

THE IMPACT OF STANDARDS-BASED MATHEMATICS CURRICULA IMPLEMENTED
HETEROGENEOUSLY ON HIGH SCHOOL STUDENT ACHIEVEMENT

By

Kari Krantz Selleck

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

K-12 Educational Administration

2012

ABSTRACT

THE IMPACT OF STANDARDS-BASED MATHEMATICS CURRICULA IMPLEMENTED HETEROGENEOUSLY ON HIGH SCHOOL STUDENT ACHIEVEMENT

By

Kari Krantz Selleck

This study examined the impact of standards-based mathematics curricula developed by the National Science Foundation, implemented within heterogeneously grouped, de-tracked high school classrooms. Four purposefully selected cohorts of high school students participated over a period of eight years. Outcome measures included two coursework measures (maximum difficulty level of math courses in which students enrolled and total number of math courses enrolled in during high school), and standardized state-level high school test results. Hierarchical regressions conducted on the sample as a whole showed no significant differences among the cohorts for the highest level of math course taken. The trends were that students in Cohort 2 (the first post-reform cohort) took slightly lower-level math courses than students in Cohort 1 (pre-reform), there was then a slight increase in Cohort 3, and finally, students in Cohort 4 took slightly higher-level math courses than students in Cohort 1. Regarding the number of courses taken, students in Cohorts 2 and 3 took fewer math courses than students in Cohort 1, and students in Cohort 4 took approximately the same number of math courses as students in Cohort 1. The results for Cohorts 2 and 3 were significant. There were negative, significant differences, although slight, for standardized tests. In other words, students post-reform performed slightly but significantly worse on standardized tests than students pre-reform. Further hierarchical regressions on the highest and two lowest-achieving quartiles (based on incoming eighth-grade state test results) showed that students at the highest proficiency level within the three post-reform cohorts fared slightly worse than those in the pre-reform cohort (-1.35, -.84, -.67) for

highest level math course with Cohort 2 significant (0.36). Highest achieving students performed worse on standardized tests with Cohort 3 and 4 significant (both at $p=.001$) Students in the lowest proficiency levels across all post-reform cohorts fared better than the pre-reform cohort in terms of level of math courses. Low-performing students in the fourth cohort (strongest treatment group) took math courses nearly two-thirds (.62) of a difficulty level higher than students in the pre-reform cohort, and this result was significant ($p = .010$). New state math course-taking requirements along with changes in content and scale scores of the state assessments during this longitudinal study posed limitations to the study. Implications for national and state mathematics policy included.

DEDICATION

To my mother and father, with love and gratitude.

ACKNOWLEDGEMENTS

This study represents the work and involvement of many different people across nearly a decade of my life. I am grateful for the assistance I received throughout the process and the collaborative spirit I encountered from so many helpful individuals. I thank each of you who contributed to the completion of this study.

First, I thank my parents, Karl and Sue Krantz, for their unending love and support from the moment I was born. I am fortunate to have been raised by parents who gave everything to help their children pursue their goals. My mother's drive and tenacity combined with my father's kindness and unwavering work ethic ensured that I would reach my goals regardless of how high I set them. Both of my parents modeled a sensitive spirit toward those less fortunate and I believe that my own passion to work with schools to ensure that equitable opportunities are available for the underprivileged is rooted in the generous acts of kindness I witnessed my parents perform.

Next, I thank my husband, Michael, who allowed our home to be transformed into a stack of papers for several years without even a word of frustration uttered. More than anyone, Michael realized the challenge of the project and always presented a plate of food or a signature "salty dog" at the most perfect time. He attended many family events without me there as his partner and never extended anything less than support. On many late-night occasions, I would hear from the bedroom, "Honey, just put a period on it." This was his way of saying it was time to give my mind a rest and come to bed. Again, he always knew when I needed to hear these words. Most importantly, I thank him for his unending pride in my accomplishments and the kindness he provided on a daily basis.

I want to also thank my nieces, Kennedy Lauren Krantz and Maddie Sue Krantz, for understanding when I needed to “work on my dissertation” and had to leave the party too early. I have loved having them see that writing is truly a process and have appreciated the many times they said, “Aunt Kari, you can do it!” Thank you, Kennedy, for the nice drawing of the four-leaf clover on the calendar marking August 16, 2012 as a very special day.

My friend, Kelly Smith, and I began our work to obtain our PhDs at the same time. There are no words appropriate to share the gratitude I have for Kelly’s friendship along with the hundreds of hours we spent riding back and forth to Saturday classes that first year along with many subsequent rides to and from evening courses over the next five years. I thank her for her strength and her dedication to improving the environment for all students in any capacity within which she has served. Her leadership was largely responsible for opening the doors of change at Midwest High School. I admire her many leadership accomplishments.

This study represents a reform effort implemented by a special group of administrators and teacher leaders who worked together for several years to improve the conditions for learning mathematics for all students, with particular work toward improving mathematics programs for those students historically marginalized in the education process. We often refer to ourselves as “the dream team” and it was the mutual admiration and genuine love we had for one another that helped us find the courage to challenge the order of things at Midwest. I am honored to have worked with this group of people and will never forget their ethics and their moral fiber, as well as their doing what was right for students versus doing what was the easiest for leaders. At critical times during the reform at Midwest, we all endured severe personal and professional criticism while doing the hard work of leading deep change, but there was a strong sense of pride and accomplishment in doing this work together. I thank John Smith, Superintendent of Midwest

School District; Frank Vyskocil, President of the Midwest Board of Education; Gary Bendall, Brenda Walworth, and John Mitosinka, also members of the Board of Education; Kelly Smith, Principal of Midwest High School, Tim Dowker, Barb Birchmeier, Steve Herrick, and Chris Curtiss, Assistant Principals at Midwest High School; Dave Moore and John Fattal, Principals of Midwest Middle School; Scott Moeller, Assistant Principal at Midwest Middle School; and Richard Ziegler, Sean McLaughlin, Andrea Tuttle, and Kathy Hurley, Elementary Principals at Midwest. An additional thank you must be extended to other important members of the “dream team” who helped organize and collect permission slips, course records, student incentives, and many other elements associated with the design of this study. These individuals include Mindy Dunn and Sherri Svrcek, Counselors at Midwest; Peggy Friess, Denise Grzibowsky, LeAnn Napier and Clara Crowe, Office Managers for several Midwest buildings.

I am grateful for the friendship and help I have received from Korrie Birchmeier, Midwest High School’s Mathematics Department Chair, who championed the reform during challenging times with peers and displayed patience and quiet resolve to successfully get things done like nobody else! Korrie remained helpful to this study during the birth of both of her children and always came through even on short notice.

Don Green, Math Specialist for Holt Public Schools, served as a great sounding board during the development of my research design. I am grateful for Don as a dear friend whose passion about mathematics and his own questions about student grouping practices and the role of intervention programs continue to shape my thinking. I will never forget the day when Don, along with Lisa Weiss, Science Specialist for Holt Public Schools, saved the day for me when I lost an important electronic file and they retyped a mathematics attitude survey for me, allowing me to meet a critical deadline during my pilot study.

I extend my heartfelt thanks to David Schulte, Superintendent of the Shiawassee Regional Education Service District. Dave worked closely with me to extract necessary student data from the county warehouse throughout the entire project. There were many, many late Friday nights and weekends when Dave made himself available to me. David's patience and his support for my study while managing far too many responsibilities of his own will never be forgotten. I can never repay Dave for his help.

I extend my deepest gratitude for the meticulous work done by my friend, Lisa Pilon, to merge and organize the hundreds of files for my study. Having a perfectionist oversee this aspect of my study gave me the ability to sleep at night. I'm not sure Lisa and I actually slept a full night during a few key stretches of the research and I have several emails to confirm our night owl propensities, but I know that Lisa's contributions to this project will never be forgotten.

Important support for my project came from a long-standing friend and mentor, Dr. Donald Weatherspoon. Don provided morale-boosting conversations related to his own doctoral journey that seemed to carry me through challenging times. He talked often about the life-changing nature of the completion of a doctorate and as I write this acknowledgment page, I realize how true his wisdom has proven to be. I appreciate the resources he gave me and his links to other experts who also ended up being invaluable for this project.

Two such contacts provided to me by Don were Wen Sun and Qiao Cui (Jay and Jenny), both experts in the field of statistics. Jay and Jenny opened their home in Dexter, Michigan to me and graciously assisted me in understanding many of the early statistical components of my study. Jay worked tirelessly to help me clean and merge data and Jenny helped me code and recode particular variables, while also helping me run many subsequent statistical analyses. I made several beautiful fall, winter, and spring drives to their home in Dexter and I am fortunate

to have had the privilege of working with and learning from Jay and Jenny during an important phase of my own learning.

I must thank my present superintendent, Dr. Andrea Tuttle, and my educational technology director, Samantha Lieberman, for understanding the challenge of the enormous amount of work necessary to complete a PhD. I have appreciated their never-ending support in helping me manage my work schedule and also for their encouragement. Samantha's technology skills have come to the rescue more than once along this journey. Andrea and Samantha are leaders who model life-long learning and I am proud to work with such fine role models. They are, indeed, dear friends.

A final thank you is extended to my dissertation committee: Dr. Steve Weiland, Dr. Michael Steele, and Dr. BetsAnn Smith. Each of them contributed to my growth and learning through coursework with and through the numerous discussions we had along the way. All are outstanding scholars at Michigan State University and I am honored to have had them as part of my mentorship team.

To Dr. Susan Printy, my patient doctoral advisor and chair: I appreciate her experience and wisdom. She opened her home to me and offered advice during important times during the course of my research. I am grateful for her immediate feedback and her pattern of follow-through, as well as her explicit help in providing me with many additional opportunities throughout my program. Susan pushed me when needed, let up when necessary, raised important points of concern as appropriate, and also extended a few necessary hugs. She has been a wonderful mentor.

TABLE OF CONTENTS

LIST OF TABLES	xiii
LIST OF FIGURES	xvi
CHAPTER 1: INTRODUCTION	1
Mathematics Reform – A Look Back	3
Standards-Based Reform	12
Academic Tracking or Ability Grouping	14
Curriculum Coherence	19
CHAPTER 2: A CASE OF SECONDARY MATHEMATICS REFORM AT THE LOCAL LEVEL – MIDWEST HIGH SCHOOL	21
Course Placement Practices	22
Curriculum Coherence	26
Mathematics Achievement	27
Reform Goals	28
Leadership	32
Summary	34
CHAPTER 3: RESEARCH MODEL, SAMPLE, AND METHODS	36
Sample	39
Context of the Mathematics Reform	40
Textbooks	40
Teacher Variances	43
Professional Development	46
Test Scores	50
Enrollment	51
Data Quality	54
Coding	58
Methods	65
Design	65
Descriptive Analyses	65
Analysis of Research Questions	66
Analyses of Lower and Highest Eighth-Grade MEAP Levels	67
Ethical Concerns	67
CHAPTER 4: RESULTS	68
Descriptive Analyses	69
Cohort	69
Student Characteristics	70
Gender	71
Free or Reduced-Price Lunch	71

Learning Disability	72
Residence	72
Summary	72
Means Analyses	72
Eighth-Grade MEAP Level.....	72
Outcome Variables.....	74
Gender.....	75
Free or Reduced-Price Lunch	77
Learning Disability	79
Residence	81
Cohort	83
Analysis of Research Questions.....	86
Correlations Among Predictor Variables.....	87
Research Question 1a: Mathematics Course Taking	89
Bivariate Correlations	89
Highest-Level Math Course Taken.....	90
Number of Math Courses Taken.....	94
Research Question 1b: Mathematics Achievement	96
Bivariate Correlations	96
Eleventh-Grade MEAP/MME	98
Eleventh-Grade ACT	101
Eleventh-Grade WorkKeys.....	105
Summary of Analyses for Research Question 1	109
Research Question 2a: Lowest Two Eighth-Grade MEAP Levels.....	109
Highest-Level Math Course Taken.....	110
Number of Math Courses Taken.....	112
Eleventh-Grade MEAP/MME	114
Summary of Analyses for Two Lowest Eighth-Grade MEAP Levels....	117
Research Question 2b: Highest Eighth-Grade MEAP Level.....	117
Highest-Level Math Course Taken.....	117
Number of Math Courses Taken.....	119
Eleventh-Grade MEAP/MME	121
Summary of Analyses for the Highest Eighth-Grade MEAP Level	124
Chapter Summary	124
Highest-Level Math Course Taken.....	124
Number of Math Courses Taken.....	125
Eleventh-Grade MEAP/MME	126
 CHAPTER 5: INTERPRETATION OF FINDINGS AND IMPLICATIONS FOR PRACTICE, POLICY, AND RESEARCH	 127
Summary of Descriptive Statistics Results.....	129
Discussion of Selected Descriptive Statistics Results	129
Middle School Results	132
Summary and Discussion of Research Question Results	133
Research Question 1	133
Research Question 2	136
General Conclusions	138

Limitations of Study	142
Data Collection	142
Mathematics Courses	143
Changes in State Assessments	145
Dependency on Standardized Assessment.....	147
Legislative Changes in State Mathematics Requirements	149
Confounding Variables	150
Summary of Limitations	150
Recommendations for Policy	150
Suggestions for Future Research	155
APPENDICES	157
Appendix A: Compiled Teacher Feedback After Year 1 Implementation	158
Appendix B: Slides from Presentation to Board of Education by the Researcher.....	164
Appendix C: Memo to John M. Smith, Superintendent of Midwest School District	173
Appendix D: 2004 News Article Included in Midwest Weekly Column in Local Newspaper.....	176
Appendix E: Letter from High School Principal to Staff, 2005	180
Appendix F: High School Mathematics Materials Review Rubric, 2003	186
Appendix G: Correspondence Concerning and Minutes of Meeting at MSU	187
REFERENCES	196

LIST OF TABLES

Table 1: A Timeline of National Mathematics Initiatives (Schoenfeld, 2004; Senk & Thompson, 2003; Wilson, 2004).....	7
Table 2: Course Placement Rubric	23
Table 3: Ninth-Grade Mathematics Enrollment Prior to Reform, 2003 (N = 186)	28
Table 4: Mathematics Program Material Review Process	30
Table 5: Textbooks and Program Materials Used in Grades 9-12 Prior to Reform, 2003-2004 ..	41
Table 6: Textbooks Used in Grades 9-12 After Five Years of Reform, 2008-2009.....	43
Table 7: Course Assignments 2003-2004 (No Reform)	45
Table 8: Midwest Students' Mathematics Achievement on State Test (Percent Proficient)	50
Table 9: Midwest Enrollment (Total Number of Students per Grade)	52
Table 10: Codes for Independent Variables.....	58
Table 11: Scale Scores, Ranges, and Performance Levels for Standardized Assessments	61
Table 12: Scale Scores, Ranges, and Performance Levels for the WorkKeys Standardized Assessment.....	62
Table 13: Math Course Codes.....	62
Table 14: Math Course Code Frame	63
Table 15: Distribution of Students by Cohort.....	70
Table 16: Demographic Characteristics of Participants, as Percentages by Cohort, and Pearson Chi-Square	71
Table 17: Eighth-Grade MEAP Level of Participants, as Percentages, by Cohort.....	73
Table 18: State-Wide Eighth-Grade MEAP Levels, by Cohort.....	74
Table 19: Means and Standard Deviations of Outcome Variables	75
Table 20: Means and Standard Deviations of Outcome Variables by Gender	76

Table 21: Means and Standard Deviations of Outcome Variables by Free or Reduced-Price Lunch	78
Table 22: Means and Standard Deviations of Outcome Variables by Learning Disability	80
Table 23: Means and Standard Deviations of Outcome Variables by Residence.....	82
Table 24: Means and Standard Deviations of Outcome Variables for Cohorts 1 through 4	84
Table 25: State Means for Eleventh-Grade MEAP/MME, By Cohort	86
Table 26: Correlations (Spearman's Rho) Among the Predictor Variables	88
Table 27: Correlations (Spearman's Rho) Between Coursework Variables and Demographics, Eighth-Grade MEAP Level, and Cohort.....	90
Table 28: Hierarchical Regression Analysis Summary for Demographic Variables, Eighth-Grade MEAP, and Cohort Predicting Highest-Level High School Math Course	91
Table 29: Hierarchical Regression Analysis Summary for Demographic Variables, Eighth-Grade MEAP, and Cohort Predicting Number of Math Courses Taken	94
Table 30: Correlations (Spearman's Rho) Between the Eleventh-Grade Standardized Test Result Variables and Demographics, Eighth-Grade MEAP Level, Coursework, and Cohort.....	97
Table 31: Hierarchical Regression Analysis Summary for Demographic Variables, 8th-Grade MEAP, Coursework through Grade 11, and Cohort Predicting Eleventh-Grade MEAP/MME Level	98
Table 32: Hierarchical Regression Analysis Summary for Demographic Variables, 8th-Grade MEAP, Coursework through Grade 11, and Cohort Predicting Eleventh-Grade ACT Level....	102
Table 33: Hierarchical Regression Analysis Summary for Demographic Variables, 8th-Grade MEAP, and Cohort Predicting Eleventh-Grade WorkKeys Level	106
Table 34: Demographic Characteristics of Participants, as Percentages by Eighth-Grade MEAP Level	110
Table 35: Hierarchical Regression Analysis Summary for Demographic Variables and Cohort Predicting Highest-Level High School Math Course Among Students at the Two Lowest Eighth-Grade MEAP Levels	111
Table 36: Hierarchical Regression Analysis Summary for Demographic Variables and Cohort Predicting Number of Math Courses Taken Among Students at the Two Lowest Eighth-Grade MEAP Levels.....	113
Table 37: Hierarchical Regression Analysis Summary for Demographic Variables and Cohort Predicting Eleventh-Grade MME/MEAP Level Among Students at the Two Lowest Eighth-Grade MEAP Levels	115

Table 38: Hierarchical Regression Analysis Summary for Demographic Variables and Cohort Predicting Highest-Level High School Math Course Among Students with Highest Eighth-Grade MEAP Level	118
Table 39: Hierarchical Regression Analysis Summary for Demographic Variables and Cohort Predicting Number of Math Courses Taken Among Students at the Highest Eighth-Grade MEAP Level	120
Table 40: Hierarchical Regression Analysis Summary for Demographic Variables, Coursework, and Cohort Predicting 11th-Grade MEAP/MME Level Among Students at the Highest Eighth-Grade MEAP Level.....	122

LIST OF FIGURES

Figure 1: Research Framework. The impact of heterogeneous classroom grouping structures using NSF-developed, standards-based curricula on achievement measured in various ways. ... 38

CHAPTER 1: INTRODUCTION

The need to improve achievement in mathematics in the United States has long been documented. National reports have urged widespread mathematics reform for K-16 schooling [McKnight, Crosswhite et al., 1987; National Commission on Mathematics and Science Teaching for the 21st Century, 2000 (TIMMS); National Council of Teachers of Mathematics (NCTM), 2000; National Research Council (NRC), 1989; Riley, 1988]. At the same time, national reports arguing for the elimination of academic tracking have proliferated for decades (Burris, 2009; NEA, 1990; Oakes, 2005; Schmidt, 2008; Silver, 1995; Welner & Burris, 2006; Wheelock, 1992). Despite these reports, too many secondary mathematics programs still utilize academic tracking practices to determine course placement for students and still provide instruction from commercially developed materials that support traditional pedagogy and the classroom structure (i.e., introduce algorithm or theory in lecture format, provide examples, allow in-class practice, give homework assignment, dismiss, repeat the following day starting with a review of homework) of decades past. Improving student achievement in mathematics at the secondary level continues to be one of the most important goals of local, state, and national education policies, but evidence of reform that coincides with the recommendations from these reports is limited. Not enough of these reports have provided supportive strategies to help school leaders and teachers combat the many barriers that are confronted during the implementation of secondary mathematics reform.

Many factors promote or limit efforts to change K-12 mathematics education practices. These complex factors are often difficult to disentangle, such that there are few existing models or prototypes for successful reform that can help practitioners to confidently and less confrontationally enact change. During the early 1990s, the National Science Foundation (NSF)

funded the development of new mathematics curricula that aligned with the content standards released by the National Council of Teachers of Mathematics (NCTM, 1999), intended to promote greater mathematical reasoning, problem-solving, mathematical connections between math topics, and accessibility for more students. The findings of studies that have examined the effect of NSF standards-based curricula on student achievement have been positive, but these findings have also been limited in a variety of ways: lack of data to determine fidelity to the intended curriculum, the impossibility of random sampling, lack of longitudinal data, and lack of information about the backgrounds of the secondary students in the studies (Batista, 1999; Le, Stecher et al., 2006; Schmidt, Huang, & Cogan, 2002; Senk & Thompson, 2003; Spillane & Zeuli, 1999). Documentation of secondary mathematics reform is limited. This study is unique in that it examines a middle school–high school reform sequence in the same district over time. Few studies have examined reform-oriented curricula and reform-oriented classroom grouping practices at the same time.

This study seeks to inform both research and practice by examining some of the elements of a local district’s secondary mathematics reform efforts. By looking closely at a case of local mathematics reform that was implemented over the course of approximately seven years at Midwest High School, I was able to compile information that will contribute to the somewhat limited current understanding of the issues at hand. In particular, this study provides helpful information for school leaders interested in improving secondary mathematics programs using inclusionary philosophies.

In this introduction, I begin by setting out a contextual basis for the reform effort by providing an account of some of the historic and national mathematic influences leading up to Midwest’s efforts. Understanding the historical context that surrounded the onset of Midwest’s

mathematics reform efforts is important as it should heighten the awareness of the social and political environment within which the change efforts were initiated.

I also discuss what is meant by *standards-based mathematics* and the curricula that developed as a result of the standards-based movement. Specifically, I present information about the mathematics curricula developed by the National Science Foundation, because one of the particular components of Midwest's reform included the adoption and implementation of K-12 NSF mathematics programs.

I then present a summary of relevant literature on the topic of tracking, as this longstanding practice is fundamentally related to some of the most important features of the reform efforts that occurred at Midwest High School. It is important to note that Midwest de-tracked its internal structure for stratification of secondary students in mathematics courses while simultaneously implementing NSF programs. One variable did not precede the other in this systems-change effort.

Finally, I conclude this introduction by describing the case of Midwest Schools' secondary mathematics improvement initiatives. This description will focus on the empirical details that have been documented to help situate the subsequent research questions outlined in this research proposal.

Mathematics Reform – A Look Back

A review of historical milestones in mathematics education and reform sets the stage for this research. Raising student achievement levels in mathematics has been a longstanding goal of educators in the United States. Data collected periodically over many years point to concerns that American students fare less well than peers from other countries. When secondary achievement data are compared internationally, American students appear to be less competent than others

[National Assessment of Educational Progress (NAEP), 1972; Second International Mathematics Study (SIMS, 1987, see Westbury & Travers, 1990); Trends in International Math and Science Study (TIMSS, see United States Department of Education), 1995]. Recommended improvements included, among other things, changing the emphasis of mathematics content, redefining purposes, examining the nature of mathematics pedagogy, releasing curriculum guidelines and general frameworks, clarifying outcomes, and developing common assessments. National recommendations like those cited above caused debate across education communities. These debates sparked a sociopolitical context that forced those involved in mathematics education to attend to the ramifications. In many cases, the accounts provided by key participants in mathematics reform efforts of the past shed light on the politically charged and often heated context surrounding these initiatives. The ongoing dispute between traditionalists and progressives served to distract many efforts (Schoenfeld, 2004; Wilson, 2003).

Although increasing levels of U.S. mathematics achievement has remained a consistent goal for national education policy, the recommended means for achieving this goal have not been uniform. Regarded by most as a necessary tool for social and economic development, mathematics education has political ties. Several notable publications have shaped ideas and initiated national responses in various forms. *A Nation at Risk* (National Commission on Excellence in Education, 1983), *Everybody Counts* (National Research Council, 1989), *Principles and Standards* (National Council of Teachers of Mathematics, 1989), followed by the release of the *Frameworks* (National Council of Teachers of Mathematics, 1999), and most recently, the unveiling of the *National Core Standards* (National Governors' Council, 2010) are all significant. These documents were released as a result of political action to rally large-scale efforts to improve mathematics achievement across the country. The release of these documents

led to a host of outcomes, including the development of new mathematics materials, state adoptions of particular programs, and, sometimes, public denouncements of particular efforts. One of the most notorious was the 1998 open letter submitted to Secretary of Education, Richard Riley, after the U.S. Department of Education released a list of recommended math programs and materials that emphasized conceptual-based mathematics over procedural-based mathematics. While the list included those programs that emphasized the ideals called for to bring about mathematics reform, nonetheless, over 200 university professors of mathematics signed the letter urging Riley's department to withdraw the recommendations.

