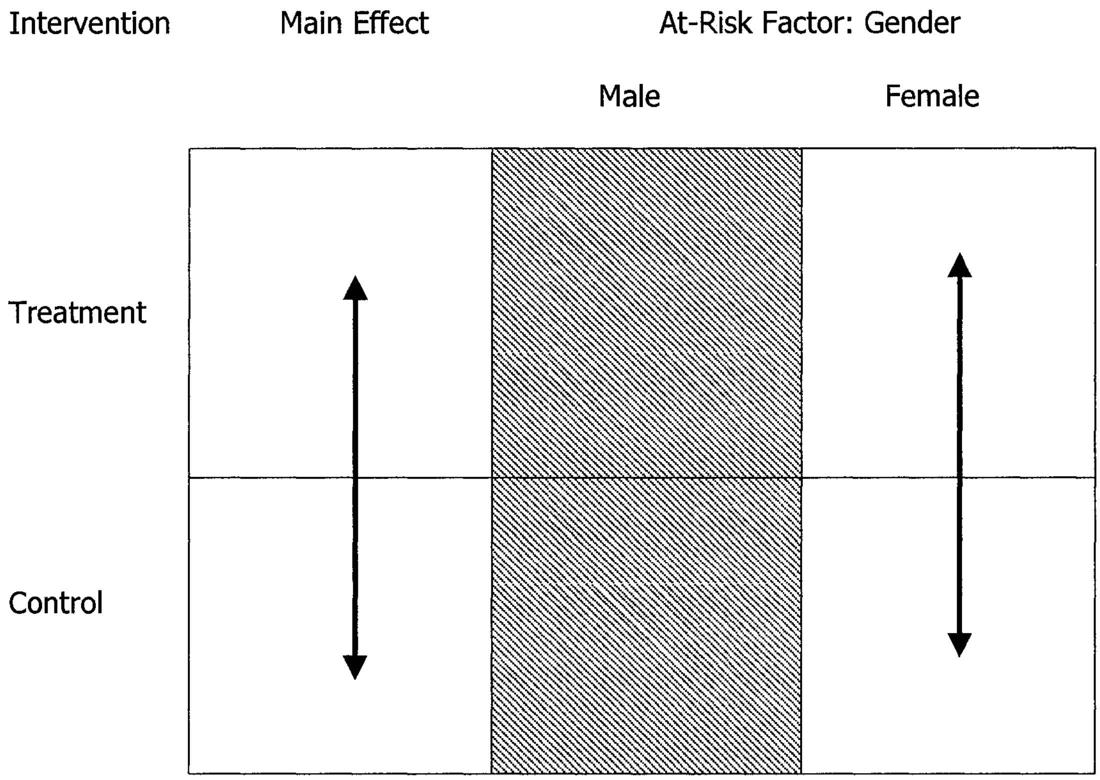

Figure 5. Conceptual map of comparison grouping of effect size estimates for the main effect and for the specific at-risk factor (gender in this example) where the disadvantaged factor is of interest.

between the treatment and control groups of the specific, disadvantaged subpopulation. Figure 5 provides a conceptual map of this unique approach to the presentation of the effect size comparisons.

It can be seen in Figure 5 that the effect size estimate for the main effect of treatment versus control provides a measure for the overall population in the study. As our interest is in the at-risk population, a second effect size is calculated based on the differences between the treatment and control groups of the identified, disadvantaged-member of the at-risk factor pair: in the example shown in Figure 5, the female student subpopulation.

Effective Interventions

At this point, the final 100 studies were sorted into groups by intervention, based on the coding of the intervention type. These intervention groups were: curricular, grouping/tracking, pedagogical/instructional, school restructuring, technology-based, and all other non-pedagogical interventions. It was apparent that the pedagogical/instruction intervention group could be divided further to include the subgroups of cognitive and metacognitive strategies, cooperative learning, peer tutoring, and other interventions. Table 18 provides a breakdown of the studies by intervention group with study counts, subgroup identification and counts, and at-risks factors identified and counts.

Pedagogical or instructional interventions was the largest group of interventions identified, accounting for 43 of the 100 studies. Technology-based interventions made up the second largest group with 21 studies identified.

School restructuring and grouping/tracking interventions had the next largest count totals with 12 and 11 studies, respectively. Eight of the studies were classified as curriculum interventions while "other, non-pedagogy" interventions accounted for five studies.

Table 18

Final intervention groupings

Intervention	Study Counts	Subgroup (Counts of Main Effects)	At-Risk Factors (Count)
Curriculum	8	Varied (8)	Gender (6), Minority (1), Special needs (1)
Grouping/Tracking	11	Ability (7), Single sex (4)	Gender (6), Low acad (3), Other (2)
Pedagogy/Instructional	43	--	--
Cognitive & metacognitive strategies	--	14	Gender (7), Low acad (6), Other (1)
Cooperative learning	--	10	Gender (6), Low acad (4)
Peer tutoring	--	5	Gender (4), Minority (1)
Other	--	14	Gender (10), Low acad (2), Other (2)
School Restructuring	12	Block scheduling (10), Other (2)	Gender (7), Lo SES (1), Low acad achieve (2), Minority (1), Other (1)
Technology-based	21	CAI (8)+ ILS (3) =(11), Calc (5), Other (5)	Gender (18), Low acad (3), Minority (1)
Other, Non-Pedagogy	5	Person/Affect (3), Other (2)	Gender (2), Special needs (2), Other (1)
Total	100		

The curriculum interventions included in the review were eight studies which varied widely in constructs. No two studies addressed the same curriculum intervention. Consequently, a meta-analysis of these studies would not be considered appropriate as each represents a different construct. This lack of adequate empirical research studies has also been an issue for the What Work Clearinghouse on their review of middle school and elementary school mathematics curriculum interventions (2004a; 2006a).

The grouping/tracking interventions accounted for eleven studies. This group could be divided further into two subgroups: grouping and tracking by ability and by gender. This group shows that three at-risk factors were identified: gender, low-academic achievement, and other at-risk factors. Unfortunately, at-risk factor data for calculating the effect size estimate was available for only three of the studies.

The largest group of interventions, pedagogical or instructional interventions, was further divided into four subgroups. Those studies identified as cognitive and metacognitive strategies accounted for 14 of these studies. Other pedagogical studies also accounted for 14 studies. This subgroup included a variety of interventions, including the use of manipulatives, double dose mathematics exposure, and anchored instruction, among others.

School restructuring interventions, which were considered to be interventions involved at the entire school, included multiple studies on block scheduling and single studies in accelerated mathematics and year-round schools. A number of at-risk factors were identified.

Technology-based interventions were the second largest grouping of interventions. Subgrouping within this intervention category was possible with eight studies involving computer-aided instruction (CAI) and three studies addressing integrated learning systems (ILS). Five studies addressed the usage of calculators, specifically graphing calculators, and mathematics achievement. Five other technology-based studies were also included in this group.

Overall, five studies could not be grouped with the other intervention groups identified. Three studies were categorized as personal or affective studies with achievement outcomes. The other two studies addressed the oral administration of standardized achievement tests and the extension of testing times for special needs populations.

A listing of each of the intervention groups and the studies included in each group is presented in Appendix F. The complete listing of the 95 study citations is located in Appendix G.

As can be seen with each of the intervention groups presented in Appendix F, each included study is identified by author(s), publication date, and intervention type. In addition, the main effect or overall effect size estimate is listed, along with the advantaged and/or disadvantaged-member of the at-risk factor pair and the effect size estimates for each of those factors (listed as primary ES and/or secondary ES).

No meta-analysis of the effect sizes for each of the intervention types is included in this review nor has an average effect size been calculated. Work is underway by the author to complete the meta-analysis for each of the intervention types identified in this systematic review. The author and the CLT-

West team are currently writing articles on each of the interventions with the intent to submit articles for publication in the fall of 2007.

CHAPTER 5

SUMMARY, CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS

Summary

A systematic review of 3,814 published and unpublished citations with abstracts identified in a literature search on interventions in secondary and early post-secondary mathematics with at-risk students was conducted. The objective of the systematic review was to identify those interventions which are most successful at improving the mathematics achievement of at-risk students while reducing the mathematics achievement gap. A multi-phased coding process resulted in 95 empirical studies meeting the inclusionary criteria. The results were summarizes through a descriptive mapping of the entire literature base and two systematic mappings of an intermediary and final set of included studies.

Conclusions

Two research questions were posed to focus this systematic review. These questions were:

1. What are the characteristics of curricular interventions that are most effective in reducing the mathematics achievement gap for at-risk secondary students?
2. What are the characteristics of instructional interventions that are most effective in reducing the mathematics achievement gap for at-risk secondary students?

In response to the first question, it is evident from the small number of curricular intervention studies that made it through the screening process (a total of eight studies) and the wide range of specific curricular efforts involved, that no curriculum interventions could be identified as being effective in reducing the mathematics achievement gap for at-risk secondary students. Each of the eight studies is considered sufficiently different that none of them can be grouped together in a meta-analysis as they represent completely different curricular approaches or constructs. It is also noted that, as explained prior, the sign of an effect size is important. Of the eight studies, half have positive effect size estimates while the other four studies have negative effect size estimates. Only three of the studies present at-risk factors with sufficient data to allow the effect size calculation by at-risk factors, though the three at-risk factors are all gender.

This finding is consistent with similar systematic review efforts being conducted by the What Works Clearinghouse (2004a; 2006a). Two factors may be involved here. The first is the lack of a sufficient number of empirical studies on each individual curriculum intervention, as noted by the What Works Clearinghouse. The second factor may be related to both the lack of empirical studies done by reviewers external to the publishing companies producing the reviews and the inadequacy of the studies on the curriculum itself.

The second research question addressed the characteristics of instructional interventions that are most effective in reducing the mathematics achievement gap for at-risk secondary students. This intervention area did produce a sufficient number of qualified studies in multiple pedagogical areas for analysis. A total of 43 studies made it through the screening process. A subgrouping of these studies resulted in three potentially effective interventions: (a) cognitive and metacognitive strategies; (b) cooperative learning; and (c) peer tutoring. Each of these subgroups will be discussed further.

Beyond the two research questions, the results of this systematic review suggest that two other interventions may be effective in mathematics achievement with secondary and early-postsecondary, at-risk students: (a) technology-based applications, (b) selected school restructuring interventions.

Three general conclusions can be reached in evaluating the information presented in this systematic review. These conclusions are: (1) The base of adequate empirical studies available in mathematics achievement with secondary and early-postsecondary students is surprisingly small and limited; (2) The overall quality of research studies conducted in the area of mathematics achievement with secondary and early-postsecondary students is questionable due to inadequate statistical and analytical reporting and inadequate research design considerations; and (3) Viable interventions for the improvement of mathematics achievement with at-risk secondary and early-postsecondary students appears to be limited to selected instructional, technology-based, and selected school-restructuring interventions. The lack of viable curricular interventions is noted.

It is recognized that not all studies conducted in the area of secondary and post-secondary mathematics achievement are intended to be empirical in nature and subjected to the rigorous scrutiny of a systematic review. Many of the studies eliminated from this review were descriptive in nature and intended to call attention to specific observations in the study of factors affecting mathematics achievement. Though they may have been eliminated from this review, their value to the knowledge base is noted.

Though mathematics education is one of the most frequently researched secondary-level, academic core areas (De Corte, 2004; Vitale & Romance, 2006), it is evident that the quality of empirical studies available in mathematics achievement with secondary and postsecondary students is limited. Of the 3,814 citations identified in the initial literature search for potential research studies, only 95 of the studies passed through the multi-phased screening process.

As described in Chapter 4, if we consider that approximately 70% of the original 3,814 citations were eliminated from this systematic review due to inappropriate study population for this review, we were left with 1,087 citations meeting the target population requirement (secondary and early-postsecondary students, grades 6 through 14). Of the remaining 1,087 citations, 495 more citations were removed for not meeting intervention, outcome, or quantitative study requirements. The remaining 592 citations represent only 15.5% of the original 3,814 citations. Further analysis of this group of citations resulted in only 137 meeting specific qualification requirements. These 137 citations represent only 3.6% of the total number of citations. This group was further reduced when the full data extraction coding was applied to the group of citations and 42 of the studies were eliminated due to inadequate statistical information, outcome validity, and intervention validity. This reduction of the final study count by an additional 30.7% brought the total number of viable studies in this review to 95 studies. These 95 studies represent 2.5% of the initial 3,814 citations identified for this review. Effectively, these 95 studies are the full collection of studies to meet all the inclusionary criteria.

Quality issues among the studies, specifically those involving research design issues and statistical handling and reporting problems, were noted. One area of concern expressed by the research team was the poor level of quality seen in the number of dissertations reviewed. These included dissertations that were apparently accepted in which attrition rates of over 50% were experienced and one in which pretest/posttest assessment scores on two different sample populations were used to measure intervention success.

Of primary interest in this review are the conclusions that can be drawn from those interventions identified as effective in general, and those identified as being effective for selected at-risk factors. Based on the analysis, three key intervention categories (a) technology-based applications, (b) pedagogical or instructional interventions, and (c) selected school restructuring interventions were identified as potentially effective in general. Further analysis, including the complete meta-analysis of these interventions, will be conducted in the future.

Discussion

A discussion of all the interventions identified during the systematic review follows. A listing of all the interventions by group, along with selected characteristics and an effect size estimate for each of the studies in each group, is presented in Appendix F.

Interventions

Six intervention types with studies and effect size estimates are listed in Appendix F as follows: (a) curriculum interventions; (b) grouping/tracking interventions; (c) pedagogical/instructional interventions; (d) school restructuring interventions; (e) technology-based interventions; and (f) other, non-pedagogical interventions. Using a general "rule-of-thumb" from the Campbell Collaboration, it is typically considered inappropriate to do a meta-analysis of fewer than ten studies on the same construct. Based on this, of the intervention types listed, only four interventions may have sufficient studies for a meta-analysis to be conducted. These are: (a) grouping/tracking interventions; (b) pedagogical/instruction interventions; (c) school restructuring interventions; and (e) technology-based interventions. However, two subgroups in the pedagogical/instructional intervention group may have sufficient numbers of studies to warrant a meta-analysis.

Curriculum Interventions

The curriculum interventions listed in Appendix F1 include eight studies addressing a wide range of specific interventions, from one focused on curriculum standards to seven completely different curriculum approaches, such as Saxon mathematics compared to a traditional mathematics curriculum. Each of these eight studies is considered sufficiently different that none of them can be grouped together in a meta-analysis as they represent completely different curricular approaches or constructs. It is also noted that, as explained prior, the sign of an effect size is important. In the column labeled "Overall ES", four of the studies have positive effect size estimates while the other four studies have negative effect size estimates. Only three of the studies present at-risk factors with sufficient data available to allow the effect size calculation by at-risk factors, though the three at-risk factors are all gender.

Grouping/Tracking Interventions

The grouping/tracking intervention results show eleven studies included in the listing in Appendix F2. Two different interventions made up this intervention group. Seven of the studies addressed grouping or tracking by ability group while four studies looked at grouping by gender: single-sex classrooms. Of these studies, four yielded positive "overall ES" estimates, 5 were negative estimates, and two were regression discontinuity measures, where the results given were in positive standard deviation measures for both studies. Studies of grouping or tracking by ability showed three negative effect size estimates, two positive estimates, and two positive regression discontinuity measures. The four grouping or tracking studies by gender showed two positive and two negative effect size estimates. These mixed results are consistent with other reviews done in the topic area (Betts & Shkolnik, 2000; Hanushek & Wobmann, 2006; Ireson & Hallam, 1999; Jimenez & Lockheed, 1989; Kulik & Kulik, 1982; LePore & Warren, 1997; Slavin, 1990).

Only four of the grouping/tracking intervention studies provided adequate data the calculation of effect size estimates for at-risk factors included. Three of these studies addressed gender as the at-risk factor, with two of the three effect size estimates being negative values for the disadvantaged pair of the at-risk factor: females. One study did listed females only and showed a strong, positive effect size estimate. The fourth study provided data on low-academic achievement and yielded a huge effect size estimate.

Pedagogical/Instructional Interventions

The pedagogical or instructional interventions provided four separate subgroups of interventions as shown in Appendix F3 through F6. The first subgroup was that grouped by cognitive and metacognitive strategies. Cognition is the mental process of learning and knowing. Metacognition is considered to be "one's knowledge concerning one's own cognitive processes" (Flavell, 1976). One of the pedagogical interventions identified as potentially effective in this review is the teaching of cognitive and metacognitive strategies to students. These strategies take two different forms. With the teaching of cognitive strategies in mathematics, the primary focus is on strategies for problem-solving, which are typically skill-based (Mayer, 1998). With instruction in metacognitive strategies, focus is primarily on the decision-making process to determine the appropriateness of a cognitive skill application to solve a problem.

As is evident in Appendix F3, fourteen studies were included in the final collection of qualified studies. Thirteen of these studies show positive overall effect size estimates while one is negative. Eleven of the studies provided sufficient information for calculation of the at-risk effect size estimate. Six of the studies addressed low-academic achievement as the at-risk factor, with all six showing a positive effect size estimate. Four studies provided effect size estimates for gender, with three of the estimates being positive and one negative for the female comparison of the at-risk factor pairing. One other at-risk factor, identified as "multiple at-risk" factors, showed a large difference between the at-risk factor pair.

Appendix F4 provides information on those studies qualified in the cooperative learning subgroup of the pedagogical or instructional intervention. Ten studies are listed, with eight of the studies being between group designs and two being within subject designs. Nine of the ten studies show positive effect size calculations. Six of the studies included gender as the at-risk factor, with four of the six studies providing sufficient information for both male and female effect size estimates. The effect size estimates are mixed with one of the four being a negative value and three of the four being smaller than the effect size estimate for the male. Three studies included low-academic achievement at-risk factor effect size estimates. Of the three, two studies showed that the population identified as low-academic achieving students had a negative effect size estimate compared to the overall effect size. Only one study showed that low-academic achievement students benefited more from the cooperative learning intervention.

The third pedagogical or instructional intervention subgroup was focused on peer tutoring and the summary of studies is shown in Appendix F5. Five studies were included in the final group of studies from the review in this intervention subgroup. This is probably an insignificant number of studies for the inclusion of a meta-analysis on the studies. Of interest however is the fact that all five studies provided positive effect size estimates, though only one sufficient data for calculation of an effect size estimate with an at-risk factor. Four of the studies identified gender as the at-risk factor while the fifth study identified a high-percentage of minority students.

The final subgroup of pedagogical or instructional interventions included all other interventions. Though fourteen of these studies were included, only a few of the interventions identified have more than one qualified study. Seven of the fourteen studies yielded positive overall effect size estimates while the other seven were negative. Ten of the studies identified gender as the at-risk factor while low-academic achievement was identified as the at-risk factor in two of the studies. Six of the studies provided sufficient data for the calculation of the effect size estimates for the at-risk factor. Five of these studies yielded effect size estimates for the female pair of the at-risk factor, with four of them being positive values. One low-academic achievement factor yielded a negative effect size estimate.

School Restructuring Interventions

School restructuring interventions were identified in twelve studies. The summary of these studies is located in Appendix F7. Of the twelve studies, ten were studies on block scheduling as a school restructuring intervention. The other two interventions were identified as year-round schools and the restructuring of the school to accommodate a focus on accelerated math.

Seven of the ten block scheduling studies yielded positive effect size estimates while three were negative. The other two interventions yielded positive effect size estimates. Eight of the twelve studies provided sufficient information for the calculation of an effect size estimate for the at-risk factor. Four gender at-risk factors were noted. Low-academic achievement, with two effect size estimates calculated, and one each for a high-percentage of minority students and for multiple at-risk factors was shown.

Technology-based Interventions

Twenty-one studies are included in the technology-based intervention summary shown in Appendix F8. Eight of the studies concerned computer-aided instruction (CAI) and three studies concerned integrated learning systems (ILS) approaches. Five studies focused on the use of calculators as an intervention while the remaining five studies were identified as either general technology-based interventions or single topic interventions, which included one study on the Internet and one on the usage of the Logo computer language.

Four of the eight studies concerning CAI interventions proved to yield negative effect size estimates. Two of the ILS studies showed negative effect sizes also. Four of the calculator studies yielded positive effect size estimates while the one comparison of the usage of a calculator versus the Maples software yielded a negative effect size estimate. For the remaining five studies all yielded positive effect size estimates.

Sixteen of the twenty-one studies in the technology-based intervention group provided sufficient data for the calculation of at-risk factor effect sizes. Twelve of the sixteen were the gender at-risk factor, with seven of the female at-risk factor effect size estimates proving to be positive and five were negative. Low-academic achievement yielded three effect size estimates, two of which were negative, while the at-risk factor of English as a second language (ESL) yielded one positive effect size estimate.

Other, Non-Pedagogical Interventions

Five studies were grouped into this intervention group. Three of these studies were personal or affective intervention studies while the other two were associated with interventions aimed at attempts to increase at-risk students' scores on standardized achievement tests. Four of the five studies yielded positive overall effect size estimates. Three of the five studies identified at-risk factors of gender and students with learning disabilities provided sufficient data for at-risk factor effect size calculations. Due to the wide differences in constructs included in this group, little can be done with these studies.

At-Risk Factors

Analyzing the results of this review from the perspective of those at-risk factors which yielded effect size estimates instead of from an intervention perspective, allows us to consider how specific at-risk populations might fare compared to the general population, as measured by the main effect size estimate. For example, a question such as "How does mathematics achievement with female secondary students compare to that of the general population in a cooperative learning intervention?", can be addressed. Tables 19 and 20 provide an analysis of this perspective.

Table 19 presents a summary of the at-risk factor effect size estimates by intervention. As can be seen, those eight curriculum intervention studies that passed through the screening process yielded three studies that sufficient data for the calculation of the effect size estimate for the gender at-risk factor, of which the female student population is considered the disadvantaged factor.

Table 20 presents the number of at-risk factor effect size estimates with values above and below the main effect size estimate for gender and low-academic achievement by intervention. From this table, we can see that, of the three studies providing an at-risk effect size estimate, one of the estimates was greater than the main effect size estimate while two estimates were less than the main effect size estimate.

Table 19

Number of at-risk factor effect size estimates by intervention

Intervention	Gender	Low-Acad	Hi-Minority	Low SES	Special Needs	Other	Total
Curriculum	3	0	0	0	0	0	3
Grouping/Tracking	4	1	0	0	0	0	5
Pedagogy/Instructional							
Cognitive & metacognitive	4	6	0	0	0	1	11
Cooperative learning	4	3	0	0	0	0	7
Peer tutoring	1	0	0	0	0	0	1
Other	5	1	0	0	0	0	6
School Restructuring	4	2	2	1	0	1	10
Technology-based	12	3	1	0	0	0	16
Other, Non-Pedagogy	2	0	0	0	1	0	3
Total	39	16	3	1	1	2	62

From the at-risk factor perspective, we can see from Table 20 that the female gender at-risk factor may benefit from the usage of technology-based interventions more than the male gender at-risk factor. Also, we see that the teaching of cognitive and metacognitive strategies may be more detrimental to the low-academic achievement at-risk population than to the general population.

As the comparisons made between the effect size estimates of the at-risk factor and the main effect in Table 20 was done to further demonstrate the relationship involved, caution should be used when interpreting the table and its intent. Though four studies show the female at-risk factor effect size estimates below the main effect size estimate for technology-based interventions that does not mean that those four studies were ineffective or detrimental for the disadvantaged population.

Table 20

Number of at-risk factor effect size estimates above and below the main effect size estimate for gender and low-academic achievement by intervention

	Gender(F) ES_F		Low-Acad ES_L		Total ES	
Intervention	> ES_M	< ES_M	> ES_M	< ES_M	> ES_M	< ES_M
Curriculum	1	2	0	0	1	2
Grouping/Tracking	2	2	1	0	3	2
Pedagogy/Instructional						
Cognitive & metacognitive	3	1	0	5	3	6
Cooperative learning	2	2	1	2	3	4
Peer tutoring	1	0	0	0	1	0
Other	4	1	0	1	4	2
School Restructuring	2	2	1	1	3	3
Technology-based	8	4	1	2	9	6
Other, Non-Pedagogy	1	1	0	0	1	1
Total	24	15	4	11	28	26

Recommendations

Four recommendations are advanced based on researcher observations and planned future extension of this research effort. Inherent in these recommendations is the understanding that the full meta-analysis of the 100 final studies qualified in this review is being completed by the CLT-West research team and the author over the next few months.

First of all, it is evident that, even though a sufficient amount of research has been done in this area, more high-quality empirical studies need to be done on the topic of mathematics interventions with at-risk populations. Of most interest here would be the development of a randomized control trial approach to investigate instructional interventions with the at-risk population including those of low-academic achievement and those with multiple at-risk factors.

Secondly, the importance of this type of study in the social sciences further illustrates the growing need for primary researchers to adequately document and communicate their results, including detailed information about the research populations studied. As stated prior, a systematic review is a secondary research methodology. Consequently, it is limited by the level of information provided in the original research. It is commonplace for practitioners of the systematic review methodology to focus on internal validity, including equivalence of groups and the control of extraneous factors, but to virtually ignore external validity.

Morgan, Gliner, and Harmon (2006) make the case that external validity, which addresses the generalizability of the results, in just as important as internal validity. However, in many systematic reviews, external validity measures and the resulting generalization of interventions by population and setting is limited. Of critical importance here is the need for more adequate reporting of subgroup data. As at-risk factors were a key part of this review, it was evident that the general absence of subgroup demographic and assessment data hampered this research effort. If primary researchers would better document their studies with subgroup information, even if only raw data are provided, the benefit of improved generalizability for at-risk students groups could be realized. Additionally, increased confidence in the generalizability of results would be possible with increased sample counts based on adequate reporting of subgroup data.

A third recommendation is that an update to the literature search and the systematic review be conducted to identify empirical studies that have been completed and released since the end of 2004, including those released late in 2004 but were not available when the original search was conducted. Based an average of about 40 potentially qualified studies released per year, it is anticipated that another 100 qualified studies could be added to the review with the result being about 20-25 more studies being fully qualified for inclusion in the meta-analysis.

The fourth and final recommendation is that the work done in this systematic review needs to be conducted in the area of science achievement. Though a more complex academic subject area with far fewer empirical studies available the research literature base, this area of study has tremendous importance for not only education but for the U.S. economy in general.

The final results of this systematic review will be a series of research articles describing the most effective interventions for student populations with specific at-risk factors. This work is currently underway. In addition, the results of this systematic review can be used to direct future research efforts by defining investigations into specific interventions with specific populations while informing the knowledge base.

REFERENCES

Abrami, P. C., Bernard, R. M., Cobb, C. B., Nye, C., Turner, H., & Wade, A. (2005, April 12, 2005). *Producing C2 systematic reviews to inform evidenced-based policy and practice in education.* Paper presented at the American Educational Research Association, Montreal, Canada.

Badian, N. A. (1983). Dyscalculia and nonverbal disorders of learning. In H. R. Myklebust (Ed.), *Progress in Learning Disabilities* (Vol. 5, pp. 235-264). New York: Grune & Stratton.

Baker, S., Gersten, R., & Lee, D.-S. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *The Elementary School Journal, 103*(1), 51-73.

Barley, Z., Lauer, P. A., Arens, S. A., Apthorp, H. A., Englert, K. S., & Snow, D. A., M. (2002). *Helping at-risk students meet standards: A synthesis of evidence-based classroom practices.* Aurora, CO: Mid-continent Research for Education and Learning.

Barton, P. E. (2002). *Raising achievement and reducing gaps: Reporting progress toward goals for academic achievement in mathematics.* Washington, DC: National Education Goals Panel.

Bennett, J., Lubben, F., Hogarth, S., Campbell, B., & Robinson, A. (2005). *A systematic review of the nature of small-group discussions aimed at improving students' understanding of evidence in science.* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Berends, M., Lucas, S. R., Sullivan, T., & Briggs, R. J. (2005). *Examining gaps in mathematics achievement among racial-ethnic groups 1972-1992.* Santa Monica, CA: RAND Corporation.

Betts, J. R., & Shkolnik, J. L. (2000). The effects of ability grouping on student achievement and resource allocation in secondary schools. *Economics of Education Review, 19*, 1-15.

Bol, L., & Berry III, R. Q. (2005). Secondary mathematics teachers' perceptions of the achievement gap. *The High School Journal* (April/May), 32-45.

Brian, T., Bay, M., Lopez-Reyna, N., & Donahue, M. (1991). Characteristics of students with learning disabilities: A summary of the extant data base and its implications for educational programs. In J. W. Lloyd, N. Nirbhay & A.

C. Repp (Eds.), *The Regular Education Initiative: Alternative Perspectives on Concepts, Issues, and Models.* Sycamore, IL: Sycamore.

Campbell, J. R. (1991). So what do we do with the poor, non-White female? Issues of gender, race and social class in mathematics and equity. *Peabody Journal of Education, 66*, 95-112.

Carpenter, R. L. (1985). Mathematics instruction in resource room: Instruction time and teacher competence. *Learning Disabilities Quarterly, 8*, 95-100.

Casto, G., & Mastropieri, M. A. (1986). The efficacy of early intervention programs: A meta-analysis. *Exceptional Children, 52*, 417-424.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* (2nd ed.). Hillsdate, NJ: Erlbaum.

College Board. (1999). *Reaching the top: A report on the national task force on minority high achievement.* New York: Author.

Cook, D. H., Mulrow, C. D., & Haynes, R. B. (1997). Systematic reviews: Synthesis of best evidence for clinical decisions. *Annuals of Internal Medicine, 126*(5), 376-380.

Cooper, H., & Hedges, L. V. (Eds.). (1994). *The Handbook of Research Synthesis.* New York: Russell Sage Foundation.

D'Amico, J. J. (2001). A closer look at the minority achievement gap. *ERS Spectrum* (Spring), 4-10.

Darling-Hammonds, L., & Sykes, G. (2003). Wanted: A national teacher supply policy for education: the right way to meet the "highly qualified teacher" challenge. *Educational Policy Analysis Archives, 11*(33).

De Corte, E. (2004). Mainstreams and perspectives in research on learning (mathematics) from instruction. *Applied Psychology: An International Review, 53*(2), 279-310.

Domestic Policy Council. (2006). *American competitiveness initiative: Leading the world in innovation.* Washington, DC: Office of Science and Technology Policy.