Dr. Suzanne Wilson (2003) provided a detailed account of the political nature of the mathematics reform efforts initiated by the California State Board of Education during the 1980s and 1990s. Her account resembles a spy novel, often including primary sources that demonstrate the many covert and overt operatives working to dismantle these systemic efforts. Conservative groups worked to raise fear among parents and the wider community that standards-based mathematics curricula threatened foundational knowledge for students. In several California school districts, parents responded, as expected, by launching effectively organized protests against the mathematics reform.

Although somewhat simplified, it is not a stretch to say that the mathematical "camps" that formed were the result of extreme positions on both ends of the education philosophy spectrum. Traditionalists, holding firm to behaviorist assumptions (knowing facts, procedures, and conceptual understanding) opposed notions of constructivist assumptions (learning procedures and conceptual understanding while experiencing and constructing mathematics). Some have aptly referred to the contentious battles that ensued during this time in the history of education as "the math wars" (Schoenfeld, 2004).

When the history is reviewed, one can see a steady drive for improvement in mathematics achievement that became more inclusive of expanded segments of students. The late 1950s represented a period in which mathematics education targeted only a select few. The 1970s brought a charge for the development of challenging materials for both non-college and college-bound students, with the notion of separate pathways. The 1980s marked a time for educators to expand mathematical thinking abilities beyond computation, yet the concept of high levels of achievement in mathematics for all students did not really emerge until 1989 with the release of *Everybody Counts*. In this document, the nation's future economic development was discussed in the context of improving mathematics education for all students. With the onset of the 21st century, national policy began to attach stringent accountability measures with the call that all children would achieve higher levels of knowledge of and competency in mathematics (No Child Left Behind, 2001). Historically speaking, it is a relatively recent phenomenon to expect high levels of mathematics achievement for all students.

Cyclical patterns of change in philosophy and emphasis can be seen in the references presented in Table 1. An emphasis on basic arithmetic and procedural knowledge in the 1950s followed by “new math” in the 1960s, followed by the 1970s “back to basics” movement, which was followed by the “cognitive revolution” of the 1980s, which is now being followed by an era focusing on narrowly defined grade-level content expectations assessable through tests made up of multiple-choice questions. Although the NCTM's standards for mathematics instruction and evaluation released in 1999 articulated principles that attend to both conceptual and procedural understanding, a balanced outcome is atypical. In practice, one might describe the pattern as a back and forth swing between the procedural and the cognitive. The framework below includes highlights of important mathematics initiatives and milestones over several decades leading to

the present. Commercially developed mathematics program materials have largely followed the trends noted below.

Table 1: A Timeline of National Mathematics Initiatives (Schoenfeld, 2004; Senk & Thompson, 2003; Wilson, 2004)

	Initiative	Purpose	Trend
1890-1940s			High school for some Limited expectations for students to need mathematics knowledge beyond elementary school
1950	National Science Foundation (NSF) established by U.S. Congress	Development of national policies for math and science education	
1957 1960s 1964	Sputnik NSF First International Math and Science Study (FIMSS)	Launched by Soviet Union Funded curricula development to focus improved mathematics learning for the best and brightest Achievement scores demonstrated that U.S. secondary students fell below international averages.	“Modern Mathematics” (New Math) Parents and community did not understand “New Math”
1970		Focus on arithmetic computation and algebraic skills	“Back to Basics Movement” – Emphasis on skill practice with limited application

	Initiative	Purpose	Trend
1972	National Assessment of Education Progress (NAEP)	1st administration	Scores were low.
1977	National Council of Supervisors of Mathematics formed	Skills must be more broadly defined (inclusion of calculators and computers)	Call to develop challenging materials for both non-college and college-bound students (dual tracking system)
1980	<i>An Agenda for Action</i> released by the National Council of Teachers of Mathematics (non-Federal)	Call for students to think mathematically	“The Cognitive Revolution”
1981	National Commission on Excellence in Education formed		
1983 1984	<i>A Nation At Risk</i> <i>A Report on the Crisis in Mathematics and Science Education: What Can Be Done Now?</i> (American Association for the Advancement of Science)	Call for increased rigor	
1987	Second International Math and Science Study (SIMSS)	Results show U.S. students failing to score above international averages Report released based on findings, <i>The Underachieving Curriculum</i>	

	Initiative	Purpose	Trend
1989	<i>Everybody Counts</i> (National Research Council)	Call for excellence for all (equity).	“Mathematics for All”
	Curriculum and Evaluation Standards for School Mathematics (NCTM)	Released to encourage instruction and evaluation to promote mathematical reasoning, problem-solving, mathematical thinking	Mathematical Thinking/Reasoning
1991	Professional Standards for Teaching Mathematics (NCTM)		Engaged learning situations and greater emphasis on real-life applications to math
1994			
1990-1995	Curriculum Frameworks (NCTM) released	Rooted in Constructivist principles	
1995	NSF funds programs	Development of several standards-based instructional programs	Mathematical reasoning, mathematical thinking emphasized.
1995		Field testing and implementation begin	Procedural algorithms situated in meaningful context
1999		Released by NCTM	
	Assessment Standards for School Mathematics		Traditionalists emerged to induce a national anti-reform movement
	Third International Math and Science	Results show U.S. secondary students below international	“Fuzzy Math”

	Initiative	Purpose	Trend
	Study U.S. Department of Education	averages List of recommended mathematics texts released Open letter to Sec. of Ed. signed by over 200 university mathematicians urging withdrawal of recommendations	Controversy surrounds release of list
2001	No Child Left Behind National Policy (NCLB)	Increased accountability at the school level	
2006 2008 2009	Curriculum Focal Points for PreK-8 Mathematics: A Quest for Coherence (NCTM) Grade Level Content Expectations developed by each state (GLCEs) NCLB reauthorization window opens Response to Intervention (RTI)	Broadly defined concepts for focus Use of narrowly defined assessments, implementation of sanctions for failure to meet Adequate Yearly Progress measures Increase need/use of data and focus on targeted, research-based interventions What Works	Assessable content for Grades 3-8, 11 Proliferation of packaged intervention programs (procedural fluency) Back-to-basics trend

	Initiative	Purpose	Trend
2010		Clearinghouse	Scientific research based programs
		School-wide screening and progress monitoring	Tiered intervention programs
			Academic tracking practices regain momentum to provide avenue for interventions for targeted students
	“Race to the Top” (Federal Initiative for States to benefit under specific criteria)	Teacher evaluation tools linked to student growth	
	<i>National Core Standards</i> released (National Governors’ Council)	Improving standard alignment across states	Coherency of curriculum at national level
	Focus in High School Mathematics: Reasoning and Sense Making – Joint Statement (NCTM, NCSM, AMTE)	Increase rigor while improving thinking and reasoning	Higher levels of mathematics thinking

Leaders should be prepared to understand the politically charged debate that still surrounds many efforts to reform mathematics education. The potential for difficult backlash from teachers, parents, and community is present when trying to change traditional mathematics education. For example, havoc can erupt when strategic means are not in place to allow parents and the wider community to understand the purpose for reform and to have opportunities to gain ample knowledge and tools to be prepared to help their children with homework that most likely

looks very different from traditional examples. This means that leaders who want to implement district or school level curricular changes in mathematics should attend to the social and cultural context surrounding the change.

A logical position for leadership to assume to ensure that both traditionalist and reformist values are included may be a rational middle ground. Mathematics education efforts emphasizing both skills and processes appear to be a healthy position. The NCTM's (2000) document, *Principles and Standards for School Mathematics*, provided the nation with a middle ground. However, as schools implement national and state regulations associated with No Child Left Behind, the potential certainly exists for unintended consequences, such as renewed support and social justification for academic tracking.

In reviewing the events and the measures taken by experts across the field of mathematics as noted in the above outline, it is apparent that debate persists over the degree of emphasis that should be placed on skill instruction within the curriculum. Knowledge of the nation's history of mathematics education, including its political underpinnings, can help leaders proceed with informed caution while also remembering that parents, teachers, administrators, and students undoubtedly possess particular mental constructs shaped by their own education in mathematics. These constructs serve as cognitive lenses that may limit one's ability and willingness to accept or support change initiatives.

Standards-Based Reform

Following the release of the *Curriculum and Evaluation Standards for School Mathematics* by the National Council of Teachers of Mathematics (1989), the National Science Foundation (NSF) funded the development of several K-12 curricula with emphases on particular over-arching principles. These curricula aimed to help all students see the relevance of

mathematics, communicate mathematically, display confidence about their abilities in mathematics, and reason mathematically (NCTM, 1989). Importantly, standards-based mathematics curricula were designed to increase the content knowledge of teachers through accompanying professional development. Further, standards-based mathematics curricula were designed to include interactions of heterogeneous groupings of students. Standards-based classroom instruction emphasized students working together toward mathematical solutions derived from multiple ways of knowing and thinking. The classroom structures inherent in the curricula that were developed from the standards movement promoted small-group interactions, so that the task groups encountered situations that promoted mathematical thinking and reasoning (Stein, 2000).

Standards-based curricula developed through the National Science Foundation have been implemented in thousands of schools across the U.S. Still, studies that provide statistically significant evidence of benefit face traditional scrutiny surrounding educational research. One of the criticisms of NSF math curricula stems from the fear that if students representing prior mathematics achievement levels are instructed together, those at the lowest and highest ends of the learning spectrum will be shortchanged. However, the research does not support this widely held view.

Research conducted on the effects of NSF-developed curricula on student achievement has shown that, when compared with students receiving traditional instruction, students in these programs have made weak, but positive gains (Hamilton, McCaffrey, Stecher, Robyn, & Bugliari, 2003; Mayer, 1998). Evidence suggests that the measure used in testing whether or not reform-oriented teaching impacts student achievement is important. Therefore, multiple-choice outcome measures provide weaker (yet still positive) correlations to increased student

achievement than open-ended measures, when examining the impact of standards-based mathematics reform (Hamilton et al., 2003). When students that have been taught using reform-oriented curricula have been tested using measures that emphasize problem-solving and cognitive challenges, they have generally outperformed their peers (Senk & Thompson, 2003).

Academic Tracking or Ability Grouping

Academic tracking continues to be one of the most fundamental influences impacting achievement levels and access to higher levels of mathematics. Much research over the last few decades has demonstrated that tracking practices are entrenched in the practices of K-12 education. Persistent conclusions from numerous studies have demonstrated that, in most situations, ability grouping is neither effective nor equitable (Oakes & Guiton, 1995). In 1990, The National Education Association (NEA) was commissioned to study the practices of academic tracking in U.S. schools and found that the practice of tracking, or ability grouping, was nearly universal practice in the schools they studied (NEA, 1990). The NEA even put forth a compelling statement to address this practice:

What is clear is that rigid academic tracking creates academic problems for many students from all socioeconomic and ethnic groups and also creates student isolation by socioeconomic status and ethnicity. It is also clear that tracking is and is likely to continue to be a “way of life” in most American schools, including those that are socioeconomically and ethnically homogeneous. This situation can be altered, but tracking will not be replaced until practitioners and parents are confident that any replacement will contribute to a better school organization with a strong probability to higher student achievement (NEA, 1990).

Two decades after this statement was released, policy research organizations continue to develop guidelines urging education officials to examine the body of research available on curriculum stratification and adhere to schooling practices that utilize heterogeneous grouping with equitable access to a high-quality curriculum for all students. A December 2009 Legislation Policy Brief, *Universal Access to a Quality Education: Research and Recommendations for the Elimination of Curricular Stratification* (Burris, Welner, & Bezosa, 2009), presented case studies from three successful de-tracked schools that provide evidence that alternatives to academic and social tracking do exist. These case studies were situated within a review of decades of research supporting the elimination of academic stratification. The authors emphasized the complex nature of de-stratification and provided details to illuminate typical barriers to this kind of schooling reform. The authors discussed recommendations for schools and districts wanting to begin de-tracking efforts. A viable first step toward reform is the elimination of the lowest level of stratification (i.e., Basic Math, General Math, Algebra C) and including these students in heterogeneously grouped courses. A phased-in model of elimination was presented. This report concluded by outlining specific language policymakers could use to author state level policies that would advance a vision for abolishing curricular stratification and utilizing high-quality, equitable teaching methods and materials for all groups.

Tracking can mean different things depending upon the school or the district. Referent terms used to indicate stratification options include adjectives such as *general* or *advanced*. Others make reference to vocational and academic program tracks to ensure that a distinction is made between career readiness and college preparatory contexts (i.e., *vocational* or *college prep*). Other referents or descriptors refer to stratifications of ability, including words such as *remedial*, *low*, *middle*, and *high*. It is this final structure that appeared most closely aligned to the

actual arrangement of students, unlike the aforementioned studies in which the lines blurred between tracks for vocational and general in terms of how they related to actual ability level (Gamoran & Berends, 1987).

Studies include various reasons that schools make particular decisions about academic pathways. Based on participant observations and interviews with 19 teachers at a typical suburban high school, Finley (1984) was able to characterize the tracking decisions observed there as largely serving the staff, rather than as responses to the diverse needs of the students. Several excerpts from interviews demonstrated that decisions to place students in particular courses often had more to do with student motivation than true mathematics ability. This study showed that curricular decisions relative to the development of new course offerings and course teaching assignment preferences were often reflective of teachers' self-interests (Finley, 1984).

Local practices that determine which students will be placed into particular streams also vary. Teacher recommendations, test scores, prior course-taking records, and other criteria are frequently cited as being used to determine accurate placements. Elizabeth Useem, having examined policies and practices of 52 administrators and mathematics leaders across many districts, found that placement decisions and the criteria established for those decisions varied widely between different districts (Useem, 1992). It was not uncommon for a student who was filtered out of an accelerated or advanced math program in one district to qualify for placement in these very programs in another district (Useem, 1992). The patterns of policy and practice across the districts examined in this study were inconsistent. Notions of equity, or lack thereof, in placement decisions made within and across these districts were raised by the studies conducted by Useem (1992) and Finley (1984).

Efforts to replace ability grouping in mathematics programs require unpacking important and longstanding beliefs and values. Despite more recent professional acknowledgement that the discipline and content of mathematics is much more ill-structured than once regarded, aligning with nontraditional mathematics course structures, practice has been slow or reluctant to follow research findings (Horn, 2005). Mathematics is still viewed by many as linear, hierarchical, and based on acquired knowledge, rather than as a constructive, complex, divergent, and multifaceted subject. These underlying beliefs about learning math make it difficult for many to understand how students might learn mathematics within heterogeneous classroom structures outside the norm of traditional academic tracking.

Another belief that guides grouping and instructional practices is the notion that secondary curriculum accommodates pre-existing abilities rather than seeking to alter these abilities (Oakes, 1992). Student placement in tracks for mathematics often begins in elementary schools. If not at the elementary level, students are regularly placed in a particular stream for mathematics instruction beginning in middle school. Data from the National Educational Longitudinal Study (NELS), in which 1,000 schools provided information concerning approximately 25,000 students, showed that 86% of students were placed in some form of stratified learning in secondary school (Hattie, 2009).

The large percentage of schools utilizing some variation of academic stratification increases the likelihood of finding documented evidence that supports the continuation of ability grouping. Although the research supportive of ability grouping is certainly much less prolific than the research supportive of heterogeneous grouping, an examination of ethnographic and survey research on the impact of tracking on achievement revealed negative effects for the highest achieving students (Gamoran & Berends, 1987). Some researchers have found a high

track benefit that surpassed the lower track deficits (Hoffer, 1992) and others have found that when particular variables are controlled, the effects of academic tracking are unremarkable (Alexander & Cook, 1982).

Deeply held notions of achievement in mathematics as being an inheritable trait appear to contribute to this practice of early tracking. Such can be evidenced by professionals who laudably profess that they “didn’t inherit the math gene.” The irony is that few adults would dare to exclaim with such confidence that they “didn’t inherit the reading gene.” The absurdity of that statement raises important considerations about how teaching practices might differ if they reflected a belief that learning mathematics has as much to do with explicit teaching practices as with students’ inherited predispositions.

The subcommittee commissioned by the NEA (1990) to examine the effects of tracking made several suggestions to help de-track American schools. One suggestion was to begin making tracking changes at the elementary levels first. They also suggested that reducing class size might allow educators to meet diverse academic needs within individual classrooms. Cooperative learning was a classroom structure that the subcommittee recommended as an effective practice to use in conjunction with de-tracking. The commission maintained that if grouping practices were to persist, they should be flexible, address specific learning difficulties, maintain high expectations for all students, provide accountability to the system, and avoid negative stereotyping (NEA, 1990). Although the commission acknowledged that efforts to de-track would require substantial technical and political support, nonetheless the report remained limited in that it merely reiterated the problems surrounding tracking that were largely already known.

Some researchers have identified particular organizational and conceptual factors that have accompanied “full-fledged mathematics de-tracking” in particular schools. For example, these schools held beliefs about the subject of mathematics that moved away from traditional assumptions of mathematics as sequential and hierarchical. These schools shifted the focus of math instruction from breadth to depth and from procedural knowledge to rigor and relevance. These schools also provided clear direction and boundaries to teachers regarding curriculum, yet encouraged professional decision-making. Finally, successful de-tracked schools have challenged traditionally implemented practices involving homework, final exams, and other barriers to real opportunities to demonstrate mathematical learning (Alvarez & Meehan, 2006; Horn, 2005; Oakes, 2005; Wheelock, 1992). Not surprising, it appears that those de-tracking efforts that have succeeded have included many concurrent changes.

Curriculum Coherence

Inequalities in the materials and curriculum associated with mathematics instruction are widespread. American educational practices have been largely influenced by teacher autonomy (Lortie, 1975). In the context of mathematics and tracking, the notion of teacher autonomy as contributing to independent and individual variations in curriculum and instruction is further supported. For one district, tracking can mean that students are placed into slower-paced courses that use the same curriculum and materials as the faster-paced courses. For another district, tracking can mean that students in lower tracks receive a very different set of instructional materials than those in higher tracks. The term, "content tracking" (Cogan, Schmidt, & Wiley, 2001) refers to the uniquely American practice of providing alternative sets of curricula to different students, depending upon the course and path they follow. Thus, within the same district, students might be placed in a series of math classes that would use one particular set of

materials, while students placed in a different series of math classes could use another set of materials. Secondary courses often utilize many different programs and textbooks, which creates a lack of coherence within distinct tracks and also between them. It is not surprising that a patchwork of instructional practices and programs often exists both vertically and horizontally (Schmidt, 2008; Supovitz, 2006) within districts.

To compound the issue, math course titles are frequently found to not match the textbook being used as the primary resource for the course. In some cases, the textbook being used reflects content that would be more suitable for a higher- or lower-level math course. For example, in many situations, "Pre-Algebra" classes were using textbooks titled *Algebra*. In other situations, classes titled "Algebra" were using basic mathematics books. For nearly one-third of the representative sample used in Cogan's research examining instructional coherence, a mismatch existed between course title and textbook title. The findings demonstrate that there is little coherence in the existing mathematics curricula and instruction in the United States (Cogan et al., 2001).

Moving from the review of relevant literature to an overview of the case of Midwest High School allows readers to gain insights into the conditions leading up to this secondary mathematics reform effort, consider some of the underlying factors present during the reform, and better understand the variables included in the design of this study.

CHAPTER 2: A CASE OF SECONDARY MATHEMATICS REFORM AT THE LOCAL LEVEL – MIDWEST HIGH SCHOOL

Midwest High School provides the context for this study of local reform in which several fundamental changes were implemented in the existing program, so that more secondary students might reach higher levels of mathematics achievement.

The total district-wide population at Midwest is typically about 2,400 students. Approximately 600 students enroll in the Midwest District under school-of-choice, non-residence provisions. The majority of the non-resident students reside within the districts sharing a contiguous border to the west and east. Midwest Middle School typically enrolls approximately 600 6th-8th grade students. Midwest High School typically enrolls approximately 800 9th-12th grade students. Approximately 20% of these students are identified as non-residents. Parental surveys conducted yearly by the Midwest School District provide data which identify “traditional values,” “a strong academic core,” and “strong discipline” as some of the most important reasons why parents chose to send their children to the Midwest District. This expectation for *tradition* among parents and community members who choose Midwest as non-resident consumers presented an additional challenge to leaders during the reform. At times, they worried that moving from a traditional secondary mathematics program to one that involved changing student-grouping practices and mathematics pedagogy had the potential to negatively impact the large number of parents opting to enroll their children in the district from outside its boundaries. District leaders were mindful that unsuccessful implementation of a secondary mathematics reform could have negative financial implications (associated with any decline in enrollment). They were particularly concerned with making sure that students currently enrolled in Midwest

as school-of-choice remained and that potential school-of-choice students would continue to seek enrollment in Midwest.

School leaders at Midwest were compelled to initiate a comprehensive improvement of their existing K-12 mathematics program due to several significant conditions and factors within the organization: course placement practices, curriculum coherence, and mathematics achievement.

Course Placement Practices

The existing secondary mathematics program at Midwest reflected a traditional stratification of mathematics courses designed to meet the needs of students across ability levels. While K-5 elementary classrooms historically served students in self-contained, heterogeneous groups, tracking started in middle school (Grades 6-8) and continued throughout high school (Grades 9-12). As early as sixth grade, students were placed in tracks titled, “Low/Co-Taught Math” (lowest-achieving general education students combined with special education students taught by a general education teacher and a special education teacher), “Advanced Math” (students recommended by fifth-grade teachers as most able in math), or the “General” track (the remaining students in the low/average math achievement range). These three middle-school stratifications fed into a high school mathematics pathway with similar levels. Basic Math, General Math, Introduction to Algebra, and College Prep Algebra were options for incoming freshmen based on prior math performance ability and their middle school feeder track.

Each spring, counselors from Midwest’s Middle School talked to fifth-grade students about course offerings available to them in their first year of middle school. Midwest High School counselors visited eighth-grade students to describe the course offerings available to students as entering freshmen, referring to the Midwest High School Course Guide. Middle

school counselors and teachers worked in conjunction with Midwest High School staff to guide students' course-taking decisions. Prior to the mathematics reform begun in 2004, students were placed in various mathematics courses based on criteria established by school personnel. Table 2 is a sample rubric developed by middle school math teachers and counselors and given to fifth-grade elementary teachers each spring to assist them in making appropriate placement recommendations.

Table 2: Course Placement Rubric

Student Name	Current State Assessment Level (1, 2, 3, 4)	District Math Assessment Score (% correct)	Final Math Course Grade	Math Teacher Comments	Other Notes
--------------	--	---	-------------------------	-----------------------	-------------

Teachers completed the rubric and forwarded the information gathered to the middle school. Once math placements were made, written letters indicating the title of the math course the student would be taking the following year were hand-delivered or mailed to exiting students before summer vacation. Teachers frequently noted that this annual practice could prompt tears among some disappointed students not deemed worthy of gaining entrance into the top-tier math classes or cheers among students who were pleased with their placement. The following email response was given when a fifth-grade teacher was asked to provide a copy of the rubric she had to complete to recommend placement decisions for middle school. Although she refers to an English placement example, her disdain for the placement practices across all subjects is obvious.

If you are referring to the way we used to place kids in academically-talented classes, I'm sorry, but I have nothing to give you. I totally disliked the system anyway. So many times I found that my recommendations for those classes were not heeded. Once I went right

into the Middle School with copies of a student's writing to plead my case. I was also frustrated handing out those letters (they should be sent in the mail!), because some kids were crushed when they “didn't get invited to the party” when they thought they were capable. -Midwest Elementary fifth-grade teacher

Similar reflections have been elicited from eighth-grade teachers as they describe painful experiences associated with high school math placement letters. Even though a rubric was utilized to increase levels of fairness in placement decisions, the validity of placement decisions was frequently questioned by students, parents, and instructors. The social and emotional outcomes of the existing course placement decisions led Midwest leaders to unpack the layers of thought perpetuating this traditional practice.

For example, some parents who were disappointed that their child was not placed into “honors mathematics” courses started requesting to see the evidence that warranted these decisions. In some cases, the evidence provided to parents left doubt in their minds about the fairness of the process. It was also common for teachers to question the judgment of their own colleagues if a student was placed in a particular stratified course but later struggled. Requests for a math class change after the start of a new school year or new semester were initiated by any number of people. Parents, students, and teachers commonly made appeals to the principal or counselors that an incorrect placement had been made. Counselors were frustrated with mathematics placement changes after the onset of a new semester due to the scheduling challenges these late decisions created.

Administrators began asking difficult questions as a result of course-placement decisions. What criteria did fifth-grade teachers use to determine which students should be placed in lower-achieving, average-achieving, and highest-achieving groupings for placement in middle school

stratified math classes? What criteria did eighth-grade teachers use to determine which students would be placed in the lowest achieving, co-taught math sections in high school and which students would be placed in the advanced honors section? Were these placements based on course grades? Were these placements based on work habits? What percentage of students' grades was based on completion of homework or performance on exams? How confident were teachers and administrators that grades reflected consistent criteria across all schools and all teachers in the district? Importantly, do early challenges in learning mathematics equate to lifetime challenges in learning mathematics? When these questions were asked, various and inconsistent answers surfaced, such that Midwest leaders realized that longstanding organizational norms and practices surrounding tracking, which were perceived by well-intended educators to be appropriate, needed to be fundamentally reconsidered.

Adult conflicts were also related to the institutional practice of tracking. Staffing decisions made by school leaders at the end of each school year frequently created additional tensions that warranted further examination. Which teachers would be assigned to teach the math classes in which the lowest-achieving students were placed? Which teachers would be assigned to teach the honors sections? An unwritten, yet recognizable norm that had developed over time at Midwest was that the most veteran teachers earned the right to teach the most able students. Newer staff members were often assigned the classes in which the lowest-achieving students were placed. This created a leadership challenge for Midwest High School's new principal (hired in 2003), when she began assigning a few of the most experienced staff members to weaker classes. The concerns raised by staff and parents after some of these organizational decisions were made provided several opportunities for members of Midwest's learning community to participate in professional discussions about equitable access to quality teaching and instruction.

These discussions also opened the door to much-needed dialogue and learning about preconceived notions of course assignment entitlements attached to teacher seniority and other longstanding norms surrounding issues of students' equality of access to quality instruction.

Curriculum Coherence

At Midwest, stratified math courses based on perceived ability led to conflicts for students and teachers as described above. The stratification and variety of courses available led to other program quality issues as well. Teaching materials and textbooks were purchased to align with the various levels of need represented in different courses. Decisions were made to purchase materials for various stratifications based on an outdated curriculum review cycle that unintentionally fostered an institutional norm to review and purchase materials for subjects whether or not they were needed. The process was originally established to ensure that each subject area within the curriculum was reviewed in a timely manner. Unfortunately, this review cycle encouraged spending funds on new materials regardless of need or fit, since the next window for review would not occur for another six years. Because funds allocated to accompany this review cycle were frequently limited, a piecemeal set of math materials resulted. For example, during the math review cycle opportunity, the math department chair might request that a new edition of Holt Algebra 1 be purchased since the company had recently released a new version. The same department chair might also initiate a review of a new textbook for use in the lower-achieving algebra courses claiming that the existing textbook is “just too difficult” or “these books are in terrible condition.” In many cases, curriculum coherence was not a necessarily identified rationale for the purchase of mathematics materials. The apparent hodgepodge of content, the inconsistent methodology/pedagogy encouraged through the textbooks and teacher resources utilized, and the overall lack of an intentional design for

teaching complex mathematical concepts required the plans for reform to include attention to curriculum coherence.

Mathematics Achievement

Student performance in mathematics at Midwest was consistently near the state average, with scores on state assessments given in fourth and eighth grades consistently a bit above the state average and scores for eleventh grade students consistently a bit below the state average. A pattern of achievement decline in mathematics occurred as students progressed from elementary through high school.

Another important factor leading to Midwest's reform was the increasingly selective environment and culture that existed as one moved higher up the trajectory of mathematics courses. A small percentage of high school students advanced to the highest levels of math learning. Many students did not see themselves as having the necessary prerequisite skills to learn additional mathematics. Data collected periodically from interest, attitude, and perception surveys provided additional evidence that many Midwest students did not like math nor did they feel adequate in this domain. Moreover, since only two math credits were required for students prior to the graduating class of 2011 (math graduation requirements at the State level increased to four credits starting with the graduating class of 2011), students could opt to take only two classes in mathematics.

Based on prerequisite guidelines, sophomores could take a second course to fulfill the two credits of math required at the time from among the following courses: Algebra I, Algebra II, Intro to Geometry, Geometry, Technical Math, or Business Math. Junior and senior math courses included Pre-Calculus/Trigonometry and AP Calculus.

The entry-level mathematics course options available to ninth and tenth graders, in particular, perpetuated a traditional system which placed students into multiple levels. Table 3 presents the 2003 mathematics course enrollment of all ninth-graders prior to the reform. A leveled system is present.

Table 3: Ninth-Grade Mathematics Enrollment Prior to Reform, 2003 (N = 186)

Course Title	Total Ninth-Grade Students Enrolled
Basic Math	8, Lowest-achieving (special ed. students)
General Math	29, Low-achieving students (C, D, E students)
Intro Algebra	66, Average-achieving students (B students)
Geometry	2, High-achieving students (A, B students)
College Prep Algebra I	53, High-achieving students (A, B students)
College Prep Algebra II	19, Accelerated, highest-achieving students (A students)
Pre-Calc./Trigonometry	1, Accelerated, highest-achieving student (A students)

Note. Some students did not enroll in a math course as ninth-graders.

Reform Goals

Midwest leaders, including building administrators, central office administrators, and many teachers, understood the issues that had been surfacing surrounding mathematics course-placement decisions, curriculum incoherency, and secondary mathematics-achievement patterns. Goals were developed to begin a course of action to make improvements. These goals included:

- Develop a philosophy of mathematics instruction that emphasizes mathematical reasoning and thinking for all students. Pay explicit attention to procedural fluency, but increase rigor through mathematics connections and authentic, real-life applications.
- Implement a standards-based mathematics curriculum in grades K-12, using NSF-developed mathematics programs.
- Decrease the number of leveled mathematics course options for incoming sixth-graders beginning in the year 2004.

- Decrease the number of leveled mathematics course options for incoming ninth-graders beginning in the year 2004 using a phase-in process. Students previously placed in highest-achievement courses in sixth-eighth grades would continue in similar course arrangements through high school. Students previously placed in lowest-achievement courses would be mixed with students in average-achievement courses starting in ninth grade.
- Improve student attitudes and perceptions about mathematics.
- Improve student achievement in mathematics.
- Evaluate students using both standardized testing methods and assessments that evaluate mathematics thinking and reasoning.

Limited challenges existed for changing course placement practices for incoming sixth-graders. Fifth-grade teachers were accustomed to teaching heterogeneous groups of students and supported the change. Middle school teachers generally supported the change, with the exception of those teachers who were usually assigned to teach the students identified as highest-achieving. The concern for meeting the needs of those “most academically talented” within heterogeneous groupings was raised by a few teachers.

Due to the phased-in approach used for incoming ninth-graders, the level of concern raised by middle school teachers relating to highest achieving students was simply delayed for high school teachers. High school teachers’ concerns surfaced later in the reform process.

Concurrently, Midwest adopted new standards-based mathematics curricula developed by the National Science Foundation for all its K-12 math courses. The heterogeneous structure was supported by curricula and materials that were used across all courses. The selection of mathematics program materials was such that all stakeholders who would eventually use these

materials were expected to participate in an equally informed and thorough review process that guided the final consensus decisions made at each juncture. The fact that we made it a priority to involve all members who would eventually teach using the potential choices required additional time and resources. Table 4 below outlines the opportunities provided to stakeholders to be involved in the process leading to the implementation of the reform.

Table 4: Mathematics Program Material Review Process

Review Activity	Stakeholders Involved	Level of Involvement
Review of relevant research literature focusing on need for improvement in K-12 mathematics achievement	<p>All K-12 administrators</p> <p>All K-12 curriculum chairs (grade level and department leaders)</p> <p>All elementary teachers All middle-school math teachers All high-school math teachers</p> <p>Board of Education</p> <p>K-12 parents</p>	<p>All research literature</p> <p>Summary research pieces</p> <p>All research literature</p> <p>Summary research presentations provided during Board meetings</p> <p>Short reviews included in school newsletters, weekly community paper</p>
<p>Review of existing practices</p> <p>Curriculum audits conducted internally</p> <p>Achievement trends</p>	<p>Individual grade levels teams</p> <p>All stakeholders</p>	<p>Thorough analysis, deep involvement, full days for released time with substitute teacher coverage</p> <p>Deep analysis by teachers and administrators</p> <p>Summary data provided to</p>

Review Activity	Stakeholders Involved	Level of Involvement
		parents, Board
Review criteria established for new program analysis	All administrators, all teachers	Criteria shared with all
Selection of research-based programs to be reviewed	Administration, curriculum leaders	Programs selected for review minimally publicized in advance of review process
Program material review process (No piloting)	All administrators, all teachers	Selected programs for review (components reviewed equally and thoroughly by all) Extensive review conducted (e.g., all “Tables of contents” reviewed simultaneously for all programs)
Consensus decision making	All administrators, all teachers	Consensus model for decision making included all review participants Results shared with Board

Consensus was reached by each group. K-5 teachers agreed to implement *Math Trailblazers* (Bieler, 1997). Teachers of grades 6-8 reached a consensus to implement *Connected Mathematics* (Lappan, 1990) and teachers of grades 9-12 determined that *Math Connections* (Berlinghoff et al., 2000) would serve as the chosen NSF program for Midwest High School students.

An additional challenge to Midwest leaders wanting to increase curriculum coherence using NSF-developed materials was that a K-12 program was not available. NSF funding had been set up to support program development in categories of elementary, middle, and high school. This factor was considered carefully by Midwest leaders. Senior-level program trainers were contacted, and in the case of *Connected Mathematics*, program authors helped consider the question, “How can Midwest ensure curriculum coherence while adopting three separately

developed NSF programs?” Particular attention was paid to examinations of the content in transition years (i.e., fifth-grade *Math Trailblazers* to sixth-grade *Connected Mathematics* and eighth-grade *Connected Mathematics* to ninth-grade *Math Connections*). Leaders were confident, after consultation, that all NSF programs had been developed with the same philosophical foundations and goals, so with the exception of some slight content overlaps requiring teaching adjustments, the level of curriculum coherency would remain high.

Every possible effort was made to provide sustained professional development for all teachers and leaders over several years to support their learning. This support was targeted to the particular programs being implemented and included training by program experts on both content and pedagogy. Teachers were also given support to plan lessons with colleagues during intermittently scheduled release time for the first few years of implementation.

Internal checks and balances were implemented to encourage the faithful implementation of each mathematics program at Midwest. Self-reports, feedback sessions (see Appendix A), classroom observations by administrators, minimal access to former textbooks, an upfront and ongoing commitment to minimizing tendencies to use supplemental methods, and periodic reviews of common unit assessment data provided important information to district leaders regarding the degree of fidelity to the intended curriculum. This allowed leaders to provide safeguards, as needed, to keep the initiatives on track.

Leadership

It is unlikely that a change effort spanning nearly a decade could withstand the tribulations that accompany second-order change without the unified efforts of a strong leadership team. The role leadership played in the mathematics reform at Midwest cannot be

underscored; however, it is not the focus of this research. A snapshot of the kind of leadership team in place during 2003-2010 follows.

In general, the superintendent and members of the Board of Education were supportive of the reform efforts. The superintendent was skilled at keeping the Board of Education informed, yet keeping them far enough away from day-to-day operations to limit any tendencies toward micromanagement. Board members were included at various points during the change process and promoted public confidence when formal presentations were given at board meetings (see Appendix B and Appendix C).

I held the position of curriculum director from 2002-2010. This role included the oversight of all curriculum, instruction, assessment, and professional development efforts from preschool through grade 12 for the Midwest district. I worked closely with all leaders during the reform and facilitated regularly scheduled meetings with building principals. I tried to connect communication about curriculum and instruction between all levels of the organization. Careful attention was given to communicating with the public at large, including the utilization of Midwest's weekly column in the local newspaper reaching the entire county (see Appendix D).

The high school principal served as a champion of change throughout the process, having assumed the responsibilities for leading Midwest High School in 2004 after serving as principal of Midwest's largest elementary school for the previous four years. On several occasions, Midwest's principal voiced her stance about the mathematics reform. During a critical point in the reform, she composed a letter to her staff that included her thoughts on the issue. This letter highlights her level of involvement (see Appendix E).

Two assistant high school principals shared the responsibilities for instructional leadership with the principal. During the seven-year period of math reform, there were

intermittent changes in assistant principals at Midwest High School. The individuals hired joined the efforts in a positive manner. Even though the assistant principals also assumed defined roles in athletics and discipline in addition to their role as instructional leaders, they were actively involved in working to support the mathematics reform process.

Midwest Middle School was led by a building principal and an assistant principal who worked cooperatively with Midwest High School's leadership team. The three elementary buildings had leadership changes during the years of the math reform and, like the new members of the Midwest High School administrative team, these elementary leaders merged well with the direction of the district.

A fairly positive picture of the leadership team in the district of Midwest emerges. Of course, all was not perfect. Interpersonal dynamics were evident between Midwest leaders during this time. Strong personalities and competition between key individuals created, at times, distractions to progress. As expected, when changes in Midwest leadership resulted in new members of the leadership team, novice status on occasion caused some slight missteps in the reform. Yet, for the most part, the leadership team at Midwest can best be described as cohesive and positively involved in the change effort.

Summary

What can we learn from Midwest? The case appears to represent improvements that have long been encouraged by national mathematics organizations. De-tracking efforts appear to have created a more equitable opportunity for more students to reach higher levels of mathematics. The NSF curricula being implemented appear to meet the criteria for exemplary mathematics programs, as called for by organizations such as the National Council of Teachers of Mathematics. Teachers and leaders appear to have been given ongoing support to learn how to

teach these curricula and the district appears to have stayed the course for a significant period of time. But, appearances must be further examined using scientific research methods. Several significant questions remain unanswered. Did students have improved access to a rigorous mathematics program as a result of the reform? Did the mathematics curriculum at Midwest become more coherent as a result of the reform? Did student achievement improve as a result of the reform? Did students who otherwise would not have taken more than two credits of mathematics prior to reform now choose to do so?

The case described in this chapter provides the context for my research study. The data collected were purposively selected and will be further described in the subsequent chapter, along with the measures I developed and the analytical approach I subsequently used.

CHAPTER 3: RESEARCH MODEL, SAMPLE, AND METHODS

This chapter presents the framework that I used to examine the impact of implementing new norms of practice for grouping students in mathematics courses using heterogeneous methods while also implementing standards-based mathematics curricula developed by the National Science Foundation. This chapter is made up of several sections. First, I provide a description of the research study, articulating the independent and dependent variables. Next, I outline the particular research questions that guided my inquiry and describe the sample and the data collected. The various measures that were developed in this investigation are carefully explained and a summary of the methods and analytic approach used is presented in the final portion of this chapter.

The efforts made by leaders and teachers in Midwest Public Schools to reform its secondary mathematics program are worthy of scientific study for several significant reasons. First, few studies have examined the impact of implementing an NSF mathematics program at the high school level in a district in which students are heterogeneously grouped for core instruction in Algebra 1, Geometry, and Algebra II. Secondly, few studies have included a multi-year analysis of a treatment within such a setting. Another important issue in this study is the strengthened middle school to high school sequence of reform-oriented instruction represented in Cohort 4 when compared to Cohort 3. Many studies have been conducted to examine the relationship of specific variables (i.e., particular instructional practices in mathematics) to student achievement, comparing effects on pre/post cohorts after one year of implementation.

This study will explore the impact of interventions conducted with four cohorts of students over a period of nine years.¹

¹ The four cohorts of students do not represent consecutive years. Students in Cohort 1 were ninth-graders in 2003 while students in Cohort 2 were ninth-graders in 2005. Students in Cohort 3 were ninth-graders in 2006 and students in Cohort 4 were ninth-graders in 2008.

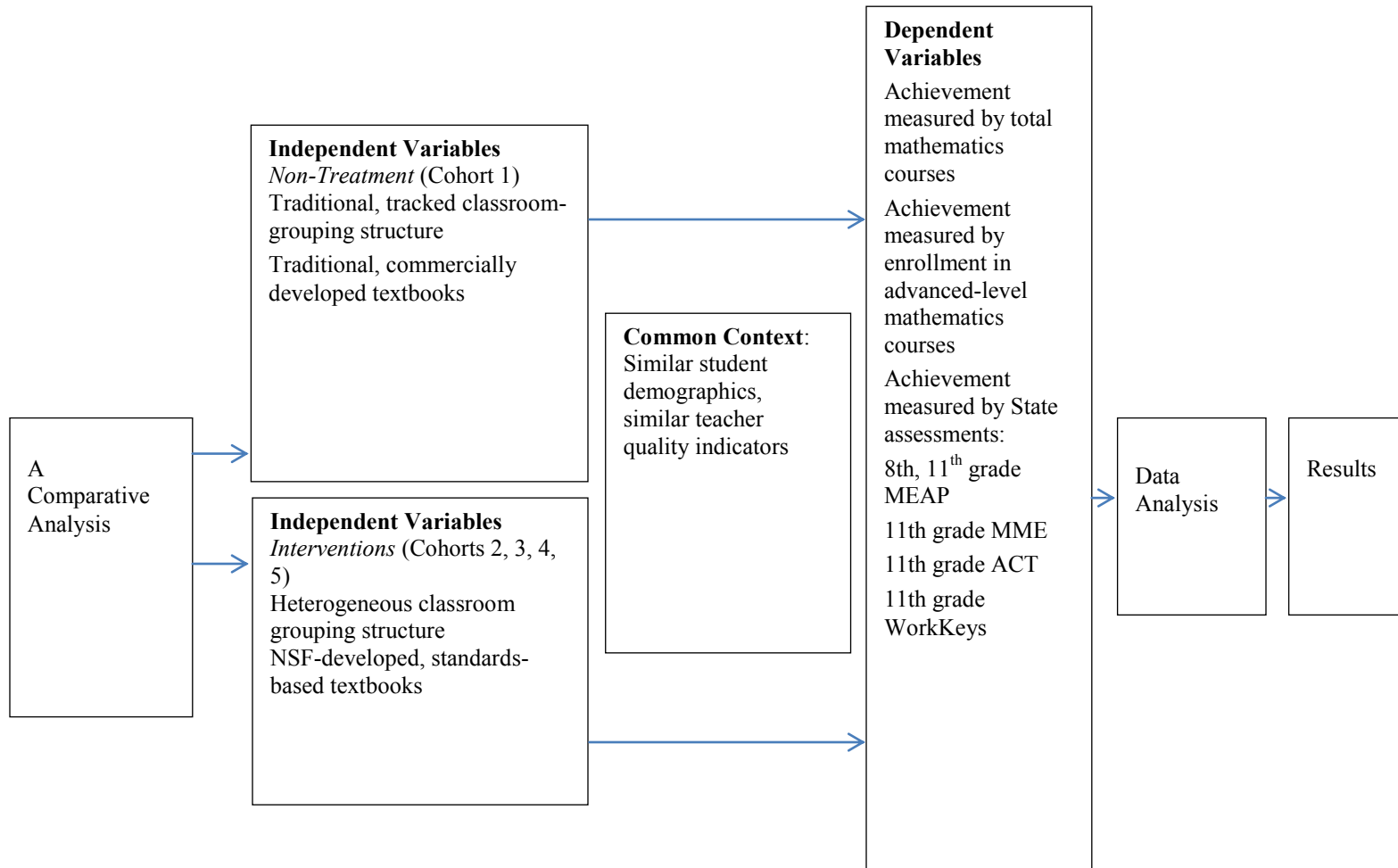


Figure 1: Research Framework. The impact of heterogeneous classroom grouping structures using NSF-developed, standards-based curricula on achievement measured in various ways.

Midwest High School implemented two purposeful interventions designed to 1) de-track students' placement into mathematics courses and 2) use National Science Foundation-developed, standards-based mathematics curricula. The following two research questions guided the study:

1. How do mathematics outcomes of Midwest High School students compare pre- and post-reform?

- 1a. How does the mathematics course taking of students compare pre- and post-reform?

- 1b. How does mathematics achievement compare pre- and post-reform?

2. How do different types of students at Midwest High School fare in the reform?

- 2a. How do students whose prior math achievement in 8th grade indicates lowest proficiency fare when compared pre- and post-reform?

- 2b. How do students whose prior math achievement in 8th grade indicates highest proficiency fare when compared pre- and post-reform?

Sample

The target sample consisted of students enrolled at Midwest High School from 2003-2012. Midwest is a rural district serving a community covering 90 square miles with a mix of manufacturing and large and small agri-business influences. Based on the 2000 census, the mean family income was \$41,000 with approximately 52% of residents owning the home in which they lived. The mean home value in the Midwest district was \$101,000, which was slightly below the national average. Midwest School District is served through the local education service district in the county along with seven other school districts.

Context of the Mathematics Reform

The total district-wide population at Midwest was about 2,400 students. Approximately 600 students were enrolled in the Midwest District under school-of-choice, non-residence provisions. The majority of the non-resident students reside within the districts sharing a contiguous border to the west and east. Midwest High School typically enrolls a total of approximately 800 students (9th -12th grade). Approximately 20% of these students are identified as school-of-choice students (non-residents).

Textbooks

Although the scope of this study does not warrant an in-depth analysis of the content and teaching methods used by specific teachers across a seven-year period, the table below provides a broad snapshot of the program materials used during the 2003 year (before treatment) as compared with those used during the 2008 year (the final cohort). As noted earlier, one criticism of mathematics education in the United States is that it resembles a “patchwork quilt” (Schmidt, 2008). In other words, students may or may not receive an aligned or cohesive mathematics curricula or program, because the choice of materials for instructional use is regularly made at the individual-school level and, often, by the individual teacher.

Course information including the teachers assigned to specific courses and the titles, publishing company, and dates of publication of the materials used in each course was collected from archival course syllabi and course descriptions. Research has found that more than teachers of any other subject matter, mathematics teachers rely most heavily on the textbook as the source for content taught (Schmidt, 2008). I made a

decision to presume, generally, that the teachers at Midwest followed this trend. These data were compiled and are displayed in Table 5.

Of note are the highlighted areas of the table, where it is evident that an assortment of materials was used for some of the less advanced courses. Table 5 also shows that the materials used in the courses titled (Intro) Algebra and College Prep Algebra I were not the same. Even without making any determination of the quality of the materials listed, it can be concluded that 4 sections of students enrolled in (Intro) Algebra were taught using a combination of multiple resources while 11 sections of students enrolled in College Prep Algebra I were taught using a single textbook. Students in these courses were taught with very different curricula but could, in fact, be enrolled in the same Algebra II course later in their schooling. It would be difficult to distinguish underlying factors leading to any differences in achievement in Algebra II given this present situation.

Table 5: Textbooks and Program Materials Used in Grades 9-12 Prior to Reform, 2003-2004

Course Title	Textbook	Publisher	Date of Publication	Sections or Classes with Assigned Materials
General Math	Various chapters/units from multiple resources	Multiple	Multiple	2 sections
Business Math	<i>Mathematics with Business Applications</i>	Glencoe	1986/1998	2 sections
Tech Math I	Teacher-developed Packets	None on record	None on record	1 section
(Intro) Algebra	Various	Multiple	Multiple	4 sections

Course Title	Textbook	Publisher	Date of Publication	Sections or Classes with Assigned Materials
	chapters/units from multiple resources			
College Prep Algebra I	<i>Algebra I</i>	Houghton Mifflin	1989	7 sections
Geometry	<i>Basic Geometry</i>	Houghton Mifflin	1990	2 sections
College Prep Geometry	<i>Geometry</i>	Holt	1986	2 sections
College Prep Algebra II	<i>Algebra II</i>	Houghton Mifflin	1990	2 sections
Trig./Pre-Calc.	<i>Advanced Mathematical Concepts Pre-Calculus with Applications</i>	Glencoe	1999	1 section
AP Calculus	<i>Calculus: Graphical, Numerical, Algebraic</i>	Addison Wesley	1995	1 section

Table 6 illustrates that after five years of the reform, curriculum coherence had been improved. Narrowing the variances across common course offerings (i.e., Algebra I and College Prep Algebra I no longer two separate courses, General Math and Technical Math eliminated) increased the likelihood that students would receive similar content and materials regardless of the course they took or the teacher they had.