Education Commission of the States. (2003). *No Child Left Behind.* Retrieved March 25, 2005, from http://nclb.ecs.org/nclb

EPPI-Centre. (2004a). *EPPI-Centre home page.* Retrieved April 10, 2004, from http://eppi.ioe.ac.uk/cms/

EPPI-Centre. (2004b). *What is a systematic review?* Retrieved December 16, 2004, from http://eppi.ioe.ac.uk/EPPIWeb/home.aspx?page=/reel/about_reviews.htm

EPPI-Centre. (2006). *EPPI-Centre methods for conducting systematic reviews, Version 1.1*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Etsey, Y. K., & Snetzler, S. (1998). *A meta-analysis of gender differences in student attitudes toward mathematics.* Paper presented at the American Educational Research Association, San Diego, CA.

Ferguson, R. F. (1998). Teachers' perceptions and expectations and the Black-White test score gap. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap*. Washington, DC: Brookings Institution Press.

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence*. Hillsdale, NJ: Erlbaum.

Forness, S. R., Kavale, K. A., Blum, I. M., & Lloyd, J. W. (1997). Mega-analysis of meta-analyses. *Teaching Exceptional Children*, 4-9.

Fuchs, L. A., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 53*, 199-208.

Fusarelli, L. D. (2004). The potential impact of the No Child Left Behind Act on equity and diversity in American education. *Educational Policy, 18*(1), 71-94.

Hanushek, E. A. (2001). Black-White achievement differences and governmental interventions, *Annual Meeting of the American Economic Association*. New Orleans, LA.

Hanushek, E. A., & Wobmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal, 116*, C63-C76.

Haskins, R., & Rouse, C. (2005). *Closing achievement gaps*. Princeton, NJ: Princeton University and the Brookings Institution.

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

Haycock, K., Jerald, C., & Huang, S. (2001). *Closing the gap: Done in a decade*. Washington, DC: The Education Trust.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107-128.

Hedges, L. V. (1982). Estimating effect size from a series of independent experiments. *Psychological Bulletin, 92*, 490-499.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hixson, J., & Tinzmann, M. B. (1990, 2004). *Who are the "at-risk" students of the 1990s?* Retrieved December 1, 2004, from http://www.ncrel.org/sdrs/areas/rpl_esys/equity.htm

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

International Association for the Evaluation of Educational Achievement. (2003). *Third international mathematics and science study repeat (TIMSS-R) 2003.* Washington, DC: Author.

Ireson, J., & Hallam, S. (1999). Raising standards: Is ability grouping the answer? *Oxford Review of Education, 25*(3), 343-358.

Jencks, C., & Phillips, M. (Eds.). (1998). *The Black-White test score gap*. Washington, DC: Brookings Institution Press.

Jimenez, E., & Lockheed, M., E. (1989). Enhancing girls' learning through single-sex education: Evidence and a policy conundrum. *Educational Evaluation and Policy Analysis, 11*(2), 117-142.

Kulik, C.-L. C., & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal, 19*(3), 415-428.

LePore, P. C., & Warren, J. R. (1997). A comparison of single-sex and coeducational Catholic secondary schooling: Evidence for the National Educational Longitudinal Study of 1988. *American Educational Research Journal, 34*(3), 485-511.

Lewin, K. (1936). *Principles of topological psychology*. New York: McGraw-Hill.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: SAGE Publications, Inc.

Lubienski, S. T. (2002). A closer look at Black-White mathematics gaps: Intersections of race and SES in NAEP achievement and instructional practices data. *Journal on Negro Education, 71*(4), 269-287.

Marzano, R. J., Gaddy, B. B., & Dean, C. (2000). *What works in classroom instruction*. Aurora, CO: Mid-continent Research for Education and Learning (McREL).

Mastropieri, M. A., & Scruggs, T. E. (1989). Constructing more meaningful relations: Mnemonic instruction for special populations. *Educational Psychology Review, 1*(2), 83-111.

Mastropieri, M. A., Scruggs, T. E., & Shiah, S. (1991). Mathematics instruction for learning disabled students: A review of research. *Learning Disabilities Research & Practice, 6*, 89-98.

Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science, 26*, 49-63.

McLeod, T., & Armstrong, S. (1982). Learning disabilities in mathematics-skill deficits and remedial approaches at the intermediate and secondary grades. *Learning Disabilities Quarterly, 5*, 305-311.

Medical Research Council of South Africa. (2004). *What is a systematic review?* Retrieved December 16, 2004, from http://www.mrc.ac.za/cochrane/systematic.htm

Miller, S. P., Butler, F. M., & Lee, K.-h. (1998). Validated practices for teaching mathematics to students with learning disabilities: A review of literature. *Focus on Exceptional Children, 31*(1), p. 24.

Morgan, G. A., Gliner, J. A., & Harmon, R. J. (2006). *Understanding and evaluating research in applied and clinical settings*. Mahwah, NJ: Lawrence Erlbaum Associates.

Moses, R. P. (1994). Remarks on the struggle for citizenship and math/science literacy. *Journal of Mathematics Behavior, 13*, 107-111.

Moses, R. P., & Cobb Jr., C. F. (2001). *Radical equations: Math, literacy and civil rights*. Boston: Beacon Press.

National Commission on Excellence in Education. (1983). *A nation at risk*. Washington, DC: National Commission on Excellence in Education.

National Education Goals Panel. (2002, March 6). *National education goals panel*. Retrieved December 1, 2004, from http://www.negp.gov

National Governors' Association. (2004). *Closing the achievement gap*. Retrieved January 24, 2005, from http://www.subnet.nga.org/educlear/achievement/index.html

National Research Council. (2000). *How people learn: Brain, mind, experience, and school. Expanded edition*. Committee on Developments in the Science of Learning and Committee on Learning Research and Educational Practice. J.D. Bransford, A.I. Brown and R.R. Cockings (Eds.). Commission on Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. (2005). *How students learn: History, mathematics, and science in the classroom*. Committee on *How People Learn*, A Target Report for Teachers. M.S. Donovan and J.D. Bransford, Editors. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Science Board. (2004). *Science and engineering indicators 2004*. Retrieved November 30, 2004, from http://www.nsf.gov/sbe/ere/seind04/c1.htm

Office of Educational Research and Improvement (OERI). (2002). *National institute on the education of at-risk students*. Retrieved December 22, 2006, from http://www.ed.gov/offices/OERI/At-Risk/index.html

Organization for Economic Co-operation and Development. (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: Author.

Phillips, D. C., & Burbules, N. C. (2000). *Postpositivism and educational research*. Lanham, MD: Rowman & Littlefield Publishers, Inc.

Robinson, T. R., Smith, S. W., Miller, M. D., & Brownell, M. T. (1999). Cognitive behavior modification of hyperactivity/impulsivity and aggression: A meta-analysis of school-based studies. *Journal of Educational Psychology, 91*, 195-203.

Rothstein, H. R., Turner III, H. M., & Lavenberg, J. G. (2004). *The Campbell Collaboration: Information retrieval policy brief*. Retrieved November 10, 2005, from http://www.campbellcollaboration.org/MG/briefs.asp

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Schneider, M. (2007). *Performance of 12th-graders*. Retrieved February 23, 2007, from http://nces.ed.gov/whatsnew/commissioner/remarks2007/2_22_2007.asp

Schoenfeld, A. H. (2002). Making mathematics work for all children: Issues of standards, testing and equity. *Educational Researcher, 31*(1), 13-25.

Sells, L. W. (1978). Mathematics: A critical filter. *Science Teacher, 45*, 28-29.

Shadish, W., & Myers, D. (2004). *Research design policy brief*. Retrieved November 3, 2004, from http://www.duke.edu/web/c2method/ResDesPolicyBrief.htm

Silva, E. (2007). *On the clock: Rethinking the way schools use time*. Washington, DC: Education Sector.

Singham, M. (2003). The achievement gap: Myths and reality. *Phi Delta Kappan, April*, 586-591.

Skiba, R., & Casey, A. (1985). Interventions for behavior disordered students: A quantitative review and methodological critique. *Behavioral Disorders, 10*, 239-252.

Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research, 60*(3), 471-499.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32*, 752-760.

Snyder, T. D., Tan, A. G., & Hoffman, C. M. (2004). *Digest of education statistics 2003*. Washington, DC: Government Printing Office.

Swanson, H. L., Hoskyn, M., & Lee, C. (1999). *Interventions for students with learning disabilities: A meta-analysis of treatment outcomes*. New York: The Guilford Press.

Talbot, E., Lloyd, J. W., & Tankersley, M. (1994). Effects of reading comprehension interventions for students with learning disabilities. *Learning Disabilities Quarterly, 17*, 223-232.

Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.

U.S. Department of Education. (2006). *The condition of education 2006*. National Center for Educational Statistics, Washington, DC: U.S. Government Printing Office.

Valentine, J. C., & Cooper, H. (2003). *Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes*. Washington, DC: What Works Clearinghouse.

Valentine, J. C., & Cooper, H. (2004). *What Works Clearinghouse study design and implementation assessment device (Version 1.1)*. Washington, DC: U.S. Department of Education.

Vitale, M. R., & Romance, N. R. (2006). Research in science education: An interdisciplinary perspective. In J. Rhoton & P. Shane (Eds.), *Teaching Science in the 21st Century*. Arlington, VA: NSTA Press.

Wachs, T. (1987). Developmental perspective on designing for development. In C. Weinstein & T. David (Eds.), *Spaces for Children* (pp. 291-307). New York: Plenum Press.

Wenglinsky, H. (2004). *Closing the racial achievement gap: The role of reforming instructional practices*. Retrieved April 25, 2005, from http://www.epaa.asu.edu/epaa/v12n64

What Works Clearinghouse. (2004a). *Curriculum-based interventions for increasing K-12 mathematics achievement*. Retrieved December 1, 2004, from http://www.whatworksclearinghouse.org

What Works Clearinghouse. (2004b). *WWC study review standards*. Retrieved November 12, 2004, from http://www.whatworksclearinghouse.org/reviewprocess/study_standards_final.pdf

What Works Clearinghouse. (2006a). *Middle School Math Curricula*. Retrieved December 10, 2006, from http://www.whatworks.ed.gov/Topic.asp?tid=03&ReturnPage=default.asp

What Works Clearinghouse. (2006b). *Technical details of WWC-conducted computations*. Retrieved March 17, 2007, from http://www.whatworks.ed.gov/reviewprocess/conducted_computations.pdf

White, W. A. T. (1986). A meta-analysis of effects of direct instruction in special education. *Education and Treatment of Children, 11*, 364-374.

Xin, Y. P., & Jitendra, A. K. (1999). The effects of instruction in solving mathematical word problems for students with learning problems: A meta-analysis. *The Journal of Special Education, 32*(4), 207-225.

APPENDIXES

Appendix A:

Abstract Review Document

APPENDIX A

Abstract Review Document

Field	Question	Options	Definition
User Def 1	Does the study sample appear to address the review questions?	Yes = Y No = N Not sure = Y	Desired sample is: Secondary or postsecondary students, grades 6–14, middle school through college, ages 12-20+ years old, involved in math through calculus. Grade 6 – secondary curriculum (grades 6-8: okay, grades 4-6: not secondary). "At-risk" students.
User Def 2	Does the study appear to examine the effects of an intervention?	Yes = Y No = N Not sure = Y	An intervention is some action taken by the researcher or other entity to change an experience or is a systematic change in the school (i.e., block scheduling). Consider the intervention an independent variable and the outcome the dependent variable. Eliminate studies that are clearly attribute only studies, as well as test validation and drug therapy studies.
User Def 3	Does the study appear to measure an outcome?	Yes = Y No = N Not sure = Y	The researcher measures the impact of the intervention on an outcome. Outcome is the dependent variable. Could be a score on a math assessment test, an attitude, retention (repeating a school grade), placement in higher math class, etc. Also could be a change in attitude, such as math appreciation.
User Def 4	Does the study appear to be a quantitative study?	Yes = Y No = N Not sure = Y	Is the study quantitative? If quantitative, uses numerical values to measure something (i.e., math achievement as measured by a standardized math assessment/test, such as CSAP, ACT, SAT, high-risk assessment). Qualitative is a description, theory generation, a case study, or "soft" measures, (no numerical values used).

Appendix B:

Keywording Document

Appendix B

Keywording Document (Version 1.7)

Section A: Source of report.

A.1 Source of report From where was the report/article obtained?	A.1.1 Citation: Article/report is identified from the reference list of another article or report. A.1.2 Contact: Article/report is identified from a conversation (electronic or verbal) with an individual(s). A.1.3 Handsearch: Article/report is identified from a hardcopy search of journals or other sources. (Example: While copying an article, another article in the same journal appears useful or article is contained in review of references in previously selected article.) A.1.4 Unknown: Article/report source is unidentifiable. Note: every attempt should be made to determine the article/report source. (Example: An article/report is left on your desk by an unknown person.) A.1.5 Electronic database (please specify): Article/report is identified from an electronic search of, e.g., ERIC, PsychInfo, Medline, Dissertation Abstracts, SSCI, other.
A.2 Status Refers to the publication status of the report/article.	A.2.1 Published: The article/report is published in a journal or book that has an ISSN or ISBN number. A.2.2 In press: The article/report will be published in a journal or book that has an ISSN or ISBN number. A.2.3 Unpublished: The article/report is not published in a journal or book that has an ISSN or ISBN number.
A.3 Linked reports Refers to whether the article/report is one of at least two articles/reports (that have already been obtained) from a singular study.	A.3.1 Not linked: The article/report is the only report obtained from a singular study. A.3.2 Linked (please provide details): The article/report is one of at least two reports obtained from a singular study. (Example: An

	article and a dissertation both report the findings of a study conducted by the researcher while completing his dissertation.) A.3.3 Don't know.

Section B: Description of the study

B.1 In which country/countries was the study carried out?	B.1.1 United States B.1.2 Other Write in which country the study was conducted. Refer to the author's description within the article/report.
B.2 Empirical Design Refers to the empirical design type and whether the interventions are active, inactive, or non-existent <u>as they are described in the article/report.</u>	B.2.1 Multi-Group Quantitative: The article/report utilizes a multi-group quantitative methodology. (Example: A study that compares the outcomes of a treatment group to a control group.) B.2.2 One Group Pretest Posttest Design: The article/report utilizes quantitative methodology but only uses a single pretest, a single treatment group, and a single posttest. . B.2.3 One Group Correlational: The article/report utilizes a correlational or associational quantitative methodology with only a single group. B.2.4: Single Subject: The article/ report utilizes a single subject quantitative design. (Example: A multiple baseline design.) B.2.5 Multiple Regression: The article/report utilizes a complex associational analysis to predict a criterion (dependent) variable from several predictor (independent) variables. B.2.6 Regression Discontinuity Design: The article/report utilizes a regression analysis with a single group associational design after participants are assigned to either the intervention program or comparison groups on the basis of a cut-off score on a specific pre-test measure. B.2.7 Interrupted Time Series Design: The

	article/report utilizes a quasi-experimental time-series design with multiple pre-tests administered prior to the intervention to establish a suitable baseline. B.2.8 Qualitative: (Please describe the study design, using some of the author's terms.) B.2.9 None of the above: The article/report is purely descriptive and does not make use of a design that allows effect size calculation or it is missing so much information about the design as to render it functionally non-empirical.
B.3 Interventions Code the type of intervention. If not one of the types listed, please describe the intervention(s) in a brief sentence or phrase.	B.3.1 Curriculum: Interventions were the content, order of presentation of content or theoretical basis of the content is altered. B.3.2 Grouping Structures/Tracking: These include strategies and techniques that involve grouping students in a classroom or school environment for delivery of instruction. B.3.3 Pedagogical/Instructional: Interventions in which the method or approach of content delivery can be altered by the teacher. In doing so, the teacher selects the pedagogical or instructional intervention. Examples are direct instruction, constructivist approach, or the teaching of problem-solving skills. B.3.4: Personal/Affective: Interventions that include such strategies as retreats designed to enhance self-esteem, classroom based discussions, counseling, classes on interpersonal relations, etc. B.3.5: Professional Development/Teacher In-service training: Interventions in which the teacher(s) is the recipient of the training. B.3.6 Technology: These interventions involve the use of a computer or graphical calculator directly for instruction or to assist instruction. B.3.7 Tutoring Strategies: These include strategies such as peer tutoring, mentoring, one-on-one instructional assistance by senior citizens or older students, etc.