Table 6: Textbooks Used in Grades 9-12 After Five Years of Reform, 2008-2009

Course Title	Textbook	Publisher	Date of Publication	Level of Use by Teachers
Business Math	<i>Mathematics with Business Applications</i>	Glencoe	1998	1 section
Algebra I	<i>Math Connections I</i>	It's About Time	2000	9 sections
Geometry	<i>Math Connections II</i>	It's About Time	2006	6 sections
Algebra II	<i>Math Connections III</i>	It's About Time	2006	7 sections
Trigonometry	<i>Algebra and Trigonometry</i>	Thomson	2007	5 sections
Pre-Calculus	<i>Advanced Mathematical Concepts Pre-Calculus with Applications</i>	Glencoe	1999	3 sections
AP Calculus AB	<i>Calculus: Graphical/Numerical/Algebraic</i>	Pearson	2007	1 section
AP Statistics	<i>Statistics Modeling the World</i>	Pearson	2007	2 sections

It is important to note that although I claim that coherency improved as a result of these changes, this dissertation is a general comparison of Midwest's curriculum pre/post reform. The scope of this paper does not attempt to examine the literature surrounding issues of coherency between state-level content standards and course content nor the questions that can be raised about differences in the intended and the enacted curriculum or the lack of agreement among states and publishers as to what math content should be taught in particular grade levels (Schmidt, Houang, & Cogan, 2002).

Teacher Variances

At the onset of my research study, I aimed to identify particular factors that could be included in a reliable "teacher-quality index." I gathered information about the education levels of the instructors, including whether they had completed an MS degree

or only a BS degree and whether or not they had completed a major or minor in mathematics. I also recorded the number of years each teacher had been teaching mathematics, and the number of years spent teaching a particular course. After further consideration, I determined that questions could be raised as to the reliability of this teacher-quality index. For example, one could argue that a teacher recently graduated from a college recognized for training its teachers using state-of-the-art math methods while also requiring an extra year of internship should receive a higher mark on the teacher-quality index than a teacher who graduated from a smaller, unknown teacher education program. Someone else might argue that the veteran teacher with a science major with 20 years of experience teaching mathematics possesses a stronger content knowledge and deeper understanding of the development of adolescents than someone with a math major and only three years of teaching experience.

I examined other considerations surrounding teacher variances over the course of this study. For example, I looked at the number of first-year teachers who were assigned to teach particular mathematics classes each year. I did this to determine whether Midwest followed the often-heard criticism that students most at-risk (i.e., those enrolled in the least advanced math classes) were frequently taught by the least experienced teachers. There was evidence to suggest that Midwest followed this trend. During the year prior to the implementation of the reform efforts, veteran teachers held all positions for math courses above the level of Basic Geometry (e.g., College Prep Geometry, Algebra II, Trig./Pre-Calculus, AP Calculus). Two mathematics teachers with three or fewer years of experience teaching at Midwest were assigned to teach General Math and Business Math. Of note, although the reasons are unknown, is the difference that existed

between the average size of the College Prep Algebra sections taught by the teacher with three years of experience (28.6) and the average size of the classes taught by the teacher with 30+ years of experience (22.5). These findings are illustrated in Table 7.

Table 7: Course Assignments 2003-2004 (No Reform)

Course Title	Teacher	Total # of Sections Assigned to Teacher	Avg. # of Students per Section	Years Teaching at Midwest
General Math	Teacher A	2	17	3
Business Math	Teacher B	2	22	1
Technical Math	Teacher C	1	15	30+
Algebra	Teacher C	4	24	30+
College Prep Alg. I	Teacher A	3	28.6	3
	Teacher D	4	22.5	30+
Geometry	Teacher B	2	22.5	1
Col. Prep Geo.	Teacher F	2	27.5	27
Col. Prep Alg. II	Teacher E	2	24	30+
Trig./Pre-Calc.	Teacher F	1	32	27
AP Calculus	Teacher D	1	14	30+

Several situations occurred that impacted teacher assignments across the timeframe of the study. Economic factors caused Midwest to offer an early retirement incentive at the conclusion of the 2004-2005 school year. Some veteran mathematics teachers were then replaced by new teachers. Natural attrition during subsequent years, combined with the loss of teachers from the early retirement incentive, resulted in at least one first-year teacher teaching mathematics each year. While the teachers hired during the reform entered prepared better to implement the kind of mathematics instruction desired, they were still early-career teachers. Long-term substitute teachers were also used during the reform to replace those mathematics teachers who were away from the

classroom for other reasons (e.g., maternity leave, medical leave). Midwest's cooperation with nearby universities also resulted in the placement of student teachers for short- and long-term practicums or internships that contributed further variance in overall teaching quality. In summary, I argue that Midwest High School experienced many outside influences affecting teacher assignments and ultimately "teacher quality".

It is impossible for me to determine whether these staffing changes were representative of most schools, but I argue that the changes in staffing were a continued source of concern and subsequent attention during the reform. Those staffing changes related most directly to the mathematics reform were intentionally monitored and explicitly supported, although staffing was certainly not held constant during the reform.

Professional Development

Few records were available to describe the professional development provided to mathematics teachers prior to 2003. Beginning in 2003, mathematics teachers² of grades 6-12 at Midwest participated in professional development that can be characterized as inclusive, ongoing, sustained, and responsive. All mathematics teachers were expected to participate in all aspects of the entire professional development plan. The year prior to implementation involved several full days of release time for mathematics teachers to read, dialogue, review existing data, and pose questions about the norms of practice at Midwest. Teachers of all grades were included to ensure that the mathematics program improvements would reflect a systems approach.

² K-5 elementary teachers were also involved in a similar effort with related professional development focused on the adoption and implementation of an NSF mathematics program (*Math Trailblazers*, Pearson, 2004).

Once a clear and compelling need for change was realized by the majority of mathematics teachers and administrators, programs were selected through careful review processes that were also inclusive. The review of materials included a collaborative process for identifying criteria to be used during the review process (see Appendix F). Teachers continued to be released from their daily classroom assignments when various presentations were made by program authors or representatives. Participation availability was carefully considered in advance of scheduling professional development sessions during or after regular school hours. Coaching commitments of key mathematics teachers frequently required alternative arrangements to be made for professional development. Stipends were paid to teachers to attend sessions outside the regular school day. Steadfast attention was paid to ensure that all teachers attended important professional development opportunities.

Once the foundation was laid for the implementation of new mathematics programs, training sessions to be conducted by program trainers were scheduled every 6-8 weeks, with planning time for various groups of teachers to meet for half and full days away from their classrooms. The professional development plan was designed to include time for specific program-related training conducted by experts in mathematics content and pedagogy related to the particular NSF mathematics program. These official training sessions were followed by time for lesson design and planning in collaborative teams, as well as time for feedback and support between training sessions. This model remained flexible and responsive to teacher needs. For example, in Year Three of the reform, teachers raised several questions about the newly de-tracked classroom structure. Teachers were compelled to move forward in lessons to maintain the kind of pace

suggested by the program, but were concerned that some students were not ready for the next concepts. Midwest teachers and administrators traveled to a local university to seek answers from one of the authors of the NSF program (see Appendix G). Although the professional development plan for this year was outlined in advance, an additional day of release time for teachers was worked into the plan for this university visit, in response to the immediate need of the teachers. Subsequent additions and revisions to the general professional development plan were made as a result of both informal and formal feedback gathered throughout the reform process.

The focus on the professional development during the reform at Midwest was meant to bring about significant changes in deeply rooted, traditional practices. This required a different approach to professional development. The kind of professional development provided during Midwest's reform is described by Dr. Mary Kennedy, in her monograph, *Form and Substance in Inservice Teacher Education* (Kennedy, 1998). She provides a model that categorizes teacher in-services (professional development) across four levels. Her research study documented an increase in effect size in student achievement that correlated to the hierarchical structure of her professional-development model. Category 1 yielded an effect size of .14 on basic skills and .10 on problem-solving, Category 2 had an effect size of .17 on basic skills and .05 on problem-solving, Category 3 had an effect size of .13 on basic skills and .50 on problem-solving, and Category 4 had an effect size of .52 on basic skills and .40 on problem-solving.

The first level categorizes professional development described as generalizable to any content (e.g., using wait time to increase student engagement). The second level includes general strategies that align with specific content. In the case of mathematics, an

example of Category 2 would be training that provided “strategies to increase basic math fact fluency.” Neither Category 1 nor 2 fit the kind of professional development provided to teachers during Midwest’s mathematics reform.

Categories 3 and 4 describe the kind of training provided at Midwest during the mathematics reform. Category 3 includes training that draws from a specific theory (e.g., a function-based approach to teaching mathematics) to connect to specific strategies within a particular program or curriculum (e.g., *Math Connections*, 2000). Midwest teachers read several research pieces that emphasized the importance of providing students with opportunities to explore important math concepts prior to learning algorithms. The units of study and the lessons within the *Math Connections* program were organized around this theory.

Category 4 in Kennedy’s model describes programs that allow teachers opportunities to examine specific content while considering the kinds of misconceptions students may or may not have while learning the content. Over the years of Midwest’s reform, teachers were given time and support to engage in this category of in-service activity. The discussions among math teachers frequently focused on student understanding of particular mathematics content within the new programs and peers would often draw from actual classroom experiences or student work to seek clarification from others or suggest a different way of reaching students. Category 3 and 4 training fostered ongoing, sustained dialogue around critical mathematics ideals and linked these ideals to particular instructional practices that generated particular student learning. This recurring process was maintained as a priority for several years at Midwest.

Test Scores

In the years before the implementation of the de-tracked, NSF-approved mathematics curriculum, state test scores for Midwest followed trends such that elementary achievement in math started out relatively high and declined consistently as students moved through middle and high school. Though the Midwest district matched the state trend, elementary math achievement scores were typically above the state averages. Middle school math achievement scores remained above the state averages, but were lower than those observed at the elementary level. At the high school level, scores were significantly below those at the elementary level, and showed an inconsistent pattern, sometimes below and sometimes above middle school and state levels. Students at Midwest High School scored much lower in achievement on state tests than elementary and middle school students. Table 8 illustrates this trend in math achievement prior to the mathematics reform.

Table 8: Midwest Students' Mathematics Achievement on State Test (Percent Proficient)

Grade 4	2001	2002	2003
Midwest	78.0	76.0	76.0
State	72.0	65.0	65.0
Grade 8	2001	2002	2003
Midwest	73.0	65.0	57.0
State	63.0	53.0	52.0
Grade 11	2001	2002	2003
Midwest	44.0	62.0	62.0
State	68.0	67.0	60.0

Enrollment

Table 9 outlines Midwest High School enrollment during the period of this study. Twelfth-grade enrollment patterns show losses from ninth grade with a fairly consistent graduation rate of approximately 85%.

Table 9: Midwest Enrollment (Total Number of Students per Grade)

Year	Grade 9	Grade 10	Grade 11	Grade 12
2003	221 Cohort 1 (start)	200	166	161
2004	208	215	193	167
2005	230 Cohort 2 (start)	211	204	188
2006	229 Cohort 3 (start)	222	197	199 Cohort 1 (end)
2007	199	232	221	183
2008	248 Cohort 4 (start)	190	218	218 Cohort 2 (end)
2009	199 ^a	243	180	200 Cohort 3 (end)
2010	223	190	228	177
2011	178	213	171	216 Cohort 4 (end)
2012				

^a Cohort 5, which was ultimately not used in the study.

From the population described above, four cohorts were purposefully selected for this study. **Cohort 1** represented students who received no treatment beyond the traditional mathematics program for middle school and high school and who graduated in the spring of 2006. Cohort 1 ($n = 199$) entered Midwest High School as ninth-graders in 2003, one year prior to the start of the comprehensive secondary mathematics reform and, during the 2003-2004 school year, received mathematics instruction that can be categorized as traditional. Students were assigned to various classes based on prior

mathematics achievement. Materials used during the four-year high school sequence for students in Cohort 1 can be described as commercially developed.

Cohort 2 was the first group of students involved in any intervention, or mathematics reform. Cohort 2 ($n = 201$) consisted of students who entered as ninth-graders during Year Two of the secondary reform effort in 2005 and experienced at least two years of heterogeneously grouped³ high school math classes that can be categorized as reform-oriented, or standards-based. Course materials used were from a National Science Foundation funded program (NSF). These interventions were concentrated at the high school level only. The middle school experience for these students still reflected the traditional trajectory and conventional materials were used for Grades 6-8. This group is of particular importance because students were not yet affected by new state legislation passed in 2006⁴ that required high school students to earn a minimum of four mathematics credits, as opposed to the formerly required minimum of two mathematics credits. In other words, there was no external pressure for students in this cohort to enroll and pass more than two classes of mathematics.

Cohort 3 represents students who entered as ninth-graders during Year Three of the secondary reform effort in 2006 and were taught from a reform-oriented, standards-

³ Two sections of students (approximately 52 students) formerly stratified in Grades 6 through 8 in honors math classes remained stratified during ninth grade while stratified grouping was phased out at the middle school one year at a time. Twelve students representing the most cognitively challenged were also stratified into a basic math class. The remaining 158 students were taught in mixed-ability classrooms as ninth and tenth graders for 2004-2005. Analysis will control for these differences.

⁴ State legislation passed in 2006 requiring all high school students beginning with the graduating class of 2011 to earn four math credits. Further, the legislation included language that specified the courses required (Algebra, Geometry, Algebra II, and one credit earned for a mathematics class completed during the twelfth grade year).

based mathematics (NSF) program in heterogeneous classes, ($n = 200$). This group had a traditional middle school experience in Grades 6-8. It is important to note that Cohort 3 was the first complete group of students entering as freshmen without any incoming course stratification. Cohort 3 was still not required by the State to earn more than two mathematics credits as a graduation requirement.

Cohort 4 was selected because it represents students who entered as ninth-graders during Year Five of the reform effort in 2008 ($n = 218$) and participated in four years of heterogeneously grouped high school math categorized as reform-oriented and standards-based (NSF). The State mathematics high school graduation requirement for Cohort 4 was increased to four credits. This is the first cohort that was required to take the additional two credits of mathematics. This cohort also participated in three full years of reform-oriented, standards-based middle school mathematics (NSF) in heterogeneous groupings in Grades 6-8. Cohort 4 will be compared to the other cohorts to determine whether the additional years of treatment in middle school had any impact.

Data Quality

Preparing the data for analysis was an intensive process. The data were pulled from the countywide data warehouse that is coordinated and managed through the regional service district agency. Although the research study examined four cohorts of high school students, the analysis required me to gather information starting in middle school (prior state test scores). The county warehouse was instituted in 2003. During the first years of data entry at the school level, inconsistency of formatting and a general lack of awareness of the importance of data specificity resulted in eventual data errors. As a

result, my data preparation required painstaking checking and cleaning to ensure that the final data to be analyzed in this study were accurate.

The county data warehouse transitioned from the data management system, SASI, to Pinnacle, in 2007, which further compounded data challenges. This transition caused additional complications that affected my ability to format and merge files. On a few occasions, data from districts outside of Midwest appeared in some files from the county warehouse. After careful checking, I removed these data. I made several on-site visits to review student transcripts to be sure correct data were confirmed, or deleted when discrepancies were evident.

I spent nearly a full year participating in ongoing communication between Midwest technology officials and the Regional Education Service District (RESA). Once all the data were gathered, I entered them into Microsoft Excel 2007. During this step, I made several decisions about renaming, merging, and deleting particular information. For example, State mathematics test scores were originally pulled and sorted by subcategories (MME Algebra, MME Geometry, MME Numeracy, etc.). For purposes of this study, I decided to use the MME (Michigan Merit Exam of the Michigan Department of Education) total math scale score rather than examine subcategories. Additionally, to attend to the changes in the scale scores in the eighth-grade MEAP (Michigan Educational Assessment Program of the Michigan Department of Education) from 2003 to 2005 and to account for the differences in the scales used by the high school MEAP test and the high school MME and ACT, I used the State's statistical procedures to recode the numerical scores to the four proficiency levels defined by the State. Similarly, I used the State's statistical procedures to recode the WorkKeys numerical scores to the seven

proficiency levels defined by the State. (The ACT and WorkKeys tests are both published by the ACT company. The ACT test is a college entrance test and the WorkKeys test measures workplace skills.)

I made several other decisions about the data. Midwest recorded course titles and course grades by marking periods and semesters during the first years of the study. Later, the course grades were recorded by semesters only due to a change in course structure as the school moved to block scheduling. In other words, at the beginning of my study, math courses were a full year long. After an A/B block schedule was instituted in 2006, former full-year courses were completed in one semester. After school officials confirmed that course content and credits earned were the same for both structures, for purposes of consistency, I decided to average the grades for any particular math course and record only the “full-course final grade” for all cohorts.

It was necessary for me to distinguish between the few cases entered as *no credit*, *incomplete*, and simply *pass*. These cases were not arbitrarily considered failing or graded. I reviewed the cases in which *no credit* or *incomplete* was recorded on an individual basis to determine the final designation. For a few cases in which *incomplete* was recorded prior to a repeated sequence of course-taking by an individual student, I excluded the *incomplete* with the final grade recorded representing the repeated course sequence. Other isolated cases of *pass* or *no pass* were easily linked to outliers within the general population of the study. Some of these designations were specific to the few students enrolled in the most advanced, dually enrolled, college-level courses, while a few others were specific to the most severely challenged special education students. Clearly, in these extreme cases, individualized education plans had been developed.

Therefore, I eliminated these cases as outliers as they were not relevant to my specific research questions.

While examining the compiled data, I identified approximately 110 problem cases in which repeated senior-level math course information was included within the freshman course column. Once again, I reviewed student transcripts by hand, re-entered, and then replaced the correct titles within the master data file.

The economic status of students was originally recorded in three separate categories. These included non-reduced, reduced, and free lunch status. For purposes of this study, I combined reduced lunch and free lunch to create simple categorical values of *free/reduced lunch* and *no free/reduced lunch*. Although it was possible that students' economic status might have changed within the four-year period during high school, I decided to capture the economic status as entering freshmen only, for all cohorts.

The number of non-Caucasian students in any subcategory was less than 1%. Based on the small number of non-Caucasian students, I decided to eliminate them from the final analysis.

Nine special education subcategories were included in the original sample from Midwest. These included the following: severely cognitively impaired, mildly cognitively impaired, autistic, blind, hearing impaired, and severely emotionally impaired. I decided to include only those students identified as having a specific learning disability, because these students were typically included in the general classroom instruction for mathematics. In all other cases, either the number was too small to warrant inclusion (e.g., hearing impaired, $N = 1$) or the students were taught within a self-contained special

education classroom (e.g., severely cognitively impaired, $N = 7$). These students were not involved in the reform efforts and were eliminated from my analysis.

Coding

Table 10 shows the codes used in the data entry of the independent variables in the research study.

Table 10: Codes for Independent Variables

Independent Variable	Code
Gender	1 = Male; 2 = Female
Alaska Native/American Indian	1 = Yes; 0 = No
Black	1 = Yes; 0 = No
Asian	1 = Yes; 0 = No
Hispanic	1 = Yes; 0 = No
White	1 = Yes; 0 = No
Special Education	1 = Yes; 0 = No
Specific Learning Disability	1 = Yes; 0 = No
Emotional Impairment	1 = Yes; 0 = No
Resident of District	1 = Yes; 0 = No
Free/Reduced-Price Lunch	1 = Yes; 0 = No
Intervention strength (cohort)	1 = No treatment 2 = Minimum 2 years high school treatment 3 = Minimum 2 years high school and 1 year middle school treatment 4 = Minimum 4 years high school and 3 years middle school treatment

I examined student achievement by looking at different outcomes. This is important, as noted in the literature review, because many of the studies conducted on the impact of standards-based math programs indicate that students taught with reform programs fare as well as students taught with commercially developed programs when tested using traditional methods (i.e., state tests), but do better than students taught using commercially developed programs when assessed using measures that ask for deeper mathematical thinking.

I acquired student achievement scores for state assessments pre/post reform from the county data warehouse. I recorded eighth-grade Michigan Education Assessment Program (MEAP) mathematics scale scores for each cohort. I also recorded the mathematics scale scores from the eleventh-grade assessments (MEAP, MME, ACT, WorkKeys) required by the State for Cohorts 1, 2, 3, and 4.

The WorkKeys assessment, by design, requires students to apply mathematical reasoning, problem-solving, and critical thinking skills to find solutions to math problems situated in real-life contexts. I chose this subtest as an important indicator of student achievement and growth in the specific target areas aligned with NSF curricula.

I gathered student achievement scores on the district selected, problem-based, End of Sophomore Year Mathematics Assessment. This assessment included questions that aligned with values and goals identified in NSF standards-based mathematics programs. This test was developed using secured items (former high school MEAP test items used prior to the adoption of the Michigan Merit Curriculum) given to me by the state assessment officer who granted special permission for the use of these questions for purposes of measuring math skills directly connected to the math reform at Midwest. I

selected these test items because they matched the kind of mathematics reasoning and problem-solving skills emphasized in the NSF curricula.

State assessment officials also provided me with a scale score equating measure to allow for comparisons to be made between scale scores on this test and state assessment scores from the high school MEAP test prior to the MME/ACT high school test first given in 2007. Total points received were recorded for both portions of the test for all students with a total test score and percent correct entered into the database.

Unfortunately, test administration procedures that were used by teachers at times when I was not able to oversee the process were compromised. During the administration of the district test for Cohort 1, several teachers of advanced math classes were reluctant to use class time to administer the assessment. This resulted in test scores for only those students in the courses then classified as “lower-level.” The district test was administered to Cohort 2, but some teachers gave extra credit for participation while others simply relied on volunteerism. Cohort 3 was split into two groups of students who participated in the test during the final hour of the day on consecutive days. Athletic teams were dismissed approximately 40 minutes early during the time when the second group was taking the test to allow them to travel to out-of-district competitions. Therefore, nearly one-third of the group was unable to complete the test. Taking all of these factors into consideration, I was not confident in the sample size for any of the cohorts and was not able to determine whether a typical cross-section of the students was represented. Regrettably, I decided to eliminate this measure from my analysis. Instead, I relied on other indicators to address the question of whether or not standard measures indicate

growth in achievement for students the students who were instructed using the interventions provided during the reform.

Table 11 presents the total possible scale scores, as well as scale score ranges for all of the assessments used in my analysis, except for the WorkKeys assessment.

Table 11: Scale Scores, Ranges, and Performance Levels for Standardized Assessments

Assessment^a	Date(s) Administered	Cohort(s)	Total Range	Level D (State Level 4) Not Proficient	Level C (State Level 3) Partially Proficient	Level B (State Level 2) Proficient	Level A (State Level 1) Advanced
8th grade Math MEAP	March 2003-2005	1-2	246-850	<499	500-529	530-558	561-850
8th grade Math MEAP ^a	October 2005-2010	3-5	636-952	636-783	784-799	800-819	820-952
11th grade Math HST	Spring 2005, ^b 2006	1	55-989	< 499	500-529	530-626	630-989
11th grade MME Math	Spring 2007-2010	2-4	950-1250	950-1088	1089-1099	1100-1127	1128-1250
11th grade ACT Math	Spring 2007-2010	2-4	1-36	1-15	16-21	22-26	27-36

^a In Fall 2005, all MEAP tests changed. Testing moved from winter to fall. Note that eighth-grade MEAP scores for Cohort 2 reflect March 2005 (winter) testing and eighth-grade MEAP scores for Cohort 3 reflect October 2005 (fall) testing.

^b No cohorts in study reflect Spring 2006 HST.

Table 12 provides the scale scores, range, and performance levels for the WorkKeys assessment. These data are presented in a separate table because this assessment included seven levels of proficiency rather than the four levels for the assessments included in Table 11.

Table 12: Scale Scores, Ranges, and Performance Levels for the WorkKeys Standardized Assessment

Assessment	Total Range						
11th Grade WorkKeys (Spring 2007-2011; Cohorts 2-4)		<3	3	4	5	6	7
Scale scores	65-90	65-70	71-74	75-77	78-81	82-86	87-90

The highest-level mathematics course taken by Midwest students served as an additional outcome measure and a proxy for achievement. Initially, I coded mathematics course titles, producing a list of 22 courses (see Table 13).

Coding for the independent variables associated with math courses required that I create two sets of tables. Both are shown below.

Table 13: Math Course Codes

Math Course	Code
Math	1
Basic Math	2
General Math	3
Functional Math	4
Technical Math	5
Business Math	6
Intro to Algebra	7
Algebra	8
Algebra I	9
Intro Geometry	10
Geometry	11
Algebra II	12
MSU Algebra II	13

Math Course	Code
Pre-Calculus	14
Pre-Calculus/Trigonometry	15
Probability/Statistics	16
Discrete Math/Statistics	17
Statistics	18
Trigonometry	19
Calculus	20
AP Calculus	21
AP Calculus AB	22
AP Statistics	23
AP Calculus BC	24

I then recoded the original numerical codes used for mathematics courses to create a leveled, hierarchical structure. I created seven groupings, as follows.

Table 14: Math Course Code Frame

Math Course	Difficulty Ranking (1-7)
Math	1
Basic Math	1
General Math	1
Functional Math	1
Technical Math	1
Business Math	1
Algebra	2
Algebra I	2
Geometry	3
College Prep Geometry	3
Algebra II	4

Math Course	Difficulty Ranking (1-7)
MSU Algebra II	4
Pre-Calculus	5
Pre-Calculus/Trigonometry	5
Statistics	5
Discrete Math/Statistics	5
Trig/Pre-Calculus	5
Trigonometry	5
Calculus	6
AP Calculus	6
AP Calculus AB	6
AP Statistics	6
AP Calculus BC	7

By reframing the mathematics courses into hierarchical difficulty levels, I could better summarize information into tables by cohort and by demographic. I derived the ranking system for the traditional core of Algebra, Geometry, and Algebra II from the state graduation requirements to which Midwest adhered. The subsequent courses followed the manner in which they were represented in Midwest’s course guide, which included narrative descriptions to help me determine the degree of difficulty associated with each course.

I deliberated carefully as to whether or not to include Geometry in Group 2 or 3. My final decision to place Geometry in Group 3 was justified by the frequent references made in research studies to Algebra as a “gateway subject” for college entrance (Silver, 1995) as well as the current increased motivation to provide early access to Algebra;

especially by eighth grade. I felt that if I isolated Algebra and Geometry on their own, I would be able to examine questions about this important content more readily.

After I completed all initial data collection, merging, cleaning, and checking the data and data structures, I imported the raw Microsoft Excel 2007 data into the Statistical Package for the Social Sciences (SPSS) Base 18.0 computer program for statistical analysis. All descriptive statistics and statistical analysis were performed using this program.

Methods

In this section, I present the research design for this study. I also present the analytic methods utilized and provide a rationale for the selection of these methods.

Design

The design of the study was quantitative. The study was longitudinal, because it involved studying cohorts of students over a number of years. It had a correlational design; the goal was to see whether the reform curriculum (indicated by the cohort variable) was related to coursework and eleventh-grade standardized math test level. Demographic variables (including gender, receiving a free or subsidized lunch, and learning disability) and eighth-grade MEAP level were used as control variables.

Descriptive Analyses

Descriptive statistics were calculated for cohort and demographic variables overall. Then, the demographic variables, eighth-grade MEAP level, and the outcome variables were each analyzed descriptively by cohort, because cohort was the main variable of interest.

Analysis of Research Questions

The research questions were addressed by means of hierarchical regression analysis, in which variables are entered into the regression equation in sequential blocks. This enabled me to enter the control variables – demographics and eighth-grade MEAP – first, in order to allow examination of the effects of the other predictor variables – cohort, and for the standardized tests, also coursework – above and beyond the effects of the control variables.

In this research, from a statistical perspective, three main questions were addressed, relating to highest math course taken, number of math courses taken, and eleventh-grade MEAP/MME level. A Bonferroni correction was used because multiple statistical tests were applied. The required p -value for each significance test was thus set at 0.05 divided by 3 (the number of main tests), or .017. Two additional questions were addressed, relating to eleventh-grade ACT level and eleventh-grade WorkKeys level. However, these were not considered main questions because, due to the fact that these tests were not administered to Cohort 1, it was not possible to compare pre- and post-reform scores but only to compare scores of the three post-reform cohorts.

A hierarchical linear regression involves many p -values. Because the focus of the study was not all changes, but only changes related to the revised curriculum (represented by cohort), the p -value used was not that of the entire model, but that of cohort, and cohort as a whole rather than the individual cohorts, as will become clearer in the discussion of the hierarchical regressions in the Results chapter.

I used theoretical and research design considerations to determine the order of entry of the variables into the hierarchical regression. Demographic variables were

entered first, as control variables. Because prior achievement is a well-documented predictor for future achievement, the variable Eighth-Grade Math MEAP was also considered a control variable. It was entered as the step following the entry of the demographic variables.

Bivariate correlations were calculated as a precursor to the regression analysis. When performing a regression analysis, it is important to determine whether any of the predictor variables are highly correlated – with correlation coefficients greater than about .8 – because, in such cases, it is difficult to distinguish the effects of these highly correlated variables from each other (though one can still study their combined effect).

Analyses of Lower and Highest Eighth-Grade MEAP Levels

In an effort to discern whether the reform curriculum helped weaker or stronger students, as distinct from whether it helped the entire sample, the hierarchical regressions were rerun for the students whose eighth-grade MEAP levels were either 1 or 2, and for the students whose eighth-grade MEAP level was 4, respectively.

Ethical Concerns

I used only ex post facto data of the students' scale scores. No names, identification numbers, or other data that may identify students were used in order to assure student anonymity. I used alphabet codes to assure teacher anonymity as well.

CHAPTER 4: RESULTS

This research study was developed to measure the impact of a secondary mathematics reform that included grouping students for instruction using a heterogeneous structure while also implementing a National Science Foundation developed mathematics curriculum. The design included a comparison of several groups of students over the course of eight years with outcome measures made up of student achievement on eleventh-grade State assessments (MEAP and MME, which included the MME subtests ACT and WorkKeys), and mathematics course-taking patterns across four years of high school.

The purpose of the research study was to discern whether and to what extent differences in student achievement occurred as a result of the implementation of the NSF mathematics curriculum and a new student grouping structure at the sample school. Investigations of achievement difference were made for groups characterized by several variables, including gender, specific learning disabilities, economic status, eighth-grade mathematics MEAP proficiency levels, and cohort year (strength of treatment increased each year from 2004-2010). The results of this research study were intended to provide meaningful information to both middle and high schools where discussions about secondary mathematics reform are taking place, particularly when reforms are being enacted due to poor performance on standardized tests.

Research cited in Chapter 1 provides evidence that schools could better serve all students by abandoning traditional student grouping practices in mathematics often found in secondary schools (Burris, 2010; Oakes, 2005; Welner, 2006). These long-standing practices frequently provide stratified course offerings in mathematics which undermine

principles of equity and curriculum coherence and lead to dramatically different instructional practices for students depending upon the particular mathematics pathway to which one is assigned (Cogan, Schmidt, & Wiley, 2001). Further, evidence also outlined in Chapter 1 shows that the use of a problem-based curriculum versus a commercially developed curriculum yields results that are similar when students are tested using a traditional method, but yields results that are significantly better when students are tested using assessments that measure higher level thinking and problem solving skills (Burris, Heubert, and Levin, 2006; Stein, 2000). This study sought to determine whether reform initiatives such as those recommended in the literature can enhance equity, curriculum coherency, and, ultimately, student achievement at the local level.

This research study targeted four cohorts of students from 2003-2011 within a small, rural school district where secondary mathematics reform was instituted. Of the 1,110 students in the original sample, approximately 810 were enrolled in the Midwest School District for the bulk of the treatment (these students were enrolled for a minimum of 36 months, or 3.5 academic years, in Midwest High School). Students who were enrolled at Midwest for less than 3.5 academic years were eliminated from the study. After a decision to exclude Cohort 5 from the analyses, data from 737 students made up the final sample for analysis.

Descriptive Analyses

Analysis began with an examination of descriptive statistics.

Cohort

The main focus of interest in this analysis was the results of implementing the new curriculum. This curriculum was implemented starting with the second cohort.

Thus, Cohort 1 represented students who participated in a traditional mathematics instructional program with use of commercially developed textbooks, within stratified student grouping structures. Cohort 2 represented students who participated in two years of reform mathematics. Cohort 3 represented students who participated in four years of high school reform mathematics Cohort 4 represented students who were required to participate in three years of high school reform mathematics. In addition, Cohort 4 participated in two years of middle school mathematics with a similar reform model.

Some data were collected for an intended fifth cohort: eighth-grade MEAP scores and coursework through eleventh grade, but not information on twelfth-grade coursework or standardized exams. It was decided not to include Cohort 5 in any analyses because the major outcome variables were unavailable.

The distribution of students by cohort is displayed in Table 15. Note that some cohorts are separated by one year and some by two years.

Table 15: Distribution of Students by Cohort

Cohort	N	Percent
Cohort 1 (Started 9/03)	158	21.4
Cohort 2 (Started 9/05)	218	29.6
Cohort 3 (Started 9/06)	177	24.0
Cohort 4 (Started 9/08)	184	25.0
Total	737	100.0

Student Characteristics

Measures of a number of student characteristics were used in the analyses as control variables. The results appear in Table 16. Because of the importance of cohort as a variable, the distribution by cohort is also included. Also included for each characteristic is the Pearson chi-square statistic, which indicates whether there was a

relationship between that characteristic and cohort; that is, whether that characteristic was distributed differently in the different cohorts.

Table 16: Demographic Characteristics of Participants, as Percentages by Cohort, and Pearson Chi-Square

Characteristic	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total	Pearson Chi-Square
Gender						0.704
Male	54.1	51.4	53.1	50.0	52.0	
Female	45.9	48.6	46.9	50.0	48.0	
Free/Reduced Lunch						3.910
Yes	20.4	21.8	23.7	28.7	23.7	
No	79.6	78.2	76.3	71.3	76.3	
Special Needs						4.335
Learning disabled	5.9	6.7	10.8	5.6	7.2	
Not learning disabled	94.1	93.3	89.2	94.4	92.8	
Residential Status						10.075*
In district	82.9	72.6	76.3	68.5	75.1	
Out of district	17.1	27.4	23.7	31.5	24.9	
N	158	218	177	184	737	

* $p < .05$.

Gender

There were roughly the same number of males (52.0%) and females (48.0%) in the sample, with virtually no differences between cohorts.

Free or Reduced-Price Lunch

Just under one-quarter (23.7%) of the sample received free or reduced-price lunches. This percentage increased gradually, from 20.4% in the first cohort to 28.7% in the last cohort, but the difference was not significant.

Learning Disability

As noted in Chapter 3, students diagnosed with a specific learning disability were included in this study because, by and large, they participated in general education classrooms for the majority of their mathematics instruction. Students with severe emotional impairments or students whose cognitive skills were extremely limited were excluded from this study because they were not included in general education classrooms for the majority of their mathematics instruction. The overall percentage of students with learning disabilities was 7.2%, with slight but not statistically significant differences between samples.

Residence

The number of out-of-district students enrolled at Midwest as school of choice increased during the years of this study. In Cohort 1, 17.1%, or about one in six, came from out of the district. By Cohort 4, the percentage was 31.5%, or nearly one-third. This difference was statistically significant and suggests that, over the period of the study, Midwest became a more highly regarded school, such that outsiders wanted to study there.

Summary

Overall, with the exception of residence, there were no significant differences in demographic characteristics between the four cohorts.

Means Analyses

Eighth-Grade MEAP Level

Because there were changes in the middle school curriculum as well as the MEAP over the course of the study, and because the main topic of interest was high

school achievement, eighth-grade MEAP scores were considered as a covariate in the study.

Table 17: Eighth-Grade MEAP Level of Participants, as Percentages, by Cohort

MEAP Level	Cohort 1	Cohort 2 ^a	Cohort 3	Cohort 4	Total
Not proficient (1)	14.6	4.7	5.2	4.7	7.0
Partially proficient (2)	27.8	17.8	15.1	12.9	18.1
Proficient (3)	52.3	73.3	41.9	40.0	52.5
Advanced (4)	5.3	4.2	37.8	42.4	22.4
Mean Level	2.48	2.77	3.12	3.20	2.90
N	151	191	172	170	684

Note. The Spearman correlation between MEAP level and cohort was 0.351 ($p = .000$).

^a MEAP data were missing for approximately 30 students in Cohort 2.

As is evident from Table 17, MEAP level rose fairly dramatically over the course of the study, with the Spearman correlation between MEAP level and cohort being .351 ($p = .000$). First, there was a decrease of approximately 20% in the percentage of students graded not proficient or partially proficient between Cohort 1 (42.4%) and Cohort 2 (22.5%). Second, there was a dramatic increase of more than 30% in the percentage of students graded advanced between Cohort 2 (4.2%) and Cohort 3 (37.8%). The mean MEAP level rose from 2.48 in Cohort 1 to 2.77 in Cohort 2 and 3.12 in Cohort 3, a total rise of approximately two-thirds of a level.

Comparing the sample results to those of the whole state strengthens this analysis (see Table 18). Whereas in Cohorts 1 and 2, average MEAP levels at Midwestern were slightly below state-wide averages, in Cohorts 3 and 4, average MEAP levels at Midwestern were somewhat above state-wide averages, which suggests an improvement over time.

Table 18: State-Wide Eighth-Grade MEAP Levels, by Cohort

Level	Cohort 1	Cohort 2 ^a	Cohort 3 ^b	Cohort 4 ^a
Not proficient (1)	23.4	16	14.2	9
Partially proficient (2)	22.4	21	22.7	19
Proficient (3)	24.4	25	32.7	30
Advanced (4)	29.4	38	30.5	41
Mean Level (calculated)	2.60	2.85	2.80	3.04
N	122,961	132,928	131,731	122,797

^a Levels for these two years were reported on the State website rounded to the nearest percent.

^b The test window was moved from winter to fall beginning in 2005 (relevant for Cohort 3). New test content reflects K-8 Grade Level Content Expectations (GLCEs) learned during the previous academic school year. GLCEs were introduced and implemented for the first time in 2004.

Outcome Variables

The outcome variables measured were of two types. The first type was high school math courses: highest course taken and number of courses taken during high school (up to and including grade 12). The second type was standardized test results from eleventh grade: MEAP/MME, ACT (only Cohorts 2 to 4) and WorkKeys (only Cohorts 2 to 4). Overall means for these outcome variables are reported in Table 19. Means and standard distributions are also reported by gender, free or reduced-price lunch, disability, residence, eighth-grade MEAP level, and cohort.

Table 19: Means and Standard Deviations of Outcome Variables

Characteristic	Range	Mean	SD	N
Highest-Level Math Course, Grades 9-12	1-7	4.02	1.48	716
Number of Math Courses, Grades 9-12	1-7	3.41	1.34	716
Eleventh-Grade MEAP/MME	1-4	2.22	1.02	590
ACT ^a	1-4	2.09	0.84	466
WorkKeys ^a	2-7	4.58	1.43	464

^a The ACT and WorkKeys were given only from Cohort 2.

Gender

Table 20 displays the mean values of the outcome variables by gender.

Table 20: Means and Standard Deviations of Outcome Variables by Gender

Variable	Males	Females	Total
Highest-Level Math Course ^a			
Mean	3.86	4.20	4.02
SD	1.50	1.42	1.47
N	373	342	715
Number of Math Courses ^b			
Mean	3.35	3.48	3.41
SD	1.38	1.29	1.34
N	373	342	715
Eleventh-Grade MEAP/MME ^c			
Mean	2.25	2.18	2.22
SD	1.06	0.98	1.02
N	302	287	589
ACT ^d			
Mean	2.09	2.09	2.09
SD	0.86	0.83	0.84
N	235	231	466
WorkKeys ^e			
Mean	4.61	4.55	4.58
SD	1.51	1.34	1.43
N	236	228	464

^a $F(1,713) = 9.345, p = .002$.

^b $F(1,713) = 1.785, p = .182$.

^c $F(1,587) = 0.573, p = .450$.

^d $F(1,464) = .001, p = .972$.

^e $F(1,462) = .162, p = .688$.

Highest-Level Math Course: There was a moderately large difference between boys and girls in terms of average highest-level math course taken: 3.86 for boys as opposed to 4.20 for girls. This difference, about one-third of a course level, was significant at $p = .002$.

Number of Math Courses: Girls had a slightly higher average number of math courses than boys: 3.48 versus 3.35. However, the difference was not significant.

Eleventh-Grade MEAP/MME: Boys had slightly higher average levels of eleventh-grade MEAP/MME than did girls. However, the difference was not significant.

Eleventh-Grade ACT: There was no difference in average ACT level between boys and girls: it was 2.09 for both.

Eleventh-Grade WorkKeys: There was only a small difference between boys and girls in average WorkKeys level, and the difference was not significant.

Summary: In terms of gender, results differed for coursework and standardized tests. Girls did somewhat better than boys on the coursework variables, but there was virtually no difference between them in terms of standardized tests.

Free or Reduced-Price Lunch

Table 21 displays the mean values of the outcome variables by free or reduced-price lunch.

Table 21: Means and Standard Deviations of Outcome Variables by Free or Reduced-Price Lunch

Variable	Free or Reduced- Priced Lunch	Regular-Price Lunch	Total
Highest-Level Math Course ^a			
Mean	3.29	4.26	4.02
SD	1.42	1.42	1.47
N	168	544	712
Number of Math Courses ^b			
Mean	2.98	3.55	3.42
SD	1.32	1.31	1.33
N	168	544	712
Eleventh-Grade MEAP/MME ^c			
Mean	1.87	2.30	2.22
SD	0.93	1.03	1.02
N	118	469	587
ACT ^d			
Mean	1.91	2.13	2.09
SD	0.79	0.85	0.85
N	93	372	465
WorkKeys ^e			
Mean	4.19	4.67	4.58
SD	1.39	1.42	1.43
N	93	370	463

^a $F(1,710) = 59.158, p = .000.$

^b $F(1,712) = 24.342, p = .000.$

^c $F(1,585) = 17.374, p = .000.$

^d $F(1,463) = 4.981, p = .026.$

^e $F(1,461) = 8.542, p = .004.$

Highest-Level Math Course: Students with free or reduced-price lunches had a much lower average highest-level math course (3.29) than students with regular-priced lunches (4.26). This difference, nearly a whole level, was significant ($p = .000$).

Number of Math Courses: On average, students with free or reduced-price lunches took more than one-half fewer math courses (2.98) than did students with regular-priced lunches (3.55). The difference was significant ($p = .000$).

Eleventh-Grade MEAP/MME: The average MEAP/MME level of students with free or reduced-price lunches was almost one-half a level lower than that of students with regular priced lunches: 1.87 versus 2.30. The difference was significant.

Eleventh-Grade ACT: The average ACT level of students with free or reduced-price lunches was 1.91, somewhat lower than that of students with regular-prices lunches, which was 2.13. The difference, though not nearly as large as the other differences, was significant ($p = .026$)

Eleventh-Grade WorkKeys: Students with free or reduced-price lunches had an average WorkKeys level that was half a level (out of six) lower than that of students with regular-priced lunches: 4.19 v. 4.67. The difference was significant ($p = .004$).

Summary: Children who received free or reduced-price lunches tended to perform significantly worse on all outcome measures than did other children.

Learning Disability

Table 22 displays the mean values of the outcome variables by learning disability.

Table 22: Means and Standard Deviations of Outcome Variables by Learning Disability

Variable	Learning Disability	No Learning Disability	Total
Highest-Level Math Course ^a			
Mean	2.66	4.20	4.09
SD	1.17	1.40	1.44
N	50	642	692
Number of Math Courses ^b			
Mean	2.86	3.49	3.45
SD	1.25	1.33	1.34
N	50	642	692
Eleventh-Grade MEAP/MME ^c			
Mean	1.05	2.30	2.22
SD	0.23	1.00	1.02
N	37	541	578
ACT ^d			
Mean	1.17	2.15	2.09
SD	0.38	0.83	0.85
N	29	430	459
WorkKeys ^e			
Mean	3.07	4.70	4.60
SD	1.02	1.38	1.42
N	30	427	460

^a $F(1,690) = 57.293, p = .000.$

^b $F(1,690) = 10.534, p = .001.$

^c $F(1,576) = 57.111, p = .000.$

^d $F(1,457) = 39.473, p = .000.$

^e $F(1,455) = 40.581, p = .000.$

Highest-Level Math Course: For students with learning disabilities, the average highest-level math course was 2.66; for students without disabilities, it was about one and a half levels higher, at 4.20. The difference was significant ($p = .000$).

Number of Math Courses: Whereas students with disabilities took an average of 2.86 math courses in high school, students without disabilities took an average of 3.49 courses, nearly two-thirds of a course more. The difference was significant ($p = .001$).

Eleventh-Grade MEAP/MME: Students with disabilities were more than one level lower on the eleventh-grade MEAP/MME than those without disabilities: 1.05 versus 2.30. The difference was significant ($p = .000$).

Eleventh-Grade ACT: Students with disabilities had a one level lower average ACT score than students without disabilities: 1.17 versus 2.15. The difference was significant ($p = .000$).

Eleventh-Grade WorkKeys: Students with disabilities were more than one and a half (out of six) levels lower on WorkKeys than students without disabilities: 3.07 versus 4.70. The difference was significant ($p = .000$).

Summary: Children with learning disabilities tended to do worse on all outcome measures than other children.

Residence

Table 23 displays the mean values of the outcome variables by residence.

Table 23: Means and Standard Deviations of Outcome Variables by Residence

Variable	In-District Residence	Out-of-District Residence	Total
Highest-Level Math Course ^a			
Mean	3.99	4.13	4.02
SD	1.48	1.47	1.48
N	535	178	713
Number of Math Courses ^b			
Mean	3.43	3.39	3.42
SD	1.34	1.31	1.33
N	535	178	713
Eleventh-Grade MEAP/MME ^c			
Mean	2.19	2.34	2.22
SD	1.03	0.99	1.02
N	445	143	588
ACT ^d			
Mean	2.08	2.12	2.09
SD	0.84	0.87	0.85
N	344	121	465
WorkKeys ^e			
Mean	4.52	4.73	4.59
SD	1.43	1.42	1.43
N	342	121	463

^a $F(1,711) = 1.211, p = .271.$

^b $F(1,711) = 0.123, p = .726.$

^c $F(1,586) = 2.605, p = .107.$

^d $F(1,463) = 0.173, p = .677.$

^e $F(1,461) = 1.831, p = .177.$

Highest-Level Math Course: There was only a small, and non-significant difference between in-district and out-of-district students on highest average level math course (3.99 versus 4.13).

Number of Math Courses: There was almost no difference between in-district and out-of-district students on average number of math courses taken (3.43 versus 3.39).

Eleventh-Grade MEAP/MME: The out-of-district students had a slightly higher average MEAP/MME level (2.34) than the in-district students (2.19), but the difference was not significant.

ACT: There was almost no difference between in-district and out-of-district students on average ACT level (2.08 versus 2.12).

WorkKeys: Out-of-district students had a slightly higher average WorkKeys level (4.73) than in-district students. The difference was not, however, significant.

Summary: In all of the outcome variables, the out-of-district students tended to have slightly "better" outcomes, but none of the differences was significant.

Cohort

Table 24 displays the outcome variable means by cohort.

Math Coursework: The first outcome of interest related to high school math courses taken. The variables measured were the highest-level math course taken and the number of math courses which each student took (see Table 24).

Table 24: Means and Standard Deviations of Outcome Variables for Cohorts 1 through 4

Variable	Cohort 1	Cohort 2	Cohort 3	Cohort 4	Total
Highest-Level Math Course ^a					
Mean	3.72	3.77	4.18	4.42	4.02
SD	1.50	1.34	1.52	1.44	1.48
N	158	206	174	178	716
Number of Math Courses ^b					
Mean	3.72	2.93	3.11	3.98	3.41
SD	1.27	1.09	1.43	1.28	1.34
N	158	206	174	178	716
Eleventh-Grade MEAP/MME ^c					
Mean	2.33	2.10	2.25	2.21	2.22
SD	1.02	1.03	1.03	1.00	1.02
N	127	172	142	149	590
ACT ^d					
Mean	—	2.04	2.11	2.12	2.09
SD	—	0.88	0.85	0.80	0.84
N	—	172	146	148	466
WorkKeys ^e					
Mean	—	4.54	4.72	4.48	4.58
SD	—	1.44	1.41	1.42	1.43
N	—	171	144	149	464

Note. ACT and WorkKeys were given only from Cohort 2.

^a $F(3,712) = 9.459, p = .000$.

^b $F(3,712) = 27.881, p = .000$

^c $F(3,586) = 1.261, p = .287$

^d $F(2,463) = 0.434, p = .648$.

^e $F(2,461) = 1.125, p = .329$.

Highest Level Math Course: The highest-level math course taken in high school started at an average of 3.72 in Cohort 1, stayed virtually unchanged (3.77) in Cohort 2, and then rose slowly to 4.42 by Cohort 4. This suggests that the new curriculum (implemented for Cohort 2) may have had a small positive effect on level of courses taken.

Number of Math Courses: The number of math courses taken over high school started at an average of 3.72 in Cohort 1, dropped to 2.93 in Cohort 2, and then rose slowly to 3.98—slightly higher than it was in Cohort 1—by Cohort 4. Whereas this is not as positive an outcome as that for highest level courses, it might be that at first the reform curriculum was associated with a decrease in total number of courses taken, but that this negative effect was eventually overcome and by the fourth year of the program (Cohort 4), students were taking very slightly more mathematics courses than before the program started.