	B.3.8: School Restructuring: Non-instructional or curriculum interventions, but rather whole school or environmental interventions such as reducing class size, school-within-a-school, block scheduling, and alternative schools. B.3.9 Other: These are non-curricular or instructional interventions that are designed to enhance academic achievement and can not be placed in the categories above. B.3.10 Indeterminate: Although an intervention is stated to have been delivered in the study to improve academic achievement, the description of the intervention type is so vague, inchoate, or incomplete that it is impossible to provide a description of it.
B.4 Outcomes Code the type of outcome. If not one of the types listed, please describe the outcome(s) in a brief sentence or phrase	B.4.1 Cognitive Outcomes: These are outcomes that are either academic achievement as measured by standardized, teacher-made, or researcher-made tests, or performance measures of cognitive skills such as grade point average, number of correct words stated, or grades on classroom assignments. B.4.2: Affective Outcomes: These are largely attitude outcomes such as attitudes towards school or learning, classmates, or certain types of instructional situations. B.4.3 Process Outcomes: These are observational measures of academically important outcomes such as frequency of interactions with peers or teachers, frequency of requests for help, time on task, or frequency of the use of self-regulating strategies. B.4.4: Enrollment in Higher Mathematics: Outcomes associated with the student completing one level of math instruction and being rewarded by enrollment in higher level math classes. B.4.5: Other Outcomes (please describe): These are outcomes that do not fit any of the areas above. B.4.6: Indeterminate - Although an outcome is

	stated to have been measured in the study, the description of the outcome is so vague, inchoate, or incomplete that it is impossible to classify it.
B.5 What was the educational setting(s) being study? In what educational setting did the study take place, as reported by the author(s)?	B.5.1 Middle/junior high school: A public or private school including grades 6 through 8 or 9. Grade 6 must be associated with middle school or junior high school curriculum – must be part of a secondary school, not part of an elementary school (i.e., if part of 4^{th} through 6^{th} grade, do not include as it is part of an elementary school curriculum). B.5.2 High school: A public or private school including grades 9 or 10 through 12. Freshmen/women, sophomores, juniors, and seniors. 9^{th} grade defaults as high school. B.5.3 Community/junior college (2 year): A public or private two-year school that awards certificates, associate degrees, and/or facilitates transfer to a 4-year institution. B.5.4 Four-year college: A public or private university or college that awards baccalaureate degrees. B.5.5 Home: Any grade levels where academic learning is primarily conducted in the home. B.5.6 Residential school: A school of any grade level in which the students reside on the premises. B.5.7 Correctional institution: A jail, prison, or juvenile detention center. B.5.8 Other educational setting (please specify): Educational setting does not fall under one of the above. B.5.9 Indeterminate or non-educational: No description of the setting was given, OR the description given was so vague, or incomplete as to render it impossible to determine if it was an educational setting.

Section C: Description of the study population	
C.1 Math Content Area	C.1.1 Number Sense, Properties and Operations C.1.2 Measurement C.1.3 Geometry and Spatial Sense C.1.4 Data Analysis, Statistics, and Probability C.1.5 Algebra and Functions C.1.6 Pre-Calculus C.1.7 Calculus C.1.8 Other:
C.2 Grade of learners Refers to the educational grade level of the participants as reported by the author(s)	C.2.1 Specify single grade or range of grades of the learners: C.2.2 Unknown
C.3 Is gender of learners specified? Refers to the biological gender of the participants as reported by the author(s). Check the sample description to see if gender is mentioned and reported in the statistical outcome tables.	C.3.1: Yes C.3.2: No
C.4 Is race/ethnicity of learners specified? Refers to the race/ethnicity of the participants as reported by the author(s). Check the sample description to see if race/ethnicity is mentioned and reported in the statistical outcome tables.	C.4.1: Yes C.4.2: No
C.5 Is other At-Risk Group specified? At-risk students are primarily those whose appearance, language, culture, values, communities, and family structures do not match those of the dominant white culture that schools were designed to serve and support. These are students for whom traditional approaches to organizing and delivering instruction have worked least well – those most "at risk" of not completing or benefiting from	C.5.1: Yes C.5.2: No

their elementary and secondary educational experiences (Hixson & Tinzmann, 1990).	

Section D: Other

D.1 Does the report meet our inclusionary criteria?	D.1.1: Yes D.1.2: No Check "No" only if you have checked B.2.9, or B.3.10, or B.4.6, or B.5.9, or <u>ALL</u> of the following are checked: C.3.2, C.4.2, C.5.2.
D.2 Keyworder(s): Refers to the person who completed the keywording of the article/report.	D.2.1 Details: Type the name or initials of the person who entered in the keywording information on an article/report.

Appendix C:

Data Extraction Document

Appendix C

Data Extraction Protocol Draft (Version 1.0)

Section A: Identification of the study

A.1 Enter EPPI unique identification number. (IT number)	A.1.1 EPPI unique number:
A.2 Name of the primary reviewer.	A.2.1 Jim Dugan A.2.2 Jerry Overmyer A.2.3 Other
A.3 Name of the secondary reviewer.	A.3.1 Jim Dugan A.3.2 Jerry Overmyer A.3.3 Other
A.4 Date of the review.	A.4.1 Review date:
A.5 Linked studies	A.5.1 Not linked A.5.2 Linked to other studies (Please identify by unique ID number -- IT XXXXX) A.5.3 Don't Know
A.6 Specify a reason why this study was excluded from this review.	A.6.1 Not applicable A.6.2 Sample not appropriate for this review (For example, did not include the appropriate grade levels, or included the appropriate grade levels but were combined with other grades and could not be separated.) A.6.3 Inadequate construct validity of the intervention (For example, not a true intervention study, the intervention was not adequately identified; the description of the intervention was inconsistent with the theoretical name it was given by the study authors.) A.6.4 Inadequate construct validity of the outcome measure (For example, no outcome measure was identified, an outcome measure was identified but the measure was not valid for this review or was inappropriate for the intervention.)

	A.6.5 Inadequate comparison information for multi-group designs (For example, no information was given to ensure comparability of groups, i.e. randomization, matched pairs, pretest covariate, etc.) A.6.6 Inadequate statistical information regardless of design (For example, adequate statistical information to calculate an effect size was not provided, or inadequate statistical information was given with which effect sizes for disenfranchised groups could be calculated.) A.6.7 Single subject study A.6.8 Qualitative study

Section B: Study purpose, problem statement, research questions and research design.

B.1 What was the author(s)' stated purpose for doing the study? List specific purpose	B.1.1 Stated (Please use author(s)' statement.) B.1.2 Not stated or unclear
B.2 What was the problem statement? List specific problem, in given. Generally, is more general or global than purpose statement	B.2.1 Problem statement explicitly stated. (Please provide - type in, copy and paste, or scan.) B.2.2 Problem statement not clear or not relevant. B.2.3 No problem statement identified.
B.3 What is/are the research question(s) or hypotheses identified for the study?	B.3.1 Research question(s) or hypotheses explicitly stated. (Please provide - type in, copy and paste, or scan.) B.3.2 Research question(s) or hypotheses not clear or not relevant. B.3.3 No research question(s) or hypotheses identified.

B.4 What is the design type?	B.4.1 Between groups (A comparison of groups) B.4.2 Within subjects (A time series design in which the same group of subjects is assessed two or more different times or a single group, pretest posttest design.) B.4.3 Mixed design (Both multiple groups and multiple data collection periods on the outcome measure of interest.) B.4.3 Multiple regression design B.4.4 Cannot be determined
B.5 If between groups or mixed design, how did the authors describe the characteristics of the comparison group(s)?	B.5.1 No treatment control (Includes "traditional class alternative") B.5.2 Alternative treatment (A second treatment different from the "traditional class alternative".) B.5.3 Wait list control (A group is told they are on the "wait list" for treatment, then used a control. Are typically offered treatment after the experiment is over.) B.5.4 Other (Please specify) B.5.5 Not stated or unclear
B.6 If the study was a single group, pretest/posttest, did the author(s) provide a reliability measure or correlation between the pretest and the posttest scores?	B.6.1 Yes B.6.2 No B.6.3 Not applicable B.6.4 Not stated or unclear
B.7 Was the study sufficiently informed by or linked to an existing body of empirical and/or theoretical evidence?	B.7.1 Yes B.7.2 No

B.8 Would you rate the quality and execution of the research approach as acceptable and appropriate?	B.8.1 Yes B.8.2 No
B.9 Given the description of the study, can it be replicated with a reasonable level of fidelity?	B.9.1 Yes B.9.2 No

Section C: Study participants, extended characteristics

C.1 Grade of learners If identified by age, calculate by using equation: (age) – 5 = grade	C.1.1 Specify single grade or range of grades of the learners (Please enter grade or range of grades.) C.1.2 Not stated or unclear
C.2 What was the total number of participants involved in the study?	C.2.1 Stated (Please specify) C.2.2 Not stated or unclear
C.3 What were the gender percentages of the participants in the study?	C.3.1 Percent male C.3.2 Not stated or unclear
C.4 What was the socio-economic status of the majority of the participants in this study?	C.4.1 Stated (Please specify) C.4.2 Not stated or unclear
C.5 What was the race/ethnicity and percentage of the majority participants in this study?	C.5.1 Percent majority: (Please specify race/ethnicity and percentage of the majority.) C.5.2 Not stated or unclear
C.6 What was the specific at-risk population factor identified in the study?	C.6.1 Gender C.6.2 Low-academic achievement C.6.3 Low-SES C.6.4 High minority percentage (non-White) C.6.5 Non-English speaker (ELL)

	C.6.6 Special needs designation C.6.7 Multiple at-risk factors C.6.8 Other (Please specify) C.6.9 Not stated or unclear
C.7 What was the attrition rate?	C.7.1 As reported for all groups (or for study population as a whole if only one group), the attrition rate was: (Please specify for each sub-group. Calculate these percentages if the author does not give them in the study text.) C.7.2 Not stated or unclear
C.8 What was the urbanicity of this study?	C.8.1 Primarily rural (Please give percentage) C.8.2 Primarily suburban (Please give percentage) C.8.3 Primarily urban (Please give percentage) C.8.4 Mixed (Please give percentages) C.8.5 Not stated or unclear

Section D: Data collection and analysis

D.1 Which method was used to collect the data on the outcome measure selected for this study? Please specify name and source of data collection method, if appropriate.	D.1.1 State standards test D.1.2 Commercially developed test D.1.3 Teacher/researcher-made test D.1.4 Questionnaire or survey D.1.5 Observation D.1.6 Multiple instruments

	D.1.7 Other D.1.8 Not stated or unclear
D.2 Was the reliability of the outcome measure selected for this study assessed and considered to be acceptable? Such as internal consistency, test - re-test, Cronbach's alpha, or interrater reliability. Standardized student assessment data collection is assumed to be acceptable as is. (r = 0.7 or higher).	D.2.1 Yes D.2.2 No
D.3 Was the validity of the outcome measure selected for this study assessed and considered to be acceptable? Such as direct mention of criterion related, face validity or construct validity	D.3.1 Yes D.3.2 No
D.4 Was sufficient evidence presented to ensure that the operationalized measure reflects the construct of interest? If the intervention dealt with the use of calculators in algebra I, did the assessment of student achievement involve the use of calculators with algebra I?	D.4.1 Yes D.4.2 No
D.5 If applicable, was interrater reliability assessed and considered to be acceptable?	D.5.1 Yes D.5.2 No D.5.3 Not applicable.
D.6 What statistics were used to analyze the data from the outcome measure selected for this study? What statistical analysis procedure was used?	D.6.1 Stated (Please specify) D.6.2 Not stated or unclear
D.7 Does the author describe methods for ensuring the validity of data analysis? As a means of justification for the statistical analysis procedure used.	D.7.1 Yes D.7.2 No

D.8 Would you rate the quality and execution of the data collection and analysis as acceptable and appropriate?	D.8.1 Yes D.8.2 No

Section E: Intervention description

E.1 Using the author(s)' words, what was the specific definition of the intervention?	E.1.1 Stated (Please use author's words.) E.1.2 Not stated or unclear
E.2 Was there a reasonably adequate description of how the intervention was implemented?	E.2.1 Yes E.2.2 No
E.3 Was the setting in which the intervention was delivered consistent with the intent of the intervention?	E.3.1 Yes E.3.2 No
E.4 Would you rate the quality and implementation of the intervention as acceptable and appropriate?	E.4.1 Yes E.4.2 No
E.5 Who delivered the intervention?	E.5.1 Researcher E.5.2 Teacher E.5.3 Other (Please specify) E.5.4 Not stated or unclear
E.6 Who was the recipient of the intervention?	E.6.1 Individual student E.6.2 Classroom of students E.6.3 Multiple classrooms E.6.4 Entire school E.6.5 Multiple schools E.6.6 Teacher

	E.6.7 Other (Please specify) E.6.8 Not stated or unclear
E.7 What was the duration of time over which the intervention was implemented?	E.7.1 Less than one day E.7.2 One day to 1 week E.7.3 Eight days to 1 month E.7.4 More than 1 month to 3 months E.7.5 More than 3 months to 6 months E.7.6 More than 6 months to 1 year E.7.7 More than 1 year to 2 years E.7.8 More than 2 years E.7.9 Other: (Please specify) E.7.10 Not Stated or unclear
E.8 What was the frequency/intensity of the intervention as delivered on a routine basis? (I.e., 55 minutes per day, one class period, three times per week?)	E.8.1 Stated (Please specify) E.8.2 Not stated or unclear
E.9 Estimate the total amount of time (total hours) involved with the intervention. (Dosage)	E.9.1 Estimated as (Please specify) E.9.2 Cannot estimate
E.10 In what year(s) was the intervention conducted?	E.10.1 Please indicate the year(s), if stated in the study. E.10.2 Not stated or unclear

Section F: Outcomes

F.1 What was the specific outcome of	F.1.1 Outcome explicitly stated: (Please

the study, presented in the author(s)' words.	provide – type in, copy & paste or scan.) F.1.2 Not stated or unclear
F.2 For multi-group studies, what was the unit of allocation (or assignment) for each intervention and control or comparison group?	F.2.1 Individuals F.2.2 Classrooms F.2.3 A school F.2.4 Multiple schools F.2.5 School districts F.2.6 Other F.2.7 Not stated or unclear
F.3 For multi-group studies, how were groups assigned?	F.3.1 Randomized F.3.2 Cluster randomized (i.e., intact classes were assigned to treatment conditions). F.3.3 Randomized matched pairs were assigned prior to implementation of treatment. F.3.4 Matched pairs were identified after treatment (ex post facto design). F.3.5 Groups were assigned according to convenience or some other non-equivalent technique. F.3.6 Not stated or unclear
F.4 For multi-group studies, was the allocation to intervention and control groups completed so that group membership was unknown to members?	F.4.1 Yes F.4.2 No
F.5 For multi-group studies, were participants aware which group they were in when the intervention was	F.5.1 Yes F.5.2 No

taking place?	
F.6 Was outcome measurement done in such a way that those conducting the measurement were unaware of group membership?	F.6.1 Yes F.6.2 No
F.7 Was the unit of data analysis the same as the unit of allocation (or assignment)? (See G.2) Were the results reported according to the unit of allocation? For example, if individuals were allocated to different groups, results from individuals should be analysed and reported.	F.7.1 Yes F.7.2 No
F.8 Are there any obvious shortcomings in the numerical reporting in this study?	F.8.1 Yes (write in) F.8.2 No
F.9 Would you rate the quality and alignment of the outcome with the intervention as acceptable and appropriate?	F.9.1 Yes F.9.2 No
F.10 For the outcome(s) of interest, are effect sizes able to be calculated?	F.10.1 Yes, as reported by the author. (Please specify for each outcome in the study.) F.10.2 Yes, it appears from the data given that the effect size(s) can be calculated. F.10.3 Not clear if the effect size(s) can be calculated.