These two results will be further explored in the regression analysis, which includes cohorts and the other variables as well.

Eleventh-Grade MEAP and MME: Unlike eighth-grade MEAP scores, eleventh-grade MEAP/MME scores remained fairly constant both for Midwestern and the state as a whole, in the range of 2.2 to 2.3 (see Table 25). The one exception was the Midwestern MME score for Cohort 2, which was only 2.10. This fits in with the findings regarding coursework, which showed a sharp dip in the first year of the new program, and then recovered. In the case of the MME, the changes did not reach statistical significance ($p = .287$).

Table 25: State Means for Eleventh-Grade MEAP/MME, By Cohort

	Cohort 1	Cohort 2	Cohort 3	Cohort 4
Mean	2.30	2.21	2.22	2.26
N	116,509	113,839	113,234	108,932

Note. State means were calculated manually from state data on distribution of levels, which were reported rounded to the nearest whole percentage.

The test window was moved from winter to fall beginning in 2005 (relevant for Cohort 3). The new test content reflected K-8 Grade Level Content Expectations (GLCEs) learned during the previous academic school year. GLCEs were introduced and implemented for the first time in 2004.

Eleventh-grade ACT: As noted above, the ACT was not administered to Cohort 1. Examining Cohorts 2 to 4, there was a slight rise in average level (from 2.04 to 2.12) between Cohort 2 and Cohort 4, but this was not significant ($p = .648$).

WorkKeys: As noted above, WorkKeys was also not administered to Cohort 1. Examining Cohorts 2 to 4, there was no obvious trend in the average levels, which ranged from 4.48 (Cohort 4) to 4.72 (Cohort 3; note that the WorkKeys levels, unlike the other tests, ranged from “1 and 2” to 7). The differences were not significant ($p = .329$).

Summary: In all three of the outcome measures that existed in Cohort 1 (highest-level math course, number of math courses, and 11th-grade MEAP/MME), there was a drop between Cohorts 1 and 2, and then a moderate rise between Cohort 2 and Cohort 3. In the case of the coursework outcomes, there was a further rise between Cohort 3 and Cohort 4, and Cohort 4 was even higher than Cohort 1. In the case of the standardized tests, Cohort 4 results were similar to (and in the case of WorkKeys, even a bit lower than) those of Cohort 3. The implications of these findings are discussed further in Chapter 5.

Analysis of Research Questions

The following two research questions guided the study:

1. How do mathematics outcomes of Midwest High School students compare pre- and post-reform?
 - 1a. How does the mathematics course taking of students compare pre- and post-reform?
 - 1b. How does mathematics achievement compare pre- and post-reform?
2. How do different types of students at Midwest High School fare in the reform?
 - 2a. How do students whose prior math achievement in 8th grade indicates lowest proficiency fare when compared pre- and post-reform?
 - 2b. How do students whose prior math achievement in 8th grade indicates highest proficiency fare when compared pre- and post-reform?

In these two questions, the relationships between the predictor (independent) variables (demographics, eighth-grade MEAP, and cohort) and the outcome (dependent) variables (high school coursework and eleventh-grade standardized test levels) are addressed.

Correlations Among Predictor Variables

Note that because some of the variables were ordinal rather than interval, Spearman's rho was used rather than Pearson correlations. As can be seen in Table 26, though several of the correlations were moderately high, the only one above 0.50 was that between the highest-level math course and the number of math courses, which was 0.68. Thus the intercorrelations were acceptably low.

Table 26: Correlations (Spearman's Rho) Among the Predictor Variables

	1	2	3	4	5	6	7	8	9	10	11
1 Disability	—										
2 Gender	-.11**	—									
3 Lunch	.13**	-.05	—								
4 Residence	.06	-.05	.03	—							
5 MEAP Gr. 8	-.23**	.01	-.11**	-.01	—						
6 Highest-Level Math Course to Grade 11	-.28**	.13**	-.27**	-.06	.52**	—					
7 Number of Math Courses to Grade 11	-.13**	.04	-.17**	-.01	.25**	.68**	—				
8 Cohort 1	-.03	-.02	-.04	.10**	-.27**	-.18**	.13**	—			
9 Cohort 2	-.01	.01	-.03*	-.03	-.13**	-.14**	-.27**	-.34**	—		
10 Cohort 3	.08*	-.01	.00	.02	.17**	.17**	-.11**	-.29**	-.36**	—	
11 Cohort 4	-.04	.02	.07	-.08*	.23**	.16**	.27**	-.30**	-.37**	-.32**	—

* $p < .05$. ** $p < .01$.

Note that Cohort was added in not as a single variable but as dummy variables. This was done because the means analysis suggested that the relationship between cohort and the outcome variables was likely to be non-linear, that is, not steadily increasing or decreasing, yet showing some pattern. Adding in the cohorts individually allowed us to discern such a pattern, if any in fact existed.

Research Question 1a: Mathematics Course Taking

Research Question 1 stated: How does the mathematics course taking of students compare pre/post reform?

Is there a statistically significant difference in mathematics course taking across four years of high school for students when taught using an NSF program within heterogeneously grouped classes versus a traditional curriculum within academically tracked classes?

Both the highest-level math course taken and the number of math courses taken were analyzed. These are discussed in the following sections.

Bivariate Correlations

As a precursor to the regression analysis, the bivariate correlations between the predictor and outcome variables were examined (Table 27). Because some of the variables were ordinal rather than interval, Spearman's rho was used rather than Pearson correlations.

Table 27: Correlations (Spearman's Rho) Between Coursework Variables and Demographics, Eighth-Grade MEAP Level, and Cohort

	Highest-Level Math Course Grades 9-12	Number of Math Courses Grades 9-12
Disability	-.27**	-.12*
Gender	.12*	.05
Free or Reduced-Price Lunch	-.28**	-.19**
Residence	-.05	.01
Eighth-Grade MEAP Level	.48**	.25**
Cohort 1	-.09	.13**
Cohort 2	-.12*	-.24**
Cohort 3	.07	-.13*
Cohort 4	.15**	.28**

* $p < .01$. ** $p < .001$.

As one can see, whereas eighth-grade MEAP clearly has the strongest bivariate correlation with coursework, all of the other variables, except for residence, have very strong ($p < .01$) correlations with at least one (and usually both) of the outcome variables. Because of the very weak correlation of residence with both of the outcome variables, combined with the results of the means analyses above, it was decided to omit this variable from the regression.

Highest-Level Math Course Taken

In a hierarchical regression, one adds the variables in blocks, often containing more than one variable, in order to see the consecutive effect of each block of variables after accounting for the previous blocks of variables. Often, one first enters variables which are expected to be statistically significant but which are not of interest in the analysis being done. In this research, it was expected that the demographic variables (or at least some of them) and eighth-grade MEAP would be significant predictors of the outcome variables, but these variables were of less interest. The primary interest was

cohort, which was used as an indicator of the new program. Therefore, the demographics and eighth-grade MEAP were entered as blocks into the regression, followed by cohort.

The results of the hierarchical regression for highest-level math course appear in Table 28.

Table 28: Hierarchical Regression Analysis Summary for Demographic Variables, Eighth-Grade MEAP, and Cohort Predicting Highest-Level High School Math Course

Step and predictor variable	<i>B</i>	<i>SE B</i>	β	<i>p</i>	R^2	ΔR^2	ΔF	<i>p</i>
Step 1					.12	.12	28.66	.000
Learning Disability	-1.33	.22	-.22	.000				
Gender	.15	.11	.05	.147				
Free/Red. Lunch	-.76	.13	-.22	.000				
Step 2					.29	.17	159.34	.000
8th-grade MEAP	.76	.06	.43	.000				
Step 3					.30	.01	1.99	.114
Cohort 2	-.17	.13	-.06	.193				
Cohort 3	-.01	.15	-.00	.931				
Cohort 4	.14	.15	.04	.320				

Step 1: As one can see, the first step included the demographic variables (excluding residence). The first line ("Step 1") indicates that these three variables explain only twelve percent of the variance ($R^2 = .12$), but still constitute a model that is significantly better than a model with no predictors ($p = .000$). The next three lines show that both having a disability and receiving a free or reduced-price lunch were significant negative predictors of highest-level math course. That is, students with a learning disability or receiving a free or reduced-price lunch tended to take lower-level math courses. Also, being female was a non-significant positive predictor of highest-level math course. That is, girls tended to take slightly higher-level math courses than boys. All of

these results fit in with the findings from the means analysis and from the bivariate correlations.

Step 2: As a second step, eighth-grade MEAP was added. This added 17% (ΔR^2) to the model for a total of 27% of the variance explained (R^2). The contribution of eighth-grade MEAP to the model, after controlling for demographics, was significant ($p = .000$). This is apparent from the following line, where the coefficient of eighth-grade MEAP is listed. The coefficient of 0.76 means that for every level higher on the eighth-grade MEAP, students reached, on average, math courses that were approximately three-quarters of a level higher.

It is important to recall, though, that eighth-grade MEAP was a control variable. That is, its value per se was not of interest. The interest was in controlling for it before adding in cohort. Or, to put it otherwise, I was interested in studying the predictor value of cohort after taking into account the predictor value of the demographic variables and eighth-grade MEAP, which were not related to high school curriculum but which had a strong relationship with highest math course taken.

Step 3: The third step was to add cohort. This step was the one in which the research question was addressed. As explained earlier, because it seemed that any predictive value of cohort might not be linear, it was decided to enter it as multiple dummy variables rather than as a single continuous variable. The variable left out of the regression was Cohort 1, which represented the situation before the curriculum was changed.⁵

⁵ In any regression, the coefficient for any particular value must have a reference point, which is the predicted value when the variable is equal to zero. In the case of dummy variables, too, it is necessary to have a zero, or reference, point. One of the dummy

Cohort as a whole contributed only about 1% (ΔR^2) to the predictive value of the model, and this contribution was not significant ($p = .114$). Since this was the p -value to be compared to .017 according to the Bonferroni discussion above, this indicated that there was no strong evidence of differences among cohorts, and particularly between Cohort 1 and the other cohorts, in the highest-level math course taken. None of Cohort 2, Cohort 3, or Cohort 4 was significantly different from Cohort 1 (none of the p -values is even close to .05).⁶ In fact, their sizes suggest that any changes were quite small.

Although the coefficients were not significant, they were, nonetheless, somewhat in line with the findings from the means analysis. That is, there was a drop between Cohort 1 and Cohort 2 (which had the largest negative coefficient, indicating lower-level courses), with Cohort 3 higher than Cohort 2 and Cohort 4 higher than Cohort 3 (that is, Cohort 3 had a smaller negative coefficient, and the coefficient of Cohort 4 was positive).

Summary: The hierarchical linear regression indicates that, after controlling for the demographics and eighth-grade MEAP level, there were no significant differences among the cohorts, nor, indeed, between Cohort 1 and any of the other cohorts, in terms of highest-level math course taken.

variables constituting the original variable is chosen, and the other values are compared to it. Consider, for example, gender in this analysis. The coding of gender was such that males were the reference point, and a positive coefficient thus indicated that females tended to score higher. In this research, since the first cohort was before the program was instituted, it was appropriate to compare the other cohorts to it. Thus, Cohort 1 was left out of the analysis, and all of the other cohort coefficients were interpreted relative to Cohort 1.

⁶ Only cohort as a whole is affected by the Bonferroni correction. The other values, which are not the "critical" ones, are still compared to $p = .05$. To be more specific: The test of the hypotheses is only on the cohort line. The individual lines serve just to see which ones are significant, but cannot in themselves confirm or refute the hypotheses. This is because the hypothesis does not relate to differences between Cohorts 1 and 3, for instance, but to cohort in general.

Number of Math Courses Taken

The results of the hierarchical regression for number of math courses taken appear in Table 29.

Table 29: Hierarchical Regression Analysis Summary for Demographic Variables, Eighth-Grade MEAP, and Cohort Predicting Number of Math Courses Taken

Step and predictor variable	<i>B</i>	<i>SE B</i>	β	<i>p</i>	R^2	ΔR^2	ΔF	<i>p</i>
Step 1					.04	.04	10.03	.000
Learning Disability	-.76	.21	-.14	.000				
Gender	-.01	.10	.00	.963				
Free/Red. Lunch	-.45	.12	-.14	.000				
Step 2					.08	.03	24.39	.000
8th-grade MEAP	.31	.06	.19	.000				
Step 3					.18	.10	27.13	.000
Cohort 2	-.91	.13	-.32	.000				
Cohort 3	-.85	.14	-.28	.000				
Cohort 4	-.09	.14	-.03	.549				

Step 1: The first step included the demographic variables (excluding residence). These three variables explained only four percent of the variance ($R^2 = .04$), even though they constituted a model that was significantly better than a model with no predictors ($p = .000$).

Both having a disability and receiving a free or reduced-price lunch were significant negative predictors of highest-level math course. That is, students with a learning disability or receiving a free or reduced-price lunch tended to take fewer math courses. Being female was almost unrelated to highest-level math course. All of these results fit in with the findings from the means analysis and from the bivariate correlations.

Step 2: As a second step, eighth-grade MEAP was added in. This added approximately 3% (ΔR^2) to the model for a total of 8% (R^2) (the numbers do not add up exactly because of rounding). The contribution of eighth-grade MEAP to the model, once controlling for demographics, was significant ($p = .000$). The coefficient of 0.31 means that for every level higher on the MEAP achieved by students, they took, on average, one-third more math courses in high school.

Step 3: The third step was to add cohort. As can be seen in "Step 3," cohort as a whole contributed 10% (ΔR^2) to the predictive value of the model, and this contribution was significant ($p = .000$).

Since this was the p -value to be compared to .017 according to the Bonferroni discussion above, this result provided strong evidence of differences between cohorts, and particularly between Cohort 1 and the other cohorts, in the number of mathematics courses taken. However, the next three lines (Cohort 2, Cohort 3, and Cohort 4) suggest that the significance was in the "wrong" direction. Both Cohorts 2 and 3 had significant ($p = .000$) negative coefficients, indicating that students in these cohorts took significantly fewer courses than students in Cohort 1, while Cohort 4 had a fairly small and non-significant negative coefficient.

While the results indicate a short-term decrease in number of mathematics courses taken, the longer-term results indicate a relative increase. That is, although there was a drop between Cohort 1 and Cohort 2 (which had the largest negative coefficient, indicating that students in Cohort 2 took the fewest courses of all the cohorts), Cohort 3 had a smaller negative coefficient than Cohort 2 (that is, students took more mathematics courses than in Cohort 2) and Cohort 4 had a fairly small coefficient, meaning that, after

controlling for the demographics and eighth-grade MEAP level, students in Cohort 4 took about the same number of courses as students in Cohort 1.

Summary: The hierarchical linear regression indicates that, after controlling for the demographics and eighth-grade MEAP level, there were significant differences among the cohorts in terms of the number of math courses taken by students. While the results are in the "wrong" direction, examining the individual cohort results does suggest that perhaps the cohorts improved over time.

Research Question 1b: Mathematics Achievement

Research Question 1b stated: How does math achievement compare pre- and post-reform?

Is there a statistically significant difference in eleventh-grade student achievement in mathematics when taught using an NSF program within heterogeneously grouped classes versus a traditional curriculum within academically tracked classes, as measured by eleventh-grade MEAP, MME, WorkKeys, and ACT?

For these analyses, highest math course through Grade 11 and number of math courses through Grade 11 were included as predictor variables, because it was anticipated that they would be related to standardized exam results. They were included as an extra step between eighth-grade MEAP and cohort. Because they were in themselves related to cohort (though the previous analysis related to coursework through twelfth grade rather than eleventh grade), they were treated as regular predictors rather than control variables.

Bivariate Correlations

As a precursor to the regression analysis, the bivariate correlations between the predictor and outcome variables were examined (see Table 30). Because some of the

variables were ordinal rather than interval, Spearman's rho was used rather than Pearson correlations.

Table 30: Correlations (Spearman's Rho) Between the Eleventh-Grade Standardized Test Result Variables and Demographics, Eighth-Grade MEAP Level, Coursework, and Cohort

Variable	11th-Grade MEAP/MME	11th-Grade ACT	11th-Grade WorkKeys
Disability	-.30**	-.30**	-.28**
Gender	-.03	.01	-.02
Lunch: Free or Reduced Lunch	-.17**	-.10	-.13*
Residence: In-District Residence	-.07	-.02	-.07
Eighth-Grade MEAP Level	.54**	.51**	.54**
Highest-Level Math Course to Grade 11	.60**	.61**	.59**
Number of Math Courses to Grade 11	.41**	.35**	.31**
Cohort ^a	.06	—	—
Cohort2	-.07	-.05	-.03
Cohort3	.02	.02	.07
Cohort4	.00	.04	-.04

^a The ACT and WorkKeys tests were given only starting with Cohort 2.

* $p < .01$. ** $p < .001$.

As can be seen in Table 30, both eighth-grade MEAP and math coursework were most strongly correlated with standardized test results. Disability and free or reduced-price lunch were also strongly correlated with standardized test results, but residence, gender, and cohort were not. Residence was therefore omitted from the hierarchical linear regression, but gender was left in for comparability with the other regression analyses, and cohort was left in because it was the variable of most interest in the study.

Eleventh-Grade MEAP/MME

The results of the hierarchical regression for eleventh-grade MEAP/MME appear in Table 31. As one can see, in the regressions for eleventh-grade test results, coursework through eleventh grade was added in as a third step in the regression.

Table 31: Hierarchical Regression Analysis Summary for Demographic Variables, 8th-Grade MEAP, Coursework through Grade 11, and Cohort Predicting Eleventh-Grade MEAP/MME Level

Step and predictor variable	<i>B</i>	<i>SE B</i>	β	<i>p</i>	R^2	ΔR^2	ΔF	<i>p</i>
Step 1					.10	.10	20.83	.000
Learning Disability	-1.220	.183	-.271	.000				
Gender	-.176	.082	-.087	.033				
Free/Red. Lunch	-.347	.103	-.136	.001				
Step 2					.32	.22	174.57	.000
8th-grade MEAP	.613	.046	.486	.000				
Step 3					.45	.13	63.65	.000
Highest level math course	.28	.04	.30	.000				
Number of math courses	.18	.04	.17	.000				
Step 4					.53	.09	33.15	.000
Cohort 2	-.40	.09	-.18	.000				
Cohort 3	-.77	.11	-.32	.000				
Cohort 4	-.89	.10	-.38	.000				

Step 1: The first step included the demographic variables (excluding residence).

The first line ("Step 1") indicates that these three variables explain ten percent of the variance ($R^2 = .10$), which means they constitute a model that is significantly better than a model with no predictors ($p = .000$).

The next three lines show that having a learning disability, being female, and receiving a free or reduced-price lunch were all significant negative predictors of eleventh-grade MEAP/MME Level. That is, students with a learning disability, who were female, or who received a free or reduced-price lunch tended to have lower-level scores on this standardized test. All of these results fit in with the findings from the means analysis and even the bivariate analysis, though the results for females were significant ($p = .003$) in the regression, whereas they were not significant in either of the other analyses.

Step 2: As a second step, eighth-grade MEAP was added in. This added approximately 22% (ΔR^2) to the model for a total of 32% (R^2). The contribution of eighth-grade MEAP to the model, once controlling for demographics, was significant ($p = .000$), which can also be seen from the following line, where the coefficient of eighth-grade MEAP is presented. The coefficient of 0.613 means that for every level higher on the eighth-grade MEAP, students achieved, on average, nearly two-thirds of level higher on the eleventh-grade MEAP/MME.

Step 3: The third step was to add coursework through Grade 11 (again, these were slightly different variables than those used in the above regressions, which went through Grade 12). These two variables added 13% to the model for a total of 45% of the variance accounted for. This contribution was significant ($p = .000$) even after controlling for the demographic variables and eighth-grade MEAP.

The two lines following Step 3 (those labeled highest-level math course and number of math courses) show that the coefficients for highest-level math course (.28) and for number of math courses (.18) were both strongly positive and significant

predictors of eleventh-grade MEAP/MME level (the coefficient of 0.28, for instance, means that, on average, for every level of highest math course, students achieved one-quarter higher level on the eleventh-grade MME/MEAP). As is shown below, however, number of math courses was generally not a significant predictor of standardized test results.

Step 4: The fourth step was to add cohort. The line "Step 4" indicates that cohort as a whole contributed 9% (ΔR^2) to the predictive value of the model. This contribution was significant ($p = .000$). Note that all the variables together accounted for 53%—approximately half—of the variance in eleventh-grade MEAP/MME scores, which makes the model as a whole relatively strong.

Since this was the p -value to be compared to .017 according to the Bonferroni discussion above, this result provided strong evidence of differences among cohorts, and particularly between Cohort 1 and the other cohorts, in eleventh-grade MEAP/MME level. However, as is shown in the subsequent three lines (Cohort 2, Cohort 3, and Cohort 4), as with number of math courses, the significance was in the "wrong" direction. All three cohorts had significant ($p = .000$) negative coefficients, indicating that students in these cohorts performed significantly worse on their eleventh-grade MEAP/MME than did students in Cohort 1. Worse still, unlike the coursework variable, in this regression each coefficient was larger negatively than the previous one, indicating that, after controlling for all the other variables, students actually performed worse on their eleventh-grade MEAP/MME test levels as time went on.

Summary: The hierarchical linear regression indicates that, after controlling for the demographics and eighth-grade MEAP level, there were significant differences

among the cohorts in eleventh-grade MEAP/MME levels, and these differences were in the "wrong" direction. There is no evidence suggesting that, beyond any effect on coursework, the new curriculum led to an improvement in eleventh-grade MEAP/MME levels.

Eleventh-Grade ACT

The results of the hierarchical regression for eleventh-grade ACT appear in Table 32.

It is crucial to note again that neither the ACT nor the WorkKeys was given to Cohort 1, so the only comparisons that are possible are to Cohort 2, which was after the reform curriculum was implemented. Nonetheless, we can see what changes occurred over the three (technically four, since there was a two-year gap between Cohorts 3 and 4) cohorts in which the curriculum was implemented. As I have mentioned, changes over the treatment cohorts are interesting to study because from Cohort 2 on they rise, although the other trend, which was that the means drop between Cohort 1 and Cohort 2, cannot be studied. However, because the ACT and WorkKeys were not given to Cohort 1, these were not included in the major hypotheses, and thus, not accounted for in the Bonferroni correction.

Table 32: Hierarchical Regression Analysis Summary for Demographic Variables, 8th-Grade MEAP, Coursework through Grade 11, and Cohort Predicting Eleventh-Grade ACT Level

Step and predictor variable	<i>B</i>	<i>SE B</i>	β	<i>p</i>	R^2	ΔR^2	ΔF	<i>p</i>
Step 1					.07	.07	11.42	.000
Learning Disability	-.94	.17	-.26	.000				
Gender	-.09	.08	-.05	.271				
Free/Red. Lunch	-.14	.10	-.07	.146				
Step 2					.27	.20	116.34	.000
Eighth-grade MEAP Level	.51	.05	.46	.000				
Step 3					.42	.15	58.04	.000
Highest-Level Math Course	.37	.04	.45	.000				
Number of Math Courses	.02	.04	.03	.586				
Step 4					.46	.04	14.90	.000
Cohort 3	-.38	.08	-.21	.000				
Cohort 4	-.36	.08	-.20	.000				

Step 1: The first step included the demographic variables (excluding residence). These three variables explain only seven percent of the variance ($R^2 = .07$), but they do constitute a model that is significantly better than a model with no predictors ($p = .000$). Having a learning disability, being female, and receiving a free or reduced-price lunch were all negative predictors of eleventh-grade ACT level. That is, female students, students with a learning disability, or students receiving a free lunch tended to have

lower-level scores on this standardized test. However, only learning disability was a significant predictor. All of these results are similar to the findings from the means analysis and bivariate analyses.

Step 2: As a second step, eighth-grade MEAP was entered. This added approximately 20% (ΔR^2) to the model for a total of 27% (R^2) of the variance explained. The contribution of eighth-grade MEAP to the model, once controlling for demographics, was significant ($p = .000$), which can also be seen in the following line, where the coefficient of eighth-grade MEAP is listed. The coefficient of 0.51 means that for every level higher on the MEAP students received, they achieved, on average, about one-half higher level on the eleventh-grade ACT.

Step 3: The third step was to add the two coursework variables, through grade 11 (again, these were slightly different variables than those used in the above regressions, which went through grade 12). These two variables added 15% to the model for a total of 42% of the variance accounted for. This contribution was significant ($p = .000$) even after controlling for the demographic variables and eighth-grade MEAP.