Appendix D:

Matrix of Prior Systematic Reviews with Bibliography

Appendix D

Matrix of Prior MA Systematic Reviews

#	Author	Study Purpose	Intervention Type Studied	Specific Intervention(s)	Study Population
1	Anderson, Yilmaz, & Washburn-Moses (2004)	To locate empirical studies that have investigated the use of an academic intervention that was effective in enhancing the academic performance of middle and high school students with learning disabilities, which could ultimately enhance the academic performance of students without LD.	Instructional	Mnemonic instruction Graphic organizer Guided notes Classwide peer tutor Coached elaboration Inquiry teaching	MS & HS, G 6-12 w/LD
2	Baker, Gersten, & Lee (2002)	To analyze findings from experimental research conducted in schools to improve the mathematics achievement of students struggling to learn math.	Instructional	Provide data to teacher and student Peer tutoring Feedback to parents Explicit instruction	Low-achieving students, K-12
3	Barley, Lauer, Arens, Apthorp, Englert, Snow, & Akiba (2002)	Synthesis on the effectiveness of strategies designed to assist low-achieving students during the school day.	Instructional	(a) General Instruction (approach), (b) Cognitively-oriented Instruction (approach), (c) Grouping structures (Intervention), (d)Tutoring (intervention), (e) Peer Tutoring (Intervention), (f) Computer-Assisted Instruction (Intervention),	K-12 students with low-achievement
4	Burns (2004)	Explored research literature, using meta-analytic techniques, to examine the effectiveness of various instructional ratios and thereby further refine the instructional level for drill tasks.	Instructional	Levels of instruction, drill ratio of <50% known, 50%-60% known, 70%-85% known, and 90% known were levels tested	K-12

Appendix D

Matrix of Prior MA Systematic Reviews

#	Author	Study Purpose	Intervention Type Studied	Specific Intervention(s)	Study Population
5	Ellington (2003)	The analysis covers the calculator's influence on student performance in (a) operational, computational and conceptual skills, and (b) general problem-solving skills including two aspects: the number of problems attempted and ability to select the appropriate problem-solving strategy.	Instructional	Calculator usage during instruction and during assessment	Precollege math students, K-12 mainstreamed students
6	Etsey & Snetzler (1998)	Gender differences in student attitudes towards mathematics	NA	NA	Grades 1-12 and postsecondary
7	Finn (1998)	Overview of recent research on the effects of class size on the academic performance and behavior of students at risk. Focuses primarily on two studies: Indiana's PRIME	Class size	Class size	K-3, early grades
8	Forness, Kavale, Blum, & Lloyd (1997)	Reports on some of the broad studies (meta-analyses) of research in special education that seem to indicate effective and noneffective teaching strategies.	All	All	Special education in general
9	Hambree (1990)	To integrate the findings of the research on mathematics anxiety, regarding its nature, effects, and relief; to describe relationships and effects with scale-invariant metrics; to reduce or resolve theoretical issues surrounding the construct.	NA	NA	Grades 3-12 and postsecondary

Appendix D

Matrix of Prior MA Systematic Reviews

#	Author	Study Purpose	Intervention Type Studied	Specific Intervention(s)	Study Population
10	Hambree (1992)	To study four regions of problem solving: characteristics of problem solvers, conditions for harder and easier problems, effects of different instructional methods on problem-solving performance and effects of classroom-related conditions on problem-solving performance.	Instructional	Cognitive problem-solving	K-12 + postsecondary
11	Jitendra & Xin (1997)	The purpose of this article was to summarize the findings of published mathematics research intervention studies designed for students with mild disabilties and these at-risk students and to provide suggestions for classroom practices and future research in mathematics word-problem-solving instruction.	Instructional	Word-problem solving including a) representational techniques, b) strategy-training procedures, c) task variations and CAI.	Age 7-65 (M=14.1) w/ mild disabilities and/or at-risk of failure
12	Kroesbergen & Van Luit (2003)	Meta-analysis was undertaken of those studies concerned with mathematics intervention for students with special education needs in both general and special elementary education.	Instructional	Systematic use of instructional strategies in prepatory math, basic skills and problem-solving	Elementary students

Appendix D

Matrix of Prior MA Systematic Reviews

#	Author	Study Purpose	Intervention Type Studied	Specific Intervention(s)	Study Population
13	Ma (1997)	Investigates the relationship between attitude towards mathematics (ATM) and achievement in mathematics (AIM).	NA	NA	K-12, General
14	Ma (1999)	Investigates the relationship between anxiety towards mathematics and achievement in mathematics (AIM).	NA	NA	K-12 students, general
15	Maccini & Hughes (1997)	The purpose of this review is threefold: (a) to determine the nature (type and focus) of the interventions, (b) to provide implications for current practice, and (c) to identify proiorities for future research.	Instructional	Instructional types are behavioral, cognitive and alternative delivery systems.	G 6-12 w/LD
16	Mastropieri, Scruggs, & Shiah (1991)	Studies that evaluated the effectiveness of mathematics teaching strategies to learning disabled students were reviewed.	Instructional	Interventions were subdivided into behavioral interventions, cognitive interventions, and alternative delivery systems.	K-12 w/LD
17	Miller, Butler, & Lee (1998)	To extend the work of Mastropieri, Scruggs & Schiah (1991) and provide an undated literature review related to validated math practices for students with learning disabilities.	Instructional	Any instructional	K-12 students with LD

Appendix D

Matrix of Prior MA Systematic Reviews

#	Author	Study Purpose	Intervention Type Studied	Specific Intervention(s)	Study Population
18	Pierce, Reid, & Epstein (2004)	To examine research that has been conducted with students with emotional and behavioral disorders (EBD) using teacher-mediated interventions for academic outcomes.	Instructional	Teacher-mediated interventions including	K-12 with EBD
19	Ryan, Reid, & Epstein (2004)	Cooperative Learning	Instructional	Peer-mediated interventions	K-12 with EBD
20	Springer, Stanne, & Donovan (1999)	This meta-analysis of research on college students in SMET is intended to facilitate a greater understanding of the effects of small-group learning in classrooms and programs at the postsecondary level.	Instructional	Cooperative learning	Grades 13 & 14, undergrad, SMET
21	Swanson & Hoskyn (1998)	To identify effective interventions to teaching reading to LD students	Instructional	Direct instruction, strategy instruction or combination of two	All ages with LD

Appendix D

Matrix of Prior MA Systematic Reviews

#	Author	Study Purpose	Intervention Type Studied	Specific Intervention(s)	Study Population
22	Swanson, Hoskyn, & Lee (1999)	Was conducted to provide a comprehensive synthesis of instructional intervention studies that address [students with learning disabilities].	Instructional	Academic, Cognitive, Behavioral domains	Students with LD
23	Weichel-Murawski & Swanson (2001)	The purpose of this study is to synthesize data-based articles pertaining to co-teaching between general and special education personnel.	Instuctional	Other, co-teaching, special education with general education	K-12
24	What Works Clearinghouse (2004)	This review focuses on curriculum-based interventions outlining the fundamentals of mathematics that students should know and be able to do, instructional programs and materials that organize the mathematical content, and assessments.	Curricular	Math 2-12, Middle School only	Middle school (elementary and high school reviews pending)
25	Xin & Jitendra (1999)	The present review uses a meta-analysis technique to aggregate intervention studies on solving word problems and examined the relationship between study characteristics and intervention outcomes.	NA	NA	K-16, students with LD

Appendix D

Matrix of Prior MA Systematic Reviews

#	Author	Time Period of Studies	Research Questions	Major Findings
1	Anderson, Yilmaz, & Washburn-Moses (2004)	1986 - 2002	NA	Those deemed most effective were: (a) Mnemonic instruction, (b) Graphic organizers, (c) Guided notes, (d) Classwide peer tutoring, (e) Coached elaboration, and (e) Inquiry teaching
2	Baker, Gersten, & Lee (2002)	1971 - 2002	NA	Effective interventions included: (a) Provide data teacher (ES = 0.57 wt, 0.71 unwt), Peer tutoring (ES = 0.62 wt, 0.66 = unwt), Feedback to parents (ES = 0.43 wt, 0.42 unwt), (d) Explicit instruction (ES = 0.65 wt, 0.58 unwt)
3	Barley, Lauer, Arens, Apthorp, Englert, Snow, & Akiba (2002)	1985 - 2002	What are effective instructional strategies that can be used in classrooms to assist low-achieving students?	(a) General Instruction (approach), Inconclusive, (b) Cognitively-oriented Instruction (approach), positive effects, (c) Grouping structures (Intervention), positive effects, (d)Tutoring (intervention), positive effects, (e) Peer Tutoring (Intervention), positive effects, (f) Computer-Assisted Instruction (Intervention), Conclusive positive effects.
4	Burns (2004)	1980 - 2002	1. Of the drill ratios found in the literature, which led to the largest effect? 2. Which type of dependent variable studied, academic outcomes or student preferences, led to the strongest effect? 3. In which stage of learning were the effects of using drill tasks strongest?	The 90% known group had the highest median effect size coefficient. Acquisition tasks had advantage over Proficience tasks, and academic outcomes had larger mean effect size than student preferences.

Appendix D

Matrix of Prior MA Systematic Reviews

#	Author	Time Period of Studies	Research Questions	Major Findings
5	Ellington (2003)	Jan 1983 through March 2002	Student achievement and student attitude	Students operational skills and problem-solving skills improved when calculators were an intergral part of testing and instruction. Students using calculators had better attitudes towards mathematics.
6	Etsey & Snetzler (1998)	1970-1995	Magnitude of gender differences in (1) general attitudes, (2) anxiety, (3) self-confidence, (4) attitude towards mathematics as a male domain, (5) usefulness, (6) moderators.	Gender differences do exist but are small
7	Finn (1998)	1987-1985	NA	Small class size has an advantage over larger class size in early primary grades. Small class advantage was consistantly greater for minority students that whites.
8	Forness, Kavale, Blum, & Lloyd (1997)	NA	NA	Those teaching strategies that work are: Mnemonic strategies (ES = 1.63), Enhanced reading comprehension (ES = 1.13), Behavior modification (ES = 0.93), Direct instruction (ES = 0.84), Cognitive behavior modification (ES = 0.71), Formative evaluation (ES = 0.70), Early intervention (ES = 0.68)
9	Hambree (1990)	All	1. Is there a causal direction in the relationship between mathematics anxiety and mathematics performance? 2. Does test anxiety subsume mathematics anxiety? 3. Are behaviors related to mathematics anxiety more pronounced in females than males?	1. Mathematics anxiety depresses performance. Higher achievement consistently accompanies reduction in mathematics anxiety. No evidence that poor performance causes mathematics anxiety. 2. It is unlikely that mathematics anxiety is purely restricted to testing. 3. Across all grades, females report higher mathematics anxiety levels than males.

Appendix D

Matrix of Prior MA Systematic Reviews

#	Author	Time Period of Studies	Research Questions	Major Findings
10	Hambree (1992)	1920-1980	NA	Direct significant links were found between problem solving and measures of basic performance, especially skills in basic mathematics. Heuristics approach proved somewhat better than other teaching approaches in middle grades 6-8 and distinctly better in high school.
11	Jitendra & Xin (1997)	1986 - 1995	N/A	1. Representational techiques (diagramming) effective only when involves relationships among components of the word problem. 2. Interventions that combine varied instruction components (i.e., cognitive and metacognitive procudures) more effective than interventions in isolation. 3. Sequencing word problems and meaningful contexts are effective. 4. CAI tutorial programs effective when empirically validated strategies and curricular designs are incorporated.
12	Kroesbergen & Van Luit (2003)	1985-2000	(1) Which domain (preparatory skills, basic skills, problem solving) most investigated, highest effect size? (2) Is there a trend in outcomes as a function of ... study characteristics, (3) What variables can explain the largest part of the between-studies variance in the total sample and in the three different domains seperately?	Q1: Of three domains, basic facts most investigated, but no significant differences in effect sizes. Q2: No significant differences found. Q3: Four variables explain 69% of between-studies variance: design (single subject= higher effect size), duration of intervention (longer duration had smaller effect size), domain (problem-solving domain had lower effect sizes), method (Self-instruction more effective)

Appendix D

Matrix of Prior MA Systematic Reviews

#	Author	Time Period of Studies	Research Questions	Major Findings
13	Ma (1997)	1966 - 1993	(1) What is the magnitude of the general relationship between ATM and AIM? (2) Is this relationship consistent across attribute variables? (3) How is this relationship affected by interaction effects among gender, grade and ethnicity? (4) What is the magnitude of the causal relatioship between ATM and AIM?	(1) Overall effect size was 0.12 (SD 0.12, 0.13).
14	Ma (1999)	1975-1999	(1) What is the magnitude of the relationship between anxiety toward mathematics and AIM? (2) How is this relationship affected by interaction effects among gender, grade and ethnicity?	The relationship between anxiety toward mathematics and AIM was r = -0.27 (d = -0.54).
15	Maccini & Hughes (1997)	July 1988 to August 1995	None	Practices that appear effective include: teacher-directed instruction, instructional design curriculum variables, three-term contingency trails, strategy instructions, self-monitoring strategies, contextualized word problems and cooperative homework teams.
16	Mastropieri, Scruggs, & Shiah (1991)	June 1975 to June 1988	None	Virtually all instructional procedures were said to be effective by the authors.
17	Miller, Butler, & Lee (1998)	Between 1988 and 1998	None	Strategy and self-regulation interventions with step-by-step processes are effective for students with learning disabilities. Direct instruction is also effective. So are use of manipulatives and computer-assisted instruction.

Appendix D

Matrix of Prior MA Systematic Reviews

#	Author	Time Period of Studies	Research Questions	Major Findings
18	Pierce, Reid, & Epstein (2004)	1963 to 2004	Listed six research questions. Addressed particpant characteristics, interventions studied, academic areas involved, types of dependent measures, reliability and validity measures and outcomes from the studies with effect sizes.	Results included: majority of studies focused on reading: most interventions were done for short periods of time; studies lacked complete participant descriptors; the majorities of outcomes were positive; and , teacher-mediated interventions were effective.
19	Ryan, Reid, & Epstein (2004)		(1) What are population characteristics, (2) Actual placement of EBD students (3) What academic outcomes were addressed, (4) What interventions were most common, (5) What were the effects of the different types of peer-mediated interventions?	Characteristics and settings defined. All content areas addressed.Variety of peer-tutoring interventions used.
20	Springer, Stanne, & Donovan (1999)	1980 - 1999	What are main effects of small-group learning on three broad categories of outcomes among SMET undergraduates: achievement, persistence, and attitudes. What effects of four moderators: potential sources for bias; differing effects of small-groups; characteristics of small-groups; and different types of outcomes.	Main effects: greater achievement with small groups (d = 0.51); greater persistance (d = 0.46); more favorable attitudes (d = 0.55). A number of moderators were discovered.
21	Swanson & Hoskyn (1998)	1963-1997	(1) test whether treatments that include components of direct, strategy or both effective, (2) test the assumption that some instructional domains are more resistant to intervention that others, (3) test whether studies that include samples of intelligence and reading scores yield difference outcomes	Effect size: Direct instruction = 0.68, strategy instruction = 0.72. Suggest use combined model of direct instruction and strategy instruction to teach reading.

Appendix D

Matrix of Prior MA Systematic Reviews

#	Author	Time Period of Studies	Research Questions	Major Findings
22	Swanson, Hoskyn, & Lee (1999)	1972-1997	None	Overall, studies that include components related to strategy instruction or direct instruction yield larger effect sizes than studies that use nonstrategy and nondirect instruction models.
23	Weichel-Murawski & Swanson (2001)	1991 - 1998	(1) Does the magnitude of co-teaching outcomes vary as a function of grade, gender, length or study, or severity or type of disability? (2) Do studies that produce the largest effect size vary from other studies as a function of the type of dependent measure of focus (e.g., grades, social outcomes, achievement)?	Overall mean effect size was 0.40, suggesting that co-teaching is a moderately effective procedure for influencing student outcomes. Insufficient data for analysis of effect on grade, gender, etc. Could not answer question (2) due to large variability range of data from studies; inconclusive.
24	What Works Clearinghouse (2004)	All	(1) Which curriculum-based interventions are effective in increasing the learning of mathematics content and skills among elementary, middle and high school students? (2) Are some interventions more effective than others for learning certain types of math content and skills? (3) Are some interventions more effective for certain types of students, particularily students who lag behind in mathematics achievement?	Found that only two computer-based algebra programs called "I CAN Learn Mathematics" and "Cognitive Tutor" had studies showing that students actually learned more with their programs compared with other programs.
25	Xin & Jitendra (1999)	1980 - 1996	1. What is the general effectiveness of word-problem-solving interventions? 2. Is intervention effectiveness related to importance of student characteristics? 3. Are treatment outcomes related to instructional features? 4. Is there a relationship between methodological features? 5. What is the effectiveness of word-problem-solving instruction in fostering skill maintenance and generalization?	1. CAI had highest effect size (d = 1.80), representation (d = 1.77), strategy (d = 0.74) and other (d = 0.00). 2. Grade/age: postsec (d = 1.68), Secondary group (d = 0.78), Elementary group (d = 0.47). IQ: <85 (d = 1.87), >85 (d = 0.51). Classification label: LD (d = 0.50), mixed disabilities (d = 1.96), at-risk (d = 1.90). (I.e., grade/age, IQ and label indicated at mediating influence on effect size. Class setting was also greater ES than pull-out, Individual instruction greater than small-group. Others indicated also.

Appendix D

Matrix of Prior MA Systematic Reviews

#	Author	Number of Studies	Statistics /Analysis	Comments
1	Anderson, Yilmaz, & Washburn-Moses (2004)	17	Not done	All disciplines (not just math) including English, math, science, social science. Used narrative sysnthesis of intervention research.
2	Baker, Gersten, & Lee (2002)	17	Effect size estimate (Cooper & Hedges, 1994)	Analysis well done
3	Barley, Lauer, Arens, Apthorp, Englert, Snow, & Akiba (2002)	118	Some, depending on sample	Extensive review. Good use of DIAD
4	Burns (2004)	13	Effect size, 55 individual analysis	

Appendix D

Matrix of Prior MA Systematic Reviews

#	Author	Number of Studies	Statistics /Analysis	Comments
5	Ellington (2003)	54	Meta-analysis with effect size (ES = q, homogenity Q, U 3%)	Effect size = standardized mean difference, difference in the experimental and control group means divided by the pooled standard deviation. Weighted effect size
6	Etsey & Snetzler (1998)	96	Effect size Cohan's *d*	
7	Finn (1998)	NA	NA	Special case: Commissioned to be an overview of recent research on the effects of class size on the academic performance and behavior of students at risk.
8	Forness, Kavale, Blum, & Lloyd (1997)	18 meta-analyses	Effect size, Cohan's *d*	Is a short review of meta-analyses.
9	Hambree (1990)	151	Effect size	

Appendix D

Matrix of Prior MA Systematic Reviews

#	Author	Number of Studies	Statistics /Analysis	Comments
10	Hambree (1992)	487	Effect size, Cohan's *d*	See Hambree 1990 for list of reviewed studies
11	Jitendra & Xin (1997)	14	None	
12	Kroesbergen & Van Luit (2003)	58	Effect size, Cohan's *d*	

Appendix D

Matrix of Prior MA Systematic Reviews

#	Author	Number of Studies	Statistics /Analysis	Comments
13	Ma (1997)	103	Effect size, Cohan's *d*	
14	Ma (1999)	26	Effect size, Cohan's *d* and correlation coefficient *r*	
15	Maccini & Hughes (1997)	19	None	Extension of Mastropieri (1991). Systematic review but no meta-analysis
16	Mastropieri, Scruggs, & Shiah (1991)	30	Not done	Considered one of the pioneering reviews in this area of interest.
17	Miller, Butler, & Lee (1998)	54	None	Literature review with no meta-analysis

Appendix D

Matrix of Prior MA Systematic Reviews

#	Author	Number of Studies	Statistics /Analysis	Comments
18	Pierce, Reid, & Epstein (2004)	24	Effect size, Cohan's *d*	Special case: Considered only students with emotional and behavorial disabilities (EBD).
19	Ryan, Reid, & Epstein (2004)	14	Effect size, Cohan's *d*	Special case: Considered only students with emotional and behavorial disabilities (EBD).
20	Springer, Stanne, & Donovan (1999)	39	Effect size, Cohan's *d*	
21	Swanson & Hoskyn (1998)	192	Effect size, Cohan's *d*	Study looked at reading domain only, no math studies.

Appendix D

Matrix of Prior MA Systematic Reviews

#	Author	Number of Studies	Statistics /Analysis	Comments
22	Swanson, Hoskyn, & Lee (1999)	275	Effect size, Cohan's *d*	Effect size in mathematics (18 studies) was 0.40, among the lowest reported. Multiple academic cores studied.
23	Weichel-Murawski & Swanson (2001)	6	Effect size	None of the studies focused on middle or junior high schools. No final results due to sample number of studies, etc.
24	What Works Clearinghouse (2004)	77 studies reporting on 44 math curricular products	Effect size, Cohan's *d*	Is the first full review released by WWC. Expect elementary and high school curricular materials in 2005.
25	Xin & Jitendra (1999)	25	Effect size	Interesting study. Note that effect sizes are high throughout. For group design studies, as most were comparisons between two interventions (not an intervention and a control), used standardized mean change for within-subject comparisons. Used Hedges' .

Appendix E

Effect Size Calculation Resources

Appendix E

Effect Size Determination

This appendix provides information on how the effect size estimates were determined for each research design and statistical analysis procedure encountered during the review. According to Lipsey and Wilson (2001, p. 5), "[t]he key to meta-analysis, therefore, is defining an effect size statistic capable of representing the quantitative findings of a set of research studies in a standardized form ...."

The primary effect size statistics used in this review was the standardized mean difference effect size. This effect size measure is defined "as the difference between mean outcome of the intervention group and the mean outcome of the comparison group divided by the pooled within-group standard deviation" (What Works Clearinghouse, 2006, p. 1). This statistic is based on the existence of a two-variable relationship in which a pre-post contrast is made on the same construct prior to exposure to the intervention and after the exposure. The difference between the pre-intervention mean and the post-intervention mean is assumed to be the result of the intervention only. This pre-post contrast is compared to a pre-post contrast measured with a control or non-treatment group which does not receive the intervention.

The sign of the effect size is meaningful. The standardized mean difference is derived by subtracting the control group pre-post difference from the treatment group pre-post difference and dividing the result by the pooled standard deviation of the measure. This results in an effect size estimate in which a positive value indicates that the treatment resulted in an increase in the outcome measure (primarily identified as student mathematics achievement in this review) over that seen in the non-treatment or control group. In turn, a negative effect size estimate would indicate that the outcome variable decreased in value due to the effect of the experimental intervention.

The basic effect size index used was Cohen's *d*, defined as the difference between the experimental and the control group means divided by the pooled standard deviation (Cooper & Hedges, 1994). Mathematically, the effect size calculation is represented as:

$$ES = \frac{X_e - X_c}{s_p}$$

where ES is effect size, X_e is the mean score of the experimental or treatment group, X_c is the mean score of the control or non-treatment group, and s_p is the pooled standard deviation of the two groups.

The process used to calculate the effect size estimate for each of the studies included in this review was as follows: (1) locating information in the article on the type of statistical analysis done by the author in the study; (2) determination of the study being a pre-post comparison or a post-test only comparison; (3) locating the data showing differences (raw score) between the pre-test and the post-test for both the treatment and the control groups; (4) determining the sample size for both the treatment and the control group; and (5) performing the effect size estimate.

The standardized mean difference effect size estimate has been shown to be upwardly biased when based on smaller sample sizes (Hedges, 1981). Hedges has provided an unbiased effect size estimate as follows:

$$ES'_{sm} = [\,1 - \tfrac{3}{4N-9}]ES_{sm}$$

where N is the total sample size ($n_1 + n_2$) and ES_{sm} is the biased standardized mean difference effect size estimate. Thus, making the adjustment for sample size, the effect size index computed is then referred to as Hedges' g. As has been explained earlier, the EPPI-Reviewer online systematic review and meta-analysis software was used for this systematic review. Inputting data from the studies into the effect size calculator available within the EPPI-Reviewer resulted in the effect size estimates being expressed as Hedges' g.

One of the complications identified in estimating effect sizes and doing comparisons across studies is the reliability of information included in the effect size estimate. As might be expected, sample sizes in multiple, real-world study comparisons vary. Larger sample sizes give better estimates of the effect size. If one intends to combine multiple studies using a meta-analysis approach, an adjustment must be made to weight each effect size estimate by its' respective sample size. The way this is done is through the usage of an "inverse variance weight" (Lipsey & Wilson, 2001, p. 36). Consequently, two values must be calculated from data presented in a study: the effect size and the inverse variance weight.

Based on this, the inverse variance weight is calculated as:

$$w_{sm} = \frac{1}{SE_{sm}^2}$$

Multiple types of statistical analyses were observed in the studies included in this review. Consequently, calculation of the standardized mean difference effect sizes for different statistical treatments by the authors was needed. Three statistical analyses beyond the simple mean differenced were encountered in the study including comparisons using independent *t*-tests, analysis of variance using *F*-values, and a few ANCOVA analyses with both adjusted and unadjusted means.

For the independent *t*-test (t) and sample sizes (n) that vary, the following equation is used:

$$ES = t\sqrt{\frac{n_1+n_2}{n_1 n_2}}$$

where t is the measure of the *t*-test and n_1 and n_2 are the sample sizes of the treatment and control groups, respectively. For calculation of the standardized mean difference effect size using an *F*-ratio (F) from a one-way ANOVA and samples sized (n) for each group, the following equation is used:

$$ES = \sqrt{\frac{F(n_1+n_2)}{n_1 n_2}}$$

where F is the *F*-ratio of the ANOVA and n_1 and n_2 are the sample sizes of the treatment and control groups, respectively.

A few studies in the final group of those qualified during the review used analysis of covariance (ANCOVA) to analyze the comparisons. For those studies where pretest assessment data was used as a covariate to adjust the mean difference, Hedges' g was calculated as:

$$ES'_{sm} = \frac{X'_1 - X'_2}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{(n_1+n_2-2)}}}$$

Appendix F

Listing of Included Studies by Interventions

Appendix F1

Summary of Intervention: Curriculum

#	Author (Year)	Intervention	Design Type	Design	Overall ES	At-risk Factor	Primary ES	Secondary ES
1	Beckett (1993)	Curr: Algebridge	Mixed	Pre/post, exp, ctrl	-0.12	Gender	X	X
2	Bottge et al (2004)	Curr: EAI vs. TBI	Mixed	Pre/post, exp, ctrl	-0.02	Spec Needs	X	X
3	Covington-Clarkson (2001)	Curr: CMP	Between	Post only, exp, ctrl	-0.45	Hi minority	X	X
4	Isley (1998)	Curr: Algebra I	Between	Post only, exp, ctrl	0.24	Gender, SES	M = 0.27	F = 0.17
5	Lafferty (1996)	Curr: Saxon vs trad	Between	Post only, exp, ctrl	0.42	Gender	M = 0.63	F = 0.05
6	Riordan (2001)	Curr: Standards	Between	Post only, exp, ctrl	0.22	Gender	X	X
7	Thurber (2001)	Curr: Acad Pgms Excell	Between	Post only, exp, ctrl	0.36	Gender	X	X
8	Watson (1996)	Curr: SCRAP	Between	Post only, exp, ctrl	-0.19	Gender, Lo SES	M = -0.44	F = 0.11

Appendix F2

Summary of Intervention: Grouping/Tracking

#	Author (Year)	Intervention	Design Type	Design	Overall ES	At-risk Factor	Primary ES	Secondary ES
1	Cahan (1996)	Group/Track: Ability	Within	Regression discont	0.20	Low acad achieve	X	X
2	Kelly (1994)	Group/Track: Ability	Between	Post only, exp, ctrl	-0.04	Gender	M = 0.06	F = -0.13
3	Linchevski et al (1998) S1	Group/Track: Ability	Between	Regression discont		Low acad achieve		
4	Linchevski et al (1998) S2	Group/Track: Ability	Between	Post only, exp, ctrl	1.37	Low acad achieve	Low ach=3.952	
5	Lockart (1996)	Group/Track: Ability	Between	Post only, exp, ctrl	-0.08	Gender	M = -0.14	F = -0.06
6	Mason (1992)	Group/Track: Ability	Between	Post only, exp, ctrl	0.21	Low acad achieve	X	X
7	Tymann (2003)	Group/Track: Ability	Between	Post only, exp, ctrl	-0.59	Low acad achieve	X	X
8	Edwards (2002)	Group/Track: Single Sex	Mixed	Pre/post, exp, ctrl	0.15	Gender	X	X
9	Mulholland et al (2004)	Group/Track: Single Sex	Between	Post only, exp, ctrl	-0.20	Gender	F only = 0.899	X
10	Wear (1997) S1	Group/Track: Single Sex	Mixed	Pre/post, exp, ctrl	-0.09	Gender (Female)		
11	Wear (1997) S2	Group/Track: Single Sex	Mixed	Pre/post, exp, ctrl	0.22	Gender (Male)		