An interesting result, somewhat different than that found with regard to eleventh-grade MEAP/MME, can be seen in the two lines labeled highest-level math course and number of math courses. Whereas these two variables were highly correlated in the bivariate analysis (Spearman's rho was 0.68), when they were both in the regression analysis, highest-level math course had a strongly positive coefficient (the coefficient of 0.38 means that for every level of math course, students achieved about just over one-third higher level on the ACT) and was a much better predictor ($p = .000$) than number of math courses. Moreover, number of math courses, which was not even close to

significant, had a near-zero coefficient. These results suggest that once one knows what level math course a student reached, it is not important (in terms of ACT level) how many courses the student took to get there.

Step 4: The line "Step 4" indicates that cohort as a whole contributed only 4% (ΔR^2) to the predictive value of the model, but this contribution was significant ($p = .000$).

Since the ACT was not administered to Cohort 1, this regression was not included as a major hypothesis, and there is no critical p -value. Nonetheless, this result indicates strong evidence of differences among the post-reform cohorts, and particularly between Cohort 2 and the other two cohorts, in ACT level. However, the results shown in the next two lines of the table (for Cohort 3 and Cohort 4) suggest that, as with eleventh grade MEAP/MME level, the significance is in the "wrong" direction. Both cohorts have significant ($p = .000$) negative coefficients, indicating that students in these cohorts performed significantly worse on the ACT test than did students in Cohort 2, which was the first cohort in which the reform curriculum was implemented. Of course, that is after controlling for coursework, and it has already been shown that taking higher-level math courses led to better results on the eleventh-grade ACT. These results are somewhat different than those of the means analysis, which did not control for coursework and found almost no change in eleventh-grade ACT levels over the course of the study.

Summary: The hierarchical linear regression indicates that, after controlling for the demographics and eighth-grade MEAP level, there were significant differences among the cohorts, but these differences were in the "wrong" direction. There is no

evidence suggesting that the new curriculum led to an improvement – even over time – in eleventh-grade ACT levels.

Eleventh-Grade WorkKeys

The results of the hierarchical regression for eleventh-grade WorkKeys appear in Table 33. Note again that neither the ACT nor the WorkKeys was given to Cohort 1, so the only comparisons that are possible are to Cohort 2, which was after the reform curriculum was implemented. Nonetheless, it is possible to examine what changes occurred over the three cohorts (technically four, since there was a two-year gap between Cohort 3 and Cohort 4) in which the curriculum was implemented and, as has been previously shown, those changes are also interesting to study.

Table 33: Hierarchical Regression Analysis Summary for Demographic Variables, 8th-Grade MEAP, and Cohort Predicting Eleventh-Grade WorkKeys Level

Step and predictor variable	<i>B</i>	<i>SE B</i>	β	<i>p</i>	R^2	ΔR^2	ΔF	<i>p</i>
Step 1					.08	.08	11.91	.000
Learning Disability	-1.5	.29	-.24	.000				
Gender	-.23	.13	-.08	.079				
Free/Red. Lunch	-.35	.16	-.10	.030				
Step 2					.34	.26	169.75	.000
Eighth-Grade MEAP Level	.99	.08	.53	.000				
Step 3					.45	.12	45.97	.000
Highest-Level Math Course	.57	.07	.42	.000				
Number of Math Courses	-.03	.06	-.02	.683				
Step 4					.50	.05	19.20	.000
Cohort 3	-.57	.13	-.19	.000				
Cohort 4	-.77	.13	-.26	.000				

Step 1: The first step included the demographic variables (excluding residence). The first line ("Step 1") indicates that these three variables explain eight percent of the variance ($R^2 = .08$), which is only a small percentage, but this does constitute a model that is significantly better than a model with no predictors ($p = .000$).

The next three lines show that having a learning disability, being female, and receiving a free or reduced-price lunch were all negative predictors of eleventh-grade WorkKeys level. That is, students with a learning disability, who were female, or who

received a free or reduced-price lunch tended to have lower-level scores on this standardized test. However, only learning disability and free or reduced-price lunch were significant predictors. All of these results fit in with the findings from the means analysis and bivariate analyses.

Step 2: As a second step, eighth-grade MEAP was entered. This added approximately 26% (ΔR^2) to the model for a total of 34% (R^2). The contribution of eighth-grade MEAP to the model, once controlling for demographics, was significant ($p = .000$), which can also be seen from the following line, where the coefficient of eighth-grade MEAP is listed. The coefficient of 0.991 means that for every level higher on the eighth-grade MEAP students received, they achieved, on average, almost one level higher on the eleventh-grade WorkKeys test (to review, unlike the other tests, the WorkKeys test has six active scoring levels).

Step 3: The third step was to add coursework through Grade 11 (again, these were slightly different variables than those used in the above regressions, which went through Grade 12). These two variables added 12% of the variance explained to the model for a total of 45% of the variance accounted for. This contribution was significant ($p = .000$) even after controlling for the demographic variables and eighth-grade MEAP.

A finding similar to but stronger than what was found with regard to eleventh-grade ACT level appears in the two lines labeled highest-level math course and number of math courses. Whereas these two variables were highly correlated in the bivariate analysis (Spearman's rho was 0.68), when they were both in the regression analysis, highest-level math course was both strongly positive (the coefficient of 0.57 means that for every level higher of math course taken, students achieved, on average, slightly more

than one-half level higher on the WorkKeys test; to review, the WorkKeys levels ranged from 2 to 7 rather than from 1 to 4) and was a much better predictor ($p = .000$) than number of math courses taken. As was the case for ACT level, the coefficient for number of math courses was almost zero and not significant, suggesting again that, on average, if a student reached a certain level of math course, it hardly mattered (in terms of WorkKeys level) how many courses the student took to get there.

Step 4: The fourth step was to add in cohort. The line "Step 4" indicates that cohort as a whole contributed only 5% (ΔR^2) to the predictive value of the model, but this contribution was significant ($p = .000$).

Since this regression was not included as a major hypothesis, there was no critical p -value. Nonetheless, this result indicates strong evidence of differences between post-reform cohorts, and particularly between Cohort 2 and the other two cohorts, in WorkKeys level. However, the results shown in the next two lines of the table (for Cohort 3 and Cohort 4) suggest that, as with eleventh grade MEAP/MME and ACT levels, the significance was in the "wrong" direction. Both cohorts had significant ($p = .000$) negative coefficients, indicating that students in these cohorts performed significantly worse on the WorkKeys test than did students in Cohort 2, which was the first cohort in which the reform curriculum was implemented. Of course, that is after controlling for coursework, and it has already been shown that taking higher-level math courses led to better results on the eleventh-grade ACT. These results are somewhat different than those of the means analysis, which did not control for coursework, and found very little change in eleventh-grade WorkKeys levels over the course of the study.

Summary: The hierarchical linear regression indicates that, after controlling for the demographics and eighth-grade MEAP level, there were significant differences among the cohorts, but these differences were in the "wrong" direction. As with ACT, there was no evidence suggesting that the new curriculum led to an improvement, even over time, in eleventh-grade WorkKeys levels.

Summary of Analyses for Research Question 1

The regressions on coursework suggest that subsequent to an initial drop that occurred just after the start of the reform curriculum, coursework results gradually improved. The regressions on standardized test results show a continual drop in level throughout the course of the study. None of these results were statistically significant.

Research Question 2a: Lowest Two Eighth-Grade MEAP Levels

I was also interested in studying whether the reform improved outcomes for different levels of students, in particular for the weakest students. Therefore, I reran the hierarchical linear regressions for the students who had scored the lowest on the eighth-grade MEAP, namely, those whose eighth-grade MEAP level was 1 (approximately 7% of the sample) or 2 (approximately 18% of the sample). Both levels were included because only a small percentage of the students had a MEAP level of 1.

Because all students in the lowest two levels had non-adequate grades, there was not enough variance in this variable to make it a useful predictor. I therefore decided to remove eighth-grade MEAP from the hierarchical linear regression. In addition, due to the fact that the ACT and WorkKeys tests were not conducted for Cohort 1, so that in any case the results were difficult to interpret (in that they included only post-reform students), the hierarchical regressions were not rerun for these two tests.

Before rerunning the regressions, I ran crosstabs between the demographic characteristics and eighth-grade MEAP level. The results appear in Table 34.

Table 34: Demographic Characteristics of Participants, as Percentages by Eighth-Grade MEAP Level

Characteristic	Less than Proficient (1-2)	Proficient (3)	Advanced (4)	Total
Gender				
Male	54.4	48.7	53.6	51.2
Female	45.6	51.3	46.4	48.8
Free/Reduced Lunch*				
Yes	68.3	79.9	80.8	77.2
No	31.7	20.1	19.2	22.8
Special Needs				
Learning-disabled*	84.5	96.0	99.3	94.0
Not learning-disabled	15.5	4.0	0.7	6.0
Total	100.0	100.0	100.0	100.0

* $p < .01$.

As can be seen, there were not large differences between boys and girls on their eighth-grade MEAP scores. However, students with special needs were somewhat more likely to have lower level eighth-grade MEAP scores, and nearly two-thirds of students with learning disabilities were at the lower two eighth-grade MEAP levels.

Highest-Level Math Course Taken

The results of the hierarchical regression for highest-level math course appear in Table 35.

Table 35: Hierarchical Regression Analysis Summary for Demographic Variables and Cohort Predicting Highest-Level High School Math Course Among Students at the Two Lowest Eighth-Grade MEAP Levels

Step and predictor variable	<i>B</i>	<i>SE B</i>	β	<i>p</i>	R^2	ΔR^2	ΔF	<i>p</i>
Step 1					.19	.19	11.72	.000
Learning Disability	-.76	.23	-.25	.001				
Gender	.42	.17	.18	.012				
Free/Red. Lunch	-.57	.18	-.23	.002				
Step 2					.23	.04	2.26	.084
Cohort 2	.18	.21	.07	.389				
Cohort 3	.19	.24	.07	.424				
Cohort 4	.62	.24	.21	.010				

Step 1: The first step included the demographic variables. The first line ("Step 1") indicates that these three variables explain 19 percent of the variance ($R^2 = .19$) for students in the lowest two eighth-grade MEAP levels, and constitute a model that is significantly better than a model with no predictors ($p = .000$). The next three lines show that both having a disability and receiving a free or reduced-price lunch were negative predictors of highest-level math course. That is, even among students with low-level eighth-grade MEAP scores (who, as has been seen, were more likely to have a learning disability or receive a free or reduced-price lunch), those with a learning disability or receiving a free or reduced-price lunch tended to take significantly lower-level math courses (the p -values were .001 for learning disability and .002 for lunch). Being female was a significant ($p = .011$) positive predictor of highest-level math course with a coefficient of .44, indicating that girls tended to take almost half a level higher math courses than boys. These results were similar to those for the entire sample, though the difference between girls and boys was more striking for the lower-performing students.

Step 2: As the second step, cohort was entered. The line "Step 2" indicates that cohort as a whole contributed approximately 4% (ΔR^2) to the predictive value of the model, and that this contribution was not significant ($p = .084$).

Looking at the individual cohorts, one can see major differences between Cohort 4 and the other cohorts in the highest-level math course taken. The lines for the individual cohorts show that the results diverged from those of the whole sample. For students in the lowest two levels on eighth-grade MEAP, all of the coefficients were positive, which means that students in these cohorts performed better than students in Cohort 1, although the differences were small and not significant. For Cohort 4, the coefficient was not only positive but significant ($p = .010$), indicating that on average, low-level students in Cohort 4 took about two-thirds (.66) of a level higher math course as their highest-level math course than did low-level students in Cohort 1.

Summary: The hierarchical linear regression indicates that, after controlling for the demographics, there were significant differences among the cohorts, mainly in that low-level students in Cohort 4 took significantly higher-level math courses than did those in Cohort 1.

Number of Math Courses Taken

The results of the hierarchical regression for number of math courses taken appear in Table 36.

Table 36: Hierarchical Regression Analysis Summary for Demographic Variables and Cohort Predicting Number of Math Courses Taken Among Students at the Two Lowest Eighth-Grade MEAP Levels

Step and predictor variable	<i>B</i>	<i>SE B</i>	β	<i>p</i>	R^2	ΔR^2	ΔF	<i>p</i>
Step 1					.02	.02	.839	.475
Learning Disability	-.27	.25	-.09	.286				
Gender	-.05	.18	-.02	.808				
Free/Red. Lunch	-.20	.20	-.08	.326				
Step 2					.10	.08	4.67	.004
Cohort 2	-.49	.22	-.19	.032				
Cohort 3	-.72	.25	-.25	.006				
Cohort 4	.22	.26	.07	.394				

Step 1: The first step included the demographic variables. The first line ("Step 1") indicates that these three variables explain only two percent of the variance ($R^2 = .02$) and constitute a model that was not significantly better than one with no predictors.

The next three lines show that both having a disability and receiving a free or reduced-price lunch were slightly negative predictors of number of math courses (that is, on average, students receiving a free or reduced-price lunch took slightly fewer math courses). However, the coefficients were not significant.

Step 2: In Step 2, cohort was entered. The line "Step 2" indicates that cohort as a whole contributed 8% (ΔR^2) to the predictive value of the model, and that this contribution was significant ($p = .004$). However, this whole model explains only 10% (R^2) of the variance.

The promising results from the analysis for highest-level math course did not extend to number of math courses. The next three lines (Cohort 2, Cohort 3, and Cohort 4) indicate that the significance is in the "wrong" direction. Cohorts 2 and 3 had significant negative coefficients, indicating that students in these cohorts took fewer math courses than students in Cohort 1, and while Cohort 4 had a slightly positive coefficient (.22), indicating that students in Cohort 4 took more math courses than students in Cohort 1, this coefficient was not significant ($p = .394$).

While the results indicate a short-term decrease in number of math courses taken, the long-term results indicate a relative increase. That is, although there was a drop between Cohort 1 and Cohorts 2 and 3 (which has the largest negative coefficient, indicating that students in Cohort 3 took the fewest courses of all the cohorts), the positive coefficient for Cohort 4 suggests that students in this cohort took slightly more (.22) math courses, although not significant, than students in Cohort 1, and more than those in Cohorts 2 and 3.

Summary: The hierarchical linear regression indicates that, after controlling for the demographics, there were significant differences among the cohorts. Although the results were in the "wrong" direction, results for the individual cohorts, results suggest that perhaps over time, students tended to take higher-level math courses. In sum, for number of math courses taken, unlike for highest-level math courses, results for the lowest-scoring students were almost the same as those for the sample as a whole.

Eleventh-Grade MEAP/MME

The results of the hierarchical regression for eleventh-grade MEAP/MME appear in Table 37. Similar to the methodology for the entire sample, in the regressions for the

three eleventh-grade standardized test results, coursework through eleventh grade was added in as a second step in the regression.

Table 37: Hierarchical Regression Analysis Summary for Demographic Variables and Cohort Predicting Eleventh-Grade MME/MEAP Level Among Students at the Two Lowest Eighth-Grade MEAP Levels

Step and predictor variable	<i>B</i>	<i>SE B</i>	β	<i>p</i>	R^2	ΔR^2	ΔF	<i>p</i>
Step 1					.09	.09	3.81	.012
Learning Disability	-.49	.18	-.25	.005				
Gender	-.24	.13	-.17	.057				
Free/Red. Lunch	-.08	.14	-.05	.566				
Step 2					.14	.06	3.92	.022
Highest-Level Math Course	.03	.07	.04	.659				
Number of Math Courses	.21	.09	.22	.016				
Step 3					.24	.10	5.20	.002
Cohort 2	-.51	.14	-.31	.001				
Cohort 3	-.37	.18	-.20	.044				
Cohort 4	-.60	.18	-.31	.001				

Step 1: The first step included the demographic variables. The first line ("Step 1") indicates that these three variables explain nine percent of the variance ($R^2 = .09$), and constitute a model that is significantly better than a model with no predictors ($p = .012$).

The next three lines show that having a learning disability, being female, and receiving a free or reduced-price lunch were all negative predictors of eleventh-grade MEAP/MME level. That is, students with a learning disability, female students, and students who received a free or reduced-price lunch tended to have lower-level scores on this standardized test. However, only learning disability was a significant predictor

($p = .005$). Being female almost reached significance ($p = .057$). All of these results fit in with the findings from the means analysis and the bivariate analysis.

Step 2: The second step was to enter coursework through Grade 11 (again, these were different variables than those used in the coursework regressions, which went through Grade 12). These two variables added 6% to the model for a total of 14% of the variance accounted for. This contribution was significant ($p = .022$) after controlling for the demographic variables. Unlike most of the analyses for the sample as whole, in this analysis it was number of courses rather than highest level math course that was a significant positive predictor of eleventh-grade MEAP/MME level ($p = .016$). The coefficient of this variable (0.21) indicates that for each additional math course a student took, he or she scored approximately one-fifth level higher level on the eleventh-grade MEAP/MME.

Step 3: As the third step, cohort was entered. The line "Step 3" indicates that cohort as a whole contributed 10% (ΔR^2) to the predictive value of the model, and that this contribution was significant ($p = .000$). As with the sample as a whole, however, the next three lines (Cohort 2, Cohort 3, and Cohort 4) suggest that, as with number of math courses, the significance is in the "wrong" direction. All three cohorts had significant negative coefficients (significance levels ranging from $p = .044$ to $p = .001$), indicating that students in these cohorts performed significantly worse on their eleventh-grade MEAP/MME than did students in Cohort 1. There were not large differences among the cohorts; the coefficients were similar to each other, ranging from $-.37$ to $-.60$.

Summary: The hierarchical linear regression indicates that, after controlling for the demographics and eighth-grade MEAP level, there were significant differences

among the cohorts in terms of eleventh-grade MEAP/MME level, but these differences were in the "wrong" direction. There is no evidence suggesting that the new curriculum led to an improvement in eleventh-grade MEAP/MME levels for students at the lowest eighth-grade MEAP levels.

Summary of Analyses for Two Lowest Eighth-Grade MEAP Levels

The analysis for highest-level math course indicated that the reform curriculum might have been more beneficial for the lowest-level students than for the students in general. However, the results of the other two analyses (for number of math courses and standardized test) were substantially the same as the analyses for the sample as a whole.

Research Question 2b: Highest Eighth-Grade MEAP Level

In addition to the lowest scoring students, hierarchical linear regressions were also re-run for the highest scoring of the eighth-grade students, those whose eighth-grade MEAP level was 4 (approximately 22% of the sample).

The methodological considerations were similar to those of the analysis of the lowest-scoring students. It was decided to remove eighth-grade MEAP from the hierarchical linear regression, this time because all the students in the subsample had the same eighth-grade MEAP level (4). Similarly, since the ACT and WorkKeys tests were not conducted for Cohort 1, making the results more difficult to interpret, the hierarchical linear regressions were not rerun for these two tests.

Highest-Level Math Course Taken

The results of the hierarchical regression for highest-level math course appear in Table 38.

Table 38: Hierarchical Regression Analysis Summary for Demographic Variables and Cohort Predicting Highest-Level High School Math Course Among Students with Highest Eighth-Grade MEAP Level

Step and predictor variable	<i>B</i>	<i>SE B</i>	β	<i>p</i>	R^2	ΔR^2	ΔF	<i>p</i>
Step 1					.04	.04	1.84	.142
Learning Disability	-2.05	1.29	-.13	.115				
Gender	.24	.21	.09	.263				
Free/Red. Lunch	-.33	.27	-.10	.213				
Step 2					.07	.03	1.68	.174
Cohort 2	-1.35	.64	-.23	.036				
Cohort 3	-.84	.48	-.32	.082				
Cohort 4	-.67	.48	0.26	.161				

Step 1: The first step included the demographic variables. The first line ("Step 1") indicates that, among students with high-level eight grade MEAP scores, these three variables explained four percent of the variance ($R^2 = .04$), and constituted a model that was not significantly better than a model with no predictors. The next three lines show that both having a disability and receiving a free or reduced-price lunch were non-significant negative predictors of highest-level math course. That is, students with a learning disability or receiving a free or reduced-price lunch tended to take lower-level math courses. Being female was a non-significant positive predictor of highest-level math course. That is, girls tended to take slightly higher-level math courses than boys. These results were similar to those for the whole sample, though for the sample as a whole, the differences were significant, and for these students they were not.

Step 2: As the second step, cohort was entered. The line "Step 2" indicates that cohort as a whole contributed approximately 3% (ΔR^2) to the predictive value of the model, and that this contribution was not significant; this model explains only 7% of the variance in highest-level math course (R^2) among students with high-level eighth-grade MEAP scores.

There were major differences among the cohorts, and particularly between Cohort 1 and the other cohorts, in the highest-level math course taken. The lines for the individual cohorts show that the results were similar in direction to, but stronger than, those of the whole sample. For students at the highest level of eighth-grade MEAP, all of the coefficients were negative and larger than those for the sample as a whole, although only the coefficient for Cohort 2 was significant.

Summary: The hierarchical linear regression indicates that, after controlling for the demographics, there were significant differences among all the cohorts for students at the highest eighth-grade MEAP level. The general result was that students in Cohorts 2 through 4 took lower-level math courses than did students in Cohort 1, though only the difference with Cohort 2 was significant. In addition, the fact that the coefficients become less negative over time suggests that if the reform had a negative impact, it might have been short-term.

Number of Math Courses Taken

The results of the hierarchical regression for number of math courses taken appear in Table 39.

Table 39: Hierarchical Regression Analysis Summary for Demographic Variables and Cohort Predicting Number of Math Courses Taken Among Students at the Highest Eighth-Grade MEAP Level

Step and predictor variable	<i>B</i>	<i>SE B</i>	β	<i>p</i>	R^2	ΔR^2	ΔF	<i>p</i>
Step 1					.02	.02	1.10	.352
Learning Disability	-1.93	1.35	-.12	.153				
Gender	.22	.22	.08	.321				
Free/Red. Lunch	-.03	.28	-.01	.919				
Step 2					.11	.09	4.75	.003
Cohort 2	-1.48	.64	-.25	.023				
Cohort 3	-1.07	.48	-.40	.029				
Cohort 4	-.38	.48	-.14	.432				

Step 1: The first line ("Step 1") indicates that, among students with high-level eighth-grade MEAP scores, these three variables explained only two percent of the variance ($R^2 = .02$), and constituted a model that was not significantly better than one with no predictors.

The next three lines show that having a disability was a negative predictor of number of math courses taken, and that being female was a positive predictor (that is, girls, on average, took slightly higher math courses), but that neither of the coefficients was significant.

Step 2: As Step 2, cohort was entered. The line "Step 2" indicates that cohort as a whole contributed 9% (ΔR^2) to the predictive value of the model, and that this contribution was significant ($p = .003$), but that this whole model explains only 11% (R^2) of the variance.

However, the next three lines (for Cohort 2, Cohort 3, and Cohort 4) again suggest that the significance is in the "wrong" direction. All three cohorts had negative coefficients (though only those of Cohort 2 and Cohort 3 were significant, and not highly so), indicating that students in these cohorts took fewer math courses than students in Cohort 1.

Whereas the results indicate a short-term decrease in number of math courses, the long-term results indicate a relative increase. That is, although there was a drop between Cohort 1 and Cohort 2, the coefficient for Cohort 3 was slightly less negative, and the coefficient for Cohort 4 was the least negative, suggesting that students in this cohort took approximately the same number of math courses as did students in Cohort 1.

Summary: The hierarchical linear regression indicates that, after controlling for the demographics, there were significant differences among the cohorts. Although the results are in the "wrong" direction, looking at the individual cohort results does suggest that perhaps students, over time, took more math courses. In short, regarding number of math courses taken, results for the high-level students were almost the same as those for the sample as a whole.

Eleventh-Grade MEAP/MME

The results of the hierarchical regression for eleventh-grade MEAP/MME among students at the highest level on the eighth-grade MEAP appear in Table 40. Similar to the methodology for the entire sample, in the regressions for the three eleventh-grade standardized test results, coursework through eleventh grade was added in as a second step in the regression.

Table 40: Hierarchical Regression Analysis Summary for Demographic Variables, Coursework, and Cohort Predicting 11th-Grade MEAP/MME Level Among Students at the Highest Eighth-Grade MEAP Level

Step and predictor variable	<i>B</i>	<i>SE B</i>	β	<i>p</i>	R^2	ΔR^2	ΔF	<i>p</i>
Step 1					.07	.07	3.30	.022
Learning Disability	-2.05	.79	-.22	.011				
Gender	.01	.14	.01	.952				
Free/Red. Lunch	-.33	.17	-.16	.062				
Step 2					.33	.26	24.66	.000
Highest level math course	.42	.08	.44	.000				
Number of math courses	.11	.06	.16	.056				
Step 3					.41	.08	5.45	.001
Cohort 2	-.38	.35	-.11	.275				
Cohort 3	-.92	.26	-.56	.001				
Cohort 4	-.88	.27	-.54	.001				

Step 1: The first step line ("Step 1") indicates that the three demographic variables explain seven percent of the variance ($R^2 = .07$), and that they constitute a model that is significantly better than a model with no predictors ($p = .022$).