Appendix F3

Summary of Intervention: Instructional -- Cognitive and Metacognitive Strategies

#	Author (Year)	Intervention	Design Type	Design	Overall ES	At-risk Factor	Primary ES	Secondary ES
1	Butler (2003)	Ped/Inst: Cogn & Meta	Mixed	Pre/post, exp, ctrl	0.22	Gender	X	X
2	Chinnappan (1996)S1	Ped/Inst: Cogn & Meta	Between	Post only, exp, ctrl	0.78	Low acad achieve	Low Ach = 0.76	
3	Chinnappan (1996)S2	Ped/Inst: Cogn & Meta	Between	Post only, exp, ctrl	0.67	Low acad achieve	Low Ach = 0.64	
4	McMann PK (1994)	Ped/Inst: Cogn & Meta	Mixed	Pre/post, exp, ctrl	0.39	Gender	M = 0.38	F = 0.47
5	Mevarech (1997) S1	Ped/Inst: Cogn & Meta	Between	Post only, exp, ctrl	0.37	Low acad achieve	Lo Ach = 0.29	
6	Mevarech (1997) S2	Ped/Inst: Cogn & Meta	Between	Post only, exp, ctrl	0.38	Low acad achieve	Lo Ach = 0.13	
7	Mevarech (1999)	Ped/Inst: Cogn & Meta	Between	Post only, exp, ctrl	0.43	Low acad achieve	Low ach = 0.25	X
8	Mevarech (2003)	Ped/Inst: Cogn & Meta	Mixed	Pre/post, exp, ctrl	0.43	Low acad achieve	Low ach = 0.62	X
9	Oladunni (1998)	Ped/Inst: Cogn & Meta	Between	Post only, exp, ctrl	0.96	Gender	X	X
10	Philips (1993)	Ped/Inst: Cogn & Meta	Between	Post only, exp, ctrl	-0.38	Gender	M = -0.499	F = -0.236
11	Schwanenberger (1991)	Ped/Inst: Cogn & Meta	Between	Post only, exp, ctrl	0.52	Gender	M = 0.51	F = 0.53
12	Shannon & Allen (1998)	Ped/Inst: Cogn & Meta	Mixed	Pre/post, exp, ctrl	0.36	Gender	M = 0.34	F = 0.32
13	Smith, JP (2000)	Ped/Inst: Cogn & Meta	Mixed	Pre/post, exp, ctrl	0.14	Gender	X	X
14	Vasquez (1998)	Ped/Inst: Cogn & Meta	Mixed	Pre/post, exp, ctrl	0.32	Multiple at-risk	Grp A = 0.580	Grp B = 0.054

Appendix F4

Summary of Intervention: Instructional -- Cooperative Learning

#	Author (Year)	Intervention	Design Type	Design	Overall ES	At-risk Factor	Primary ES	Secondary ES
1	Abrami (1993)	Ped/Inst: Coop Learn	Between	Post only, exp, ctrl	0.31	Low acad achieve	Low ach=0.795	X
2	Alkhateeb (2002)	Ped/Inst: Coop learn	Between	Post only, exp, ctrl	0.11	Gender	M = 0.37	F = -0.20
3	Barbato (2000)	Ped/Inst: Coop learn	Between	Post only, exp, ctrl	0.84	Gender	M = 0.83	F = 0.84
4	Hoek, et al (1997)	Ped/Inst: Coop Learn	Mixed	Pre/post, exp, ctrl	0.39	Low acad achieve	X	X
5	Hooper (1992)	Ped/Inst: Coop Learn	Between	Post only, exp, ctrl	0.12	Low acad achieve	Low ach=-0.28	X
6	Kinel (1994)	Ped/Inst: Coop learn	Between	Post only, exp, ctrl	0.52	Gender, others	X	X
7	Lopez (1991)	Ped/Inst: Coop Learning	Between	Post only, exp, ctrl	0.59	Gender	M = 0.84	F = 0.59
8	Lopez (1993)	Ped/Inst: Coop Learning	Between	Post only, exp, ctrl	0.76	Gender	M = 1.38	F = 0.30
9	Miller, WA (1991)	Ped/Inst: Coop learn	Mixed	Pre/post, exp, ctrl	-0.47	Low acad achieve	Low ach= -0.597	X
10	Thompson, EO (1992)	Ped/Inst: Coop learn	Mixed	Pre/post, exp, ctrl	0.28	Gender	X	X

Appendix F5

Summary of Intervention: Instructional -- Peer Tutoring

#	Author (Year)	Intervention	Design Type	Design	Overall ES	At-risk Factor	Primary ES	Secondary ES
1	Calhoon & Fuchs (2003)	Ped/Inst: Peer Tutoring	Mixed	Pre/post, exp, ctrl	0.36	Hi minority	X	X
2	Carlin (2002)	Ped/Inst: Tutoring	Mixed	Pre/post, exp, ctrl	0.02	Gender	M = -0.039	F = 0.062
3	Fields (2002)	Ped/Inst: Tutorial	Between	Post only, exp, ctrl	0.30	Gender	X	X
4	Steinhauser (2000)	Ped/Inst: Team teaching	Between	Post only, exp, ctrl	0.38	Gender	X	X
5	Watt, JX (1991)	Ped/Inst: Tutoring	Between	Post only, exp, ctrl	0.41	Gender	X	X

Appendix F6

Summary of Intervention: Instructional -- Other

#	Author (Year)	Intervention	Design Type	Design	Overall ES	At-risk Factor	Primary ES	Secondary ES
1	Bottge et al (2001)	Ped/Inst: Anchored instr	Mixed	Pre/post, exp, ctrl	-0.46	Gender, low ach	X	X
2	Brainerd (1991)	Ped/Inst: Extend LO	Mixed	Pre/post, exp, ctrl	0.07	Gender	M = -0.104	F = 0.254
3	Brown, SA (1996) S1	Ped/Inst: Writing, Alge1	Between	Post only, exp, ctrl	-0.22	Gender	X	X
4	Brown, SA (1996) S2	Ped/Inst: Writing, Basic	Between	Post only, exp, ctrl	0.24	Gender	X	X
5	Colarulli (1998)	Ped/Inst: Learning style	Mixed	Pre/post, exp, ctrl	0.24	Gender	M = 0.34	F = 0.16
6	Hawkins & Brady (1994)	Ped/Inst: Pauses	Within	Single group repeat	0.33	Gender	M = 0.20	F = 1.63
7	Kirby & Boulter (1999)	Ped/Inst: Learning style	Between	Post only, exp, ctrl	0.06	Gender	X	X
8	Lee, DW (1991)	Ped/Inst: Other	Between	Post only, exp, ctrl	-0.02	Low acad achieve	Low ach=-0.9	
9	Loiacano (1997)	Ped/Inst: Manipulatives	Between	Post only, exp, ctrl	-0.06	None	X	X
10	McCaffrey, et al (2001)	Ped/Inst: Other	Between	Post only, exp, ctrl	-0.14	Gender	X	X
11	McLaughlin (2000)	Ped/Inst: Double dose	Between	Matched, exp, ctrl	0.31	Gender, Hi minority	M = -0.27	F = 0.65
12	Smith, KC (2002)	Ped/Inst: Homework	Between	Post only, exp, ctrl	-0.06	Unknown	X	X
13	Torres (1999)	Ped/Inst: Double dose	Between	Post only, exp, ctrl	0.43	Low acad achieve	X	X
14	Weidemann, WJ (1990)	Ped/Inst: Manipulatives	Between	Post only, exp, ctrl	-0.35	Gender, Lo spatial	M = -0.61	F = -0.10

Appendix F7

Summary of Intervention: School Restructuring

#	Author (Year)	Intervention	Design Type	Design	Overall ES	At-risk Factor	Primary ES	Secondary ES
1	Algaze (1998)	School Rstr: Block Sch	Between	Post only, exp, ctrl	0.16	Multiple ?	0.30	X
2	Anderson (1995)	School Rstr: Block Sch	Mixed	Pre/post, exp, ctrl	0.06	Gender	X	X
3	Cavalos (2002)	School Rstr: Block Sch	Mixed	Pre/post, exp, ctrl	0.09	High minority	Non-W = 0.08	X
4	Crotteau (2002)	School Rstr: Block Sch	Between	Post only, exp, ctrl	-0.10	Low acad achieve	Lo Ach = -0.70	
5	Drummond (2001)	School Rstr: Block Sch	Between	Post only, exp, ctrl	0.07	Gender, Low SES, H	M = -0.03	F = 0.15
6	Fortner (2000)	School Rstr: Block Sch	Between	Post only, exp, ctrl	-0.01	Gender	X	X
7	Fowler (2001)	School Rstr: Block Sch	Between	Post only, exp, ctrl	-0.02	Gender, race	M = 0.02	F = -0.06
8	Harris (1996)	School Rstr: Block Sch	Mixed	Pre/post, exp, ctrl	0.02	Gender	M = 0.04	F = 0.01
9	Schroth (1996)	School Rstr: Block Sch	Mixed	Pre/post, exp, ctrl	0.09	Low acad achieve	0.13	X
10	Thomas, LM (2002)	School Rstr: Year round	Between	Post only, exp, ctrl	0.29	Lo SES, hi minority	Lo SES = 0.23	non-W = 0.27
11	Trlica (1998)	School Rstr: Block Sch	Mixed	Pre/post, exp, ctrl	0.14	Gender, lo SES,	X	X
12	Washington (1999)	School Rstr: accel math	Mixed	Pre/post, exp, ctrl	0.05	Gender	M= 0.11	F = 0.09

Appendix F8

Summary of Intervention: Technology-based

#	Author (Year)	Intervention	Design Type	Design	Overall ES	At-risk Factor	Primary ES	Secondary ES
1	Alberta Learn (2002)	Technology: General	Between	Post only, exp, ctrl	0.12	Low acad achieve	Low ach=0.17	X
2	Alkhateeb (2002)	Tech: Graph calculator	Mixed	Pre/post, exp, ctrl	0.11	Gender	M = -0.26	F = 0.36
3	Alkhateeb (2002)	Tech: Graph vs Maples	Mixed	Pre/post, exp, ctrl	-0.30	Gender	M = -0.05	F = -0.54
4	Baker (1998)	Technology: General	Mixed	Pre/post, exp, ctrl	0.35	Gender	M = 0.48	F = 0.24
5	Cortez (1996)	Tech: CAI	Between	Post only, exp, ctrl	0.06	Hi minority	ESL = 0.199	
6	Dugdale, DeKoven, Ju (199	Technology: General	Within	Post only, exp, ctrl	0.29	Gender	X	X
7	Ejiofor (2002)	Tech: Calculators	Between	Post only, exp, ctrl	0.27	Low acad achieve	Low ach=-0.003	
8	Forsyth (1991)	Tech: CAI	Between	Post only, exp, ctrl	0.52	Gender	M = 0.279	F = 0.909
9	Haynes (1999)	Tech: CAI	Between	Post only, exp, ctrl	-0.12	Gender	M = -0.12	X
10	McKenzie (1999)	Tech: CAI	Mixed	Pre/post, exp, ctrl	-0.98	Gender	M = -1.01	F = -0.89
11	Merckling (1999)	Tech: Graph calculator	Between	Post only, exp, ctrl	0.34	Gender	X	X
12	Rendall (2001)	Tech: CAI	Mixed	Pre/post, exp, ctrl	0.32	Gender	X	X
13	Rinaldi (1997)	Tech: CAI	Between	Post only, exp, ctrl	-0.80	Gender	M = -1.33	F = -0.19
14	Ruthven (1990)	Tech: Graphing calc	Between	Post only, exp, ctrl	1.73	Gender	M = 1.18	F = 2.35
15	Savelli-Keska (1999)	Tech: ILS	Mixed	Pre/post, exp, ctrl	-0.22	Gender	M = -0.03	F = -0.40
16	Sayavong (2004)	Tech: ILS	Between	Post only, exp, ctrl	-0.30	Gender	X	X
17	Schuck (2003)	Technology: Internet	Mixed	Pre/post, exp, ctrl	0.31	Low acad achieve	Lo Ach = -0.847	X
18	Vahey, et al (2000)	Tech: CAI	Between	Post only, exp, ctrl	0.68	Gender	M = 0.51	F = 0.84
19	West, RC (1993)	Tech: ILS	Between	Multiple regression	0.13	Gender	X	X
20	Wohlgehagen (1992)	Tech: CAI	Mixed	Pre/post, exp, ctrl	-0.41	Gender	M = -0.47	F = -0.39
21	Yusuf (1995)	Tech: Logo & Geometry	Between	Post only, exp, ctrl	2.97	Gender	M = 5.04	F = 2.06

Appendix F9

Summary of Intervention: Other, Non-Instructional

#	Author (Year)	Intervention	Design Type	Design	Overall ES	At-risk Factor	Primary ES	Secondary ES
1	Alster (1997)	Other: Extend test time	Within	Single grp pre/post	0.77	Spec Needs	LD only = 0.83	X
2	Good, et al (2003)	Personal/Affective	Between	Post only, exp, ctrl	0.90	Gender	M = 0.63	F = 1.10
3	Huynh, et al (2004)	Other: Oral admin	Between	Post only, exp, ctrl	-0.06	Spec Needs	X	X
4	Kolacinski (2003)	Personal/Affective	Mixed	Pre/post, exp, ctrl	0.28	None	X	X
5	Roulier (1999)	Personal/Affective	Mixed	Pre/post, exp, ctrl	0.47	Gender	M = 0.48	F = -0.08

Appendix G

List of included study citations

Appendix G

Mathematics Achievement Systematic Review Final Studies Included

Abrami, P. C., Chambers, B., D'Apollonia, S., Farrell, M., & De Simone, C. (1992). Group outcome: The relationship between group learning outcome, attributional style, academic achievement, and self-concept. *Contemporary Educational Psychology, 17*(3), 201-210.

Alberta Learning. (2002). *The Learning Equation (TLE) final report* (ERIC OCLS No. Clearinghouse: TM034233). Edmonton, Alberta, Canada.

Algaze, L. J. (1998). A comparison of mathematics achievement, attendance, and behavior of at-risk, minority, and female students in block-scheduled, and traditional settings. *DAI, 59*(09A), 3293.

Alkhateeb, H. M. (2002). A preliminary study of achievement, attitudes toward success in mathematics, and mathematics anxiety with technology-based instruction in brief calculus. *Psychological Reports, 90*(1), 47-57.

Alkhateeb, H. M., & Jumaa, M. (2002). Cooperative learning and algebra performance of eighth grade students in United Arab Emirates. *Psychological Reports, 90*(1), 91-100.

Alkhateeb, H. M., & Wampler, J. (2002). Graphing calculators and students' conceptions of the derivative. *Perceptual and Motor Skills, 94*(1), 165-170.

Alster, E. H. (1997). The effects of extended time on algebra test scores for college students with and without learning disabilities. *Journal of Learning Disabilities, 30*(2), 222-227.

Anderson, K. M. (1995). Block scheduling in the middle school: Does it make a difference? *DAI, 57*(02A), 519.

Baker, J. J. (1998). Effects of a generative instructional design strategy on learning mathematics and on attitudes towards achievement. *Dissertation Abstracts International Section A: Humanities & Social Sciences, 58*(7-A), Univ.

Barbato, R. A. (2000). Policy implications of cooperative learning on the achievement and attitudes of secondary school mathematics students. *DAI, 61*(06A), 2113.

Beckett, E. C. (1993). An investigation of math scores and self-esteem of Mexican-American seventh and eighth-grade students after Algebridge activities. *Dissertation Abstracts International, 53*(3-A), Univ.

Bottge, B. A., Henrichs, M., Chandler, P. A., & Serlin, R. C. (2001). Anchoring adolescents' understanding of math concepts in rich problem-solving environments. *Remedial and Special Education, 22*(5), 299-314.

Bottge, B. A., Henrichs, M., Mehta, Z. D., Rueda, E., Hung, Y.-H., & Danneker, J. (2004). Teaching mathematical problem solving to middle school students in math, technology education, and special education classrooms. *Research in Middle Level Education Online, 27*(1).

Brainerd, W. E. (1991). The effect of supplementary measuring activities on achievement and attitude of non-honors high school geometry students. *DAI, 52*(11A), 3850.

Brown, S. A. (1996). The mathematics learning log and its effects on mathematics achievement, anxiety, and communication. *Dissertation Abstracts International Section A: Humanities & Social Sciences, 57*(6-A), Univ.

Butler, F. M., Miller, S. P., Crehan, K., Babbitt, B., & Pierce, T. (2003). Fraction instruction for students with mathematics disabilities: Comparing two teaching sequences. *Learning Disabilities Research & Practice, 18*(2), 99-111.

Cahan, S., & Linchevski, L. (1996). The cummulative effect of ability grouping on mathematical achievement: A longitudinal perspective. *Studies in Educational Evaluation, 22*(1), 29-40.

Calhoon, M. B., & Fuchs, L. S. (2003). The effects of peer-assisted learning strategies and curriculum-based measurement on the mathematics performance of secondary students with disabilities. *Remedial & Special Education, 24*(4), 235-245.

Carlin, D. M. (2002). The relationship between the academic improvement plan and seventh graders' reading and mathematics achievement. *DAI, 63*(05A), 1776.

Cavazos, S. (2002). The A/B alternating block versus the modified block in the middle school. *DAI, 64*(01A), 32.

Chinnappan, M., & Lawson, M. J. (1996). The effects of training in the use of executive strategies in geometry problem solving. *Learning & Instruction, 6*(1), 1-17.

Colarulli, R. (1998). Effects of teaching mathematics to learning style perceptual preference on academic achievement of seventh-grade middle school students. *DAI, 59*(02A), 0438.

Cortez, N. F., Jr. (1996). A study of computer-assisted instruction and its effects on both ESL and non-ESL at-risk minority secondary students' achievement. *DAI, 57*(12A), 5123.

Covington Clarkson, L. M. (2001). The effects of the Connected Mathematics Project on middle school mathematics achievement. *DAI, 61*(12A), 4709.

Crotteau, R. A. (2002). The combined effect of ability and format of instruction on Algebra I achievement: A study of the superiority of accelerated block scheduling over traditional scheduling as moderated by student ability. *DAI, 63*(06A), 2189.

Drummond, P. A. (2001). A comparative analysis of students' academic achievement under block and traditional scheduling. *DAI, 63*(02A), 548.

Dugdale, S., DeKoven, E., & Ju, M.-K. (1998). Computer course enrollment, home computer access, and gender: Relationships to high school students' success with computer spreadsheet use for problem solving in pre-algebra. *Journal of Educational Computing Research, 18*(1), 49-62.

Edwards, S. R. (2002). Gender-based and mixed-sex classrooms: The relationship of mathematics anxiety, achievement, and classroom performance in female high school math students. *Dissertation Abstracts International Section A: Humanities & Social Sciences, 62*(8-A), Univ.

Ejiofor, P. C. (2002). Integrating non-programmable calculators in the Nigerian 10th-grade mathematics curriculum: Effects on student achievement and attitudes. *DAI, 63*(05A), 1755.

Fields, M. A. (2002). The effect of a three-week TAAS Blitz on overall student performance on the exit level of the Texas Assessment of Academic Skills Test. *DAI, 64*(12A), 4434.

Forsyth, E. (1991). Computer assisted instruction: Advisement and interactions with selected learner characteristics in learning 10th grade mathematics. *MAI, 31*(02), 0520.

Fortner, T. M. (2000). The effect of gender and the term a course is taught in the block schedule structure on Stanford Achievement Test mathematics scores. *DAI, 61*(03A), 864.

Fowler, R. W. (2001). The impact of block scheduling and traditional scheduling on academic achievement. *DAI, 63*(02A), 548.

Good, C., Aronson, J., & Inzlicht, M. E. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology, 24*(6), 645-662.

Harris, M. A. S. (1996). The effects of block scheduling on Alabama end of course algebra scores, attendance, and discipline. *DAI, 58*(03A), 0666.

Hawkins, J., & Brady, M. P. (1994). The effects of independent and peer guided practice during instructional pauses on the academic performance of students with mild handicaps. *Education & Treatment of Children, 17*(1), 1-28.

Haynes, L. C. (1999). Gender differences in the use of a computer-based mathematics game: Strategies, motivation, and beliefs about mathematics and computers. *Dissertation Abstracts International Section A: Humanities & Social Sciences, 60*(9-A), Univ.

Hoek, D., Terwel, J., & van den Eeden, P. (1997). Effects of training in the use of social and cognitve strategies: An intervention study in secondary mathematics in co-operative groups. *Educational Research and Evaluation, 3*(4), 364-389.

Hooper, S. (1992). The effects of group composition on achievement, interaction, and learning efficiency during computer-based cooperative instruction. *Educational Technology Research & Development, 39*(3), 27-40.

Huynh, H., Meyer, J. P., & Gallant, D. (2004). Comparability of student performance between regular and oral administrations for a high-stakes mathematics test *Applied Measurement in Education, 17*(1), 39-57.

Isley, L. S. (1998). A study of subsequent mathematical outcomes for students taught Algebra I using "Algebra I: A Process Approach". *DAI, 59*(10A), 3764.

Kelly, W. M. (1994). Gender comparisons of attitudes towards mathematics in streamed and destreamed classrooms. *MAI, 33*(06), 1647.

Kinel, R. (1994). Cooperative learning: Reducing the Black/White achievement gap in middle school mathematics. *DAI, 55*(08A), 2311.

Kirby, J. R., & Boulter, D. R. (1999). Spatial ability and transformational geometry. *European Journal of Psychology of Education, 14*(2), 283-294.

Kolacinski, J. F. (2003). Mathematics anxiety and learned helplessness. *Dissertation Abstracts International Section A: Humanities & Social Sciences, 64*(6-A), Univ.

Lafferty, J. F. (1996). The links among mathematics text, students' achievement, and students' mathematics anxiety: A comparison of the incremental development and traditional texts. *Dissertation Abstracts International Section A: Humanities & Social Sciences, 56*(8-A), Univ.

Lee, D. W. (1991). A comparison of the effectiveness between the cooperative and individual learning on students' achievement and attitudes on a computer-assisted mathematics problem-solving task. *Dissertation Abstracts International, 52*(7-A), Univ.

Linchevski, L., & Kutscher, B. (1998). Tell me with whom you're learning, and I'll tell you how much you've learned: Mixed-ability versus same-ability grouping in mathematics. *Journal for Research in Mathematics Education, 29*(5), 533-554.

Lockart, G. M. (1996). Grouping practices and their effects on middle level gifted students. *DAI, 57*(10A), 4249.

Loiacano, M. L. (1996). Differences between traditional and non-traditional algebra courses on students perceptions of instruction and academic achievement. *Dissertation Abstracts International Section A: Humanities & Social Sciences, 57*(10-A), Univ.

Lopez, C. L. (1993). Effect of personalization of instructional context on the achievement and attitudes of Hispanic students. *Educational Technology Research & Development, 40*(4), 5-13.

Lopez, C. L., & Sullivan, H. J. (1991). Effects of personalized math instruction for Hispanic students. *Contemporary Educational Psychology, 16*(1), 95-100.

Mason, D. A. (1992). Assigning average-achieving eighth graders to advanced mathematics classes in an urban junior high. *Elementary School Journal, 92*(5), 587-599.

McCaffrey, D. F., Hamilton, L. S., Stecher, B. M., Klein, S. P., Bugliari, D., & Robyn, A. (2001). Interactions among instructional practices, curriculum, and student achievement: The case of standards-based high school mathematics. *Journal for Research in Mathematics Education, 32*(5), 493-517.

McKenzie, S. Y. (1999). Achievement and affective domains of Algebra I students in traditional or self-paced computer programs. *DAI, 60*(09A), 3297.

McLaughlin, N. L. (2000). Developmentally appropriate practices: Interventions to help at-risk high school students meet math graduation requirements. *DAI, 61*(07A), 2636.

McMann, P. K. (1994). The effects of teaching practice review items and test-taking strategies on the ACT mathematics scores of second-year algebra students. *DAI, 55*(04A), 0897.

Merckling, W. J. (1999). Relationship(s) between the perceptual preferences of secondary school students and their achievement in functions using a graphing calculator. *DAI, 60*(02A), 0371.

Mevarech, Z. R. (1997). IMPROVE: A multidimensional method for teaching mathematics in heterogeneous classrooms. *American Educational Research Journal, 34*(2), 365-394.

Mevarech, Z. R. (1999). Effects of metacognitive training embedded in cooperative settings on mathematical problem solving. *The Journal of Educational Research, 92*(1), 195-205.

Mevarech, Z. R., & Kramarski, B. (2003). The effects of metacognitive training versus worked-out examples on students' mathematical reasoning. *British Journal of Educational Psychology, 73*(4), 449-471.

Miller, W. A. (1991). Effects of a cooperative learning format on content, attitude, and learning preference in mathematics for a population of junior high school summer academy students. *Dissertation Abstracts International, 51*(5-A), Univ.

Mulholland, J., Hansen, P., & Kaminski, E. (2004). Do single-gender classrooms in coeducational settings address boys' underachievement? An Australian study. *Educational Studies, 30*(1), 19-32.

Oladunni, M. O. (1998). An experimental study on the effectiveness of metacognitive and heuristic problem solving techniques on computational performance of students in mathematics. *International Journal of Mathematical Education in Science and Technology, 29*(6), 867-874.

Philips, B. B. (1993). Writing to learn mathematics: The effects of writing/discussion versus discussion-only on immediate and delayed test performance in trigonometry. *Dissertation Abstracts International, 54*(5-A), Univ.

Rendall, L. T. (2001). The effectiveness of a computer-based instruction program: A comparative study. *DAI, 61*(12A), 4743.

Rinaldi, I. L. (1997). A study of the effects of computer-assisted instruction and teacher instruction on achievement in mathematics. *MAI, 35*(05), 1124.

Riordan, J. E., & Noyce, P. E. (2001). The impact of two standards-based mathematics curricula on student achievement in Massachusetts. *Journal for Research in Mathematics Education, 32*(4), 368-398.

Roulier, L. R. (1999). Effects of goal setting procedures on students' mathematical achievement and self-efficacy. *Dissertation Abstracts International Section A: Humanities & Social Sciences, 61*(1-A), Univ.

Ruthven, K. (1990). The influence of graphic calculator use on translation from graphic to symbolic forms. *Educational Studies in Mathematics, 21*, 431-450.

Savelli-Keska, S. (1999). Differences in mathematics achievement of low-achieving students in supplemental and regular integrated learning system programs at the middle school level. *DAI, 60*(03A), 0681.

Sayavong, P. (2004). Mexican-American and Mexican immigrant adolescents' academic achievement in technology-mediated mathematics classrooms. *Dissertation Abstracts International: Section B: The Sciences & Engineering, 64*(7-B), Univ.

Schroth, G., & Dixon, J. (1996). The effects of block scheduling on student performance. *International Journal of Educational Reform, 5*(4), 472-476.

Schuck, R. W. (2003). Responses to the use of the Internet with a California standards-based middle school algebra curriculum among at-risk, Latino students. *DAI, 64*(12A), 4345.

Schwanenberger, M. C. (1991). Test-wiseness training with sixth and seventh-grade students and its effect on standardized achievement test scores. *DAI, 52*(10A), 3576.

Shannon, H. D., & Allen, T. W. (1998). The effectiveness of a REBT training program in increasing the performance of high school students in mathematics. *Journal of Rational-Emotive & Cognitive-Behavior Therapy, 16*(3), 197-209.

Smith, J. P. (2000). The effects of chess instruction on the mathematics achievement of southern, rural, Black secondary students. *Research in the Schools, 7*(1), 19-26.

Smith, K. C. (2002). The effects of teacher homework review on student academic performance in middle grades mathematics. *DAI, 63*(01A), 124.

Steinhauser, J. B. (2000). Mathematics achievement in secondary education: Team teaching and interdisciplinary curriculum. *DAI, 61*(10A), 3933.

Thomas, L. M. (2002). Year-round schooling's impact on the Texas academic excellence indicators of achievement, attendance and high school completion in the South Texas region. *DAI, 63*(12A), 4269.

Thompson, E. O. (1992). Three methods of instruction in high school geometry and the effects they have on achievement, retention, and attitude. *Dissertation Abstracts International, 53*(8-A), Univ.

Thurber, T. (2001). The impact of the Academic Programs for Excellence on promoting academic achievement at Easton High School (Maryland). *DAI, 62*(02A), 451.

Torres, M. (1999). Effects of extending mathematics instruction through a Saturday academy to limited English-proficient Latino students in low performing schools. *DAI, 60*(08A), 2844.

Trlica, K. H. (1998). The effect of alternating block scheduling on middle school mathematics achievement growth. *DAI, 59*(08A), 2810.

Tymann, R. (2003). At risk students and state exams: An RX for student success. *DAI, 64*(09A), 3158.

Vahey, P., Enyedy, N., & Gifford, B. (2000). Learning probability through the use of a collaborative, inquiry-based simulation environment. *Journal of Interactive Learning Research, 11*(1), 51-84.

Vasquez, S. (1998). Effective instructional techniques for at-risk mathematics students. *DAI, 59*(06A), 1957.

Washington, J. (2000). Achievement and social-emotional outcomes of a multigraded, accelerated mathematics program. *DAI, 61*(02A), 539.

Watson, M. L. (1996). The impact of a new program for teaching algebra to increase achievement and persistence in mathematics. *DAI, 57*(09A), 3864.

Watt, J. X. (1991). The effects of an experimental spatial ability treatment program on calculus achievement with university students. *Dissertation Abstracts International, 51*(10-A), Univ.

Wear, S. B. (1997). The effect of single-sex mathematics classes on achievement and attitude for eighth-grade students. *DAI, 59*(01A), 0116.

Weidemann, W. J. (1990). Three methods of teaching locus of points problems in high school geometry. *DAI, 52*(01A), 0105.

West, R. C. (1993). The effects of an integrated learning system (ILS) using incremental time allotments on ninth-grade algebra achievement. *Dissertation Abstracts International, 53*(4-A), Univ.

Wohlgehagen, K. S. (1992). A comparison of the attitude and achievement in mathematics of Algebra I students using computer-based instruction and traditional instructional methods. *DAI, 53*(12A), 4243.

Yusuf, M. M. (1995). The effects of Logo-based instruction. *Journal of Educational Computing Research, 12*(4), 335-362.