The next three lines show that having a learning disability and receiving a free or reduced-price lunch were negative predictors of eleventh-grade MEAP/MME level. That is, students with a learning disability or -receiving a free or reduced-price lunch tended to have lower-level scores on this standardized test. However, only learning disability was a significant predictor ($p = .005$). Being female had almost no predictive value for eleventh-grade MEAP/MME level, among students at the highest level on the eighth-grade MEAP. These results were similar to those for the whole sample.

Step 2: The second step was to add coursework through Grade 11. These two variables added 26% to the model for a total of 33% of the variance accounted for. This

contribution was significant ($p = .000$) even after controlling for the demographic variables. As with the sample as a whole, for the highest-level students, highest-level math course and number of math courses were both positive predictors, though only highest-level math course was significant ($p = .000$). These results indicate that for these students, as for the sample as a whole, both taking higher level courses and taking more math courses had a positive impact on their eleventh-grade standardized test scores (as represented by the eleventh-grade MEAP/MME).

Step 3: As the third step, cohort was entered. The line "Step 3" indicates that cohort as a whole contributed 8% (ΔR^2) to the predictive value of the model, and that this contribution was significant ($p = .001$). As with the sample as a whole, however, the next three lines (Cohort 2, Cohort 3, and Cohort 4) suggest that, as with number of math courses, the significance is in the "wrong" direction. All three cohorts had negative coefficients (with Cohort 3 and Cohort 4 significant, both at $p = .001$), indicating that students in these cohorts performed significantly worse on their eleventh-grade MEAP/MME than did students in Cohort 1. Moreover, the negative coefficients grew larger from Cohort 2 to Cohort 3, though the coefficients for Cohorts 3 and 4 were approximately equal.

Summary: The hierarchical linear regression indicates that, after controlling for the demographics and eighth-grade MEAP level, there were significant differences among the cohorts in eleventh-grade MEAP/MME level for students at the highest level of eighth-grade MEAP, but these differences were in the wrong direction. There is no evidence suggesting that the new curriculum led to an improvement in eleventh-grade MEAP/MME levels.

Summary of Analyses for the Highest Eighth-Grade MEAP Level

The regression analyses suggest that the reform curriculum was no more beneficial for, and may have had more of a negative impact on, the students at the highest level than for the sample as a whole.

Chapter Summary

The main variable of interest in this research was cohort, which represented the reform curriculum. This summary focuses on the three major outcomes that were available for all four cohorts in the study: highest-level math course taken by students, the number of math courses taken by students, and students' eleventh-grade MEAP/MME levels. Results for the other two eleventh-grade standardized tests were not available for the first cohort. The results for all three eleventh-grade standardized tests were similar, so that the results for other two tests might well be similar to those for the eleventh-grade MME/MEAP. In addition to the results for the entire sample, the results for the weakest and strongest students (as measured by eighth-grade MEAP level) are summarized here.

Highest-Level Math Course Taken

There were interesting results for highest-level math course. For the sample as a whole, the regression coefficients for Cohorts 2 and 3 were weakly negative, indicating that students in these cohorts took lower-level math courses than students in Cohort 1. The coefficient for Cohort 4 was slightly, though not significantly, positive. These results suggest that whereas the reform curriculum might have had a small negative impact at first on the level of math courses taken, the longer-term effect might be positive.

For the strongest students, the coefficients for Cohorts 2, 3, and 4 were all negative, and also became progressively less so (though the coefficients for Cohorts 3 and 4 were somewhat similar). However, the coefficients were much larger than in the regression for the sample as a whole, suggesting that the reform curriculum might have had a larger negative impact, in terms of level of math courses taken, on the strongest students than on the students overall.

On the other hand, for the weakest students, the coefficients were positive, with the largest for Cohort 4. This result suggests that for the weakest students, the reform curriculum had a positive effect on level of math courses taken, and that this effect increased over the course of the study.

Number of Math Courses Taken

The findings in terms of number of math courses taken by students were also of interest, although less so than the findings for highest-level math course taken. In all three regressions (that is, the regression for the whole sample, the regression for the strongest students, and the regression for the weakest students), Cohorts 2 and 3 had negative coefficients, meaning that students in Cohorts 2 and 3 took fewer math courses than students in Cohort 1. The coefficients for Cohort 4 in the sample as a whole and for the strongest students were very slightly negative, while the coefficient for the weakest students were very slightly positive, suggesting an increase in number of math courses taken, and slightly more math courses taken in Cohort 4 than in Cohort 1. The effects in Cohorts 2 and 3 were more negative for the strongest group, and less negative for the weakest group, than for the sample as a whole. That is, the reform appeared to have less of a negative effect on the number of courses taken by the weakest students, and more of

a negative effect on the number of courses taken by the strongest students, than on the sample as a whole.

Eleventh-Grade MEAP/MME

The findings for eleventh-grade MEAP/MME differed from the other findings. After controlling for coursework, all coefficients for cohort were negative in all three regressions. However, unlike the results for coursework, the coefficients became more negative as the study progressed, such that (with the partial exception of the strongest students) the most negative coefficients were obtained for Cohort 4. Similar to the results for coursework, the negative coefficients were generally larger for the stronger students, and smaller for the weaker students, than for the sample as a whole, though this was not totally consistent.

In the final chapter, I provide a synthesis of the findings from my study relative to limitations, implications, and general conclusions.

CHAPTER 5: INTERPRETATION OF FINDINGS AND IMPLICATIONS FOR PRACTICE, POLICY, AND RESEARCH

The purpose of this research study was to evaluate the impact of a longitudinal mathematics reform effort conducted at Midwest High School. Leaders at Midwest High School implemented two purposeful interventions designed to 1) de-track students' placement into secondary mathematics courses and 2) use standards-based secondary mathematics curricula developed by the National Science Foundation. That is, Midwest de-tracked its internal structure for stratification of secondary students in Algebra, Geometry, and Algebra II mathematics courses while implementing NSF curricula.

The following research questions guided the study:

1. How do mathematics outcomes compare pre- and post-reform?
 - 1a. How does the mathematics course taking of Midwest high school students compare pre/post-reform?
 - 1b. How does mathematics achievement as an outcome compare pre/post reform?
2. How do different types of students fare in the reform?
 - 2a. How do students whose prior math achievement in eighth grade indicates lowest proficiency fare when compared pre/post reform?
 - 2b. How do students whose prior math achievement in eighth grade indicates highest proficiency fare when compared pre/post reform?

I examined these research questions using data collected from four cohorts of high school students. Cohort 1 (2003-2007) represented students who received no treatment beyond the traditional mathematics program for middle school and high school. These students were stratified for their middle and high school coursework according to their mathematics achievement in elementary and middle school.

Cohort 2 (2005-2009) represented the first group of students involved in the intervention, or mathematics reform. With the exception of approximately 55 students still stratified for honors instruction (the last students remaining from the de-tracking

elimination process phased in starting in grades 6-8), this cohort consisted of students who entered as ninth graders during year two of the secondary reform and participated, at minimum, in heterogeneously grouped Algebra I and Geometry instruction taught with standards-based NSF materials. At the time, Cohort 2 students were required to earn a minimum of two mathematics credits to graduate.

Cohort 3 (2006-2010) represented the second study group of students who entered as ninth graders during year three of the secondary reform effort in 2006 and were taught from a reform oriented, standards-based mathematics (NSF) program within heterogeneous groupings for Algebra I and Geometry. Like Cohort 2, graduation requirements for mathematics remained at two math credits. This group participated in a traditional middle school experience in grades 6-8. This group was of special interest as it was the first group of students entering as ninth graders to be placed in fully implemented heterogeneous groups, including even those very highest achieving students for whom ability grouping still existed in Cohorts 1 and 2 as tracking became phased out one year at a time in middle school.

Cohort 4 (2008-2012) represented students who entered as ninth graders during year five of the secondary reform effort in 2008 and participated in heterogeneously grouped Algebra I, Geometry, and Algebra II taught with standards-based curriculum. For the first time ever, state mathematics high school graduation requirements in Michigan increased from two to four credits. This state legislation also required that for the first time, students enroll in a mathematics class during their senior year. This cohort also participated in three full years of reform oriented, standards-based middle school mathematics in heterogeneous groupings in grades 6-8.

In the following summary and discussion of findings section, I begin with a discussion of the descriptive statistics findings followed by a summary and discussion of each research question. Then, I present some of the general conclusions of the study, limitations of the study, implications of the study, and some recommendations for further research based on my findings.

Summary of Descriptive Statistics Results

Overall, with the exception of residence, there were no significant differences in demographic characteristics between the four cohorts. In all of the outcome variables, the out-of-district students tended to have slightly “better” outcomes, but none of the differences were significant.

In terms of gender, results differed for coursework and standardized tests. Girls did somewhat better than boys on the coursework variables, but there was virtually no difference between them in terms of standardized tests.

Discussion of Selected Descriptive Statistics Results

Information about enrollment for non-resident students is interesting and worthy of note. Midwest High School gained enrollment of school of choice students from 17.1%, or about one in six students in cohort 1 to 31.5%, or nearly one-third of students in cohort 4. This statistically significant difference gives evidence that over the course of time, more students chose to attend Midwest High School. This is important to keep in mind because as noted in the case description in Chapter 2, parents indicated in surveys prior to the reform that a strong traditional academic program was one of the reasons that they were drawn to send their children to Midwest as non-resident families. The increase in non-resident students each year provided some indication that confidence was not

impacted even though the school was in the midst of a major reform in the secondary mathematics program. In fact, there is some reason to believe that non-resident families had a heightened desire to enroll, although it is not clear what the particular factors were that may have contributed to this increase.

For all of the outcome variables, the out-of-district students fared slightly “better,” but none of the differences were significant. This slight improvement, although not significant, is still important to note because of the unique proportion of school of choice students who chose to attend Midwest. Remember that the percentage of students attending Midwest High School in Cohort 4 had reached nearly 30%. There were times when school personnel and also parents or community members made generalized statements that reflected a negative stance toward this large population. It continues to be common to hear statements such as, “If those non-resident students would just go back to their own district we would be much better.” These findings provide some evidence that in the area of secondary mathematics achievement, non-resident students had no effect on the overall achievement for Midwest High School.

In summary of the findings regarding gender, girls took about one-third of a mathematics course level higher (3.48) than boys (3.86) and this was significant. Although the difference was not large enough to be significant, girls took slightly more math courses than boys, 3.48 versus 3.35. Girls did somewhat better than boys on both coursework variables but there was no difference between girls and boys on standardized tests (MEAP/MME, ACT, and WorkKeys). One speculation is that at Midwest High school, girls saw advanced mathematics courses as viable opportunities, even when these courses were not required.

For students with disabilities, the average highest level math course was 2.66 and for students without disabilities, 4.20, nearly one and a half levels higher. The difference was significant. Students with disabilities took an average of 2.86 math courses in high school compared to students without disabilities taking an average of 3.49 courses. The difference was significant.

Students with disabilities had significant differences ($p = .001$) for all standardized tests when averages were compared with students without disabilities. On the HST/MME, which uses a 4-level system of proficiency with level 1 being the lowest category, students with disabilities scored approximately one level lower (1.05) than students without disabilities (2.30). On the ACT, which also uses a 4-level system of proficiency with level 1 being the lowest category, students with disabilities scored approximately one level lower (1.17) than students without disabilities (2.15). Students with disabilities scored more than one and a half performance levels lower average (out of six) on the WorkKeys than students without disabilities (3.07 v 4.70). These results should be reviewed in light of the fact that students with disabilities were held to the same performance standards as those without disabilities for state accountability purposes. In summary, results of the study show that students with disabilities took lower levels of math than their peers and also fewer math courses. This continues Midwest High School's prior achievement record; students with disabilities fare worse than their peers.

Students receiving free or reduced-price lunches took significantly lower level courses and fewer courses than did other students. Comparing the highest level math course between students with free or reduced-price lunches (3.29) and those with regular-

priced lunches (4.26) showed about one level of difference, which was significant. Students with free or reduced-price lunches took fewer average level courses (2.65) than students with regular-priced lunches (3.55) which was also significant. This, too, reinforces Midwest High School's prior achievement record.

Middle School Results

In spite of the fact that this study did not focus on the mathematics reform taking place simultaneously at Midwest Middle School, the data collected on the eighth-grade mathematics MEAP as an entry point for baseline achievement proved to represent information worthy of discussion. Midwest Middle School began heterogeneous grouping in mathematics classes beginning in 2004 for its incoming sixth graders. At the same time, they implemented an NSF-developed mathematics curriculum: Connected Mathematics. The reform was introduced to all sixth graders and lowest-performing math students (still grouped homogeneously) in 2005, and by 2006 all grades 6-8 mathematics classes were undergoing the full reform.

Eighth-grade math MEAP scores rose fairly dramatically between Cohorts 1 and 4. The number of students graded "not proficient" or "partially proficient" between Cohort 1 (42.4%) and Cohort 2 (22.5%) decreased, meaning fewer students were seen in the lowest performance levels. Further, there was an increase of more than 30% in the number of students scoring in the top performance category, "advanced," between Cohort 2 (4.2%) and Cohort 3 (37.8%). The new MEAP scale score changes at the state level that occurred during the test administration for Cohort 2 likely contributed to this upswing, but in order to determine if the positive changes in eighth-grade scores over time were a result of treatment rather than simply due to changes in state scale scores, I

compared the scores with those of the state. This comparison confirmed that achievement in Midwest Middle School mathematics improved over time. Eighth-grade MEAP levels for Midwest Middle School were slightly below state averages for Cohorts 1 and 2 but were above state averages for Cohorts 3 and 4. Performance changed from below state average within the first few years of treatment to above state averages after several years of treatment. In sum, eighth-grade mathematics achievement at Midwest Middle School assessed with state measures improved over time when compared with overall state performance.

Summary and Discussion of Research Question Results

Research Question 1

After controlling for the demographics and eighth-grade MEAP level, I found there were no significant differences among the cohorts, or, indeed, between Cohort 1 and any of the other cohorts, in terms of highest-level math course taken. However, this finding can be interpreted differently if the course leveling system used in my research design is further considered using a different model which might more adequately represent the difficulty of courses as the reform actually played out. The course leveling system I developed for my study used a traditional conception of difficulty using the basic mathematics hierarchy familiar to most: *General Math*, *Algebra*, *Geometry*, *Algebra II*, etc. In retrospect, this hierarchy failed to account for the increase in difficulty of the same course as a result of the new curricula. Let me illustrate this important point. College Prep Algebra 1 taught using the traditional curriculum likely centered on procedural knowledge. This same course taught using the NSF curriculum centered on conceptual knowledge. This difference, I maintain, changed both the nature of the content

as well as the rigor of the mathematics content. Therefore, the difficulty level of the Algebra course taught in 2005 (after one year of the reform) may have been more accurately represented if I had used, say, a difficulty level of 2.5 rather than holding the original coding of 2 constant. In other words, Algebra 1 for Cohort 1 was an easier Algebra course than Algebra 1 for Cohorts 2-4. This line of reasoning might perhaps provide additional logic for the downward movement in the findings for course taking outcomes, particularly since I found strong decreases between Cohorts 1 and 2 where courses with similar titles would have reflected a vast shift in difficulty. This also may explain the slight upward trend for Cohorts 3 and 4 as course difficulty would have begun to plateau over time.

After controlling for the demographics and eighth-grade MEAP level, there were significant differences among the cohorts in terms of the number of math courses taken by students. While the results were in the wrong direction, examining the individual cohort results did suggest that perhaps, subsequent to the reform, the cohorts improved over time on number of math courses taken. This finding, too, may be partially explained by the coding thoughts mentioned above, but should also be situated within the context of increased mathematics credit and course requirements affecting only Cohort 4. I return to the legislation changes impacting Cohort 4 later in this chapter.

There was no evidence suggesting that the new curriculum or changes in grouping practices led to an improvement in eleventh-grade MEAP/MME levels. The hierarchical linear regression indicated that, after controlling for the demographics and eighth-grade MEAP level, there were significant differences among the cohorts in eleventh-grade MEAP/MME levels, but these differences were in the wrong direction.

There was also no evidence suggesting that the new curriculum or student grouping practices led to an improvement in eleventh-grade ACT levels. The hierarchical linear regression indicates that, after controlling for the demographics and eighth-grade MEAP level, there were significant differences among the cohorts, but these differences were in the wrong direction.

Finally, there was no evidence suggesting that the new curriculum or student grouping practices led to an improvement in eleventh-grade WorkKeys levels. The hierarchical linear regression indicates that, after controlling for the demographics and eighth-grade MEAP level, there are significant differences among the cohorts, but these differences are in the wrong direction.

In all three of the outcome measures explored for Cohort 1 (highest-level math course, number of math courses, and eleventh-grade MEAP/MME), there was a drop between Cohorts 1 and 2, and then a moderate rise between Cohort 2 and Cohort 3. In the case of the coursework outcomes, there was a further rise between Cohort 3 and Cohort 4, and Cohort 4 was even higher than Cohort 1. In the case of the standardized tests, Cohort 4 results were similar to (and in the case of WorkKeys, even a bit lower than) those of Cohort 3.

When comparing high school eleventh- grade HST/MME standardized test scores (reported in performance levels ranging from 1 to 4) from Midwest with those for the state as a whole, mean performance levels remained fairly constant. Midwest and the state scored in the range of 2.2 to 2.3 for all years of the study. The MME score for Midwest in Cohort 2 was the only exception, when the average proficiency level dropped from Cohort 1 to Cohort 2 (from 2.33 to 2.10). Although state scores from Cohort 1 to

Cohort 2 also dropped, the negative change was only .09 (from 2.30 to 2.21). No changes in MME were significant.

Research Question 2

The impact of both elements of the reform (heterogeneous grouping and NSF curriculum) appear to have had a slightly negative impact on the achievement of students scoring in the highest quartile on the eighth-grade MEAP test (approximately 22% of the sample) as incoming ninth graders at Midwest High School.

Regarding the highest achievers, students in Cohorts 2 through 4 took significantly lower level math courses than did students in Cohort 1, as the negative regression coefficients indicated (-1.35, -.84, and -.67). Course difficulty was indicated on a scale of 1-7 with courses such as *Functional Math* and *General Math* at level 1, *Geometry* at level 3, and *AP CalculusBC* at level 7. Since all of the coefficients were negative, it seems that overall, the highest-achieving students in the post-reform cohorts fared worse than did the highest achievers in Cohort 1.

Program effects for the highest achieving students in all post-reform cohorts had negative coefficients (-1.48, -1.07, and -.38), with the differences between Cohort 1 and Cohort 2, and between Cohort 1 and Cohort 3, being significant for the number of math courses taken. The difference between Cohort 1 and Cohort 4 was not significant. These results give some evidence that cohorts improved with time, as Cohort 4 was higher than both Cohorts 2 and 3. These results can be misleading, however, as more mathematics courses may not signify the most beneficial results. I further discuss the topic of number of courses taken later in the implications section of this chapter.

This subgroup of the advanced students did not improve over the course of treatment when eleventh- grade MME scores were examined after controlling for demographics. In fact, the negative coefficients grew stronger from Cohort 2 to Cohorts 3 and 4, with Cohorts 3 and 4 being similar and significantly different from Cohort 1.

I now shift the discussion to how well students at the low end of the spectrum fare. I highlight the findings for the students achieving in the lowest two levels of the eighth-grade MEAP (level 1 included approximately 7% of the sample and level 2 included about 18%) combined, as all such students were considered not meeting the state benchmark.

The results for this subpopulation of students diverged from those of the whole sample. On average, lowest achieving incoming ninth grade students in Cohort 4 took about two-thirds (.62) of a level higher math course than did those in Cohort 1. When demographics were controlled, all coefficients were positive for these students across Cohorts 2, 3, and 4 (.18, .19, and .62) when compared to Cohort 1. For Cohort 4, the coefficient is not only positive but significant.

The results for this lowest quartile were different when I examined the results for the total number of math courses enrolled. Cohorts 2 and 3 had negative coefficients (-.49, -.72), meaning that these students took fewer math courses than did students in Cohort 1. The coefficient for Cohort 4 compared to Cohort 1 was positive (.22), suggesting more courses, but it was not significant. In summary, for students in the lowest quartile, the number of courses taken during the intervention first decreased and then increased, with the differences between Cohort 1 and Cohorts 2 and 3 being significant.

Interestingly, unlike the sample as a whole, for the lowest-achieving students, taking more or higher level courses appear to make little difference in their eleventh-grade test scores. After controlling for demographic variables, highest level math course (.03) and number of math courses (.21) added together account for only 6% of the variance. Unusually, the coefficient for number of math courses was significant. When cohort effect was added to the regression, negative coefficients resulted (-.51, -.37, -.60), and all were significant. This slight negative decrease in eleventh-grade MEAP/MME levels for this subgroup over the course of the study was in the same direction as the means analysis, where a slight though not significant decrease in eleventh-grade MEAP/MME levels was found as well.

General Conclusions

In light of the findings summarized above, several general conclusions can be made.

First, data from Midwest Middle School suggest that positive changes occurred, at least when student achievement was measured using an assessment whose content remained consistent over the course of the years of the study. Eighth-grade mathematics MEAP scores strengthened over time, moving from below state average for incoming ninth-graders in Cohorts 1 and 2 to above state average for incoming ninth-graders in Cohort 3 and Cohort 4. Although it is not clear what exact variables contributed to this positive change in standardized test scores, it is clear that students were grouped more equitably than prior to the reform and received instruction using an NSF-developed middle school mathematics program within an environment that supported faithful and authentic implementation. Teachers at Midwest Middle School were provided with

support through ongoing professional development that included substantial time for teachers to strengthen their own knowledge to better understand fundamental mathematics concepts. They were provided with regularly scheduled common planning days to encourage teachers to work together to reflect on prior lessons and to think carefully about units and lessons ahead of time. Teachers were encouraged to implement the NSF program with fidelity and to avoid any tendencies they might have to make autonomous decisions about the use of supplemental mathematics materials, to change the order of mathematics units or lessons, or to alter the program's suggested pace. Based on the changes that started at Midwest Middle School in the fall of 2005, it appears as though students at Midwest Middle School benefitted from more equitable mathematics course placement practices and a more cohesive mathematics curriculum, at least when state assessment results are used as the measure.

Like Midwest Middle School, the student placement practices for Midwest High School also changed and the instructional materials used became more coherent. In other words, students entering as ninth graders beginning in 2004 were enrolled in fewer academic tracks than prior. Ninth-grade mathematics course options for lower-level mathematics classes (*General Math*, *Functional Math*, *Business Math*, *Intro Algebra*) were decreased. Nearly all ninth graders were placed in Algebra or above beginning in 2004, the first year of the reform. Curriculum coherence likely improved as a result of the elimination of multiple versions of entry level courses and also likely improved as a result of a conscious attention by leaders to ensure that the new NSF-developed high school mathematics program was implemented with fidelity by all teachers. Teachers were expected to focus on the curriculum, as intended, and participate in the ongoing

professional development provided, which included an explicit component of time for teachers to engage in uninterrupted common unit and common lesson planning.

In spite of the increase in equitable course placements and the increase in cohesion for the curriculum, Midwest High School did not see the same test score improvements as Midwest Middle School. In fact, scores for most students, in most all subcategories, remained the same or slightly decreased. Even though Midwest High School's scores were closely in line with the trends among state averages, reasons for the lack of improvement are difficult to pinpoint. As I explained in Chapters 3 and 4, I was concerned during the development in my research design that the high school state assessments might not necessarily align with the content and focus of the NSF-developed mathematics programs. I included an alternative measure of achievement (a district-developed test) to remedy this concern. Due to test administration irregularities, I was unable to rely on the results of this assessment and therefore excluded them from my study. Aside from the apparent reasons to include this alternative assessment because it aligned with the changes in the instructional program, another important reason was the fact that it was administered during the spring of tenth grade as opposed to the state tests that are administered during the spring of eleventh grade. An earlier assessment of progress at the conclusion of tenth grade would have allowed teachers to analyze the results to inform their decisions about instructional practices after only two years of instruction. It was an unfortunate setback to my research that I had to discard the results from this instrument. Because of these reasons, the state assessment instrument became my only assessment tool.

My study was negatively impacted by my need to rely so heavily on state assessment measures. The negative impact was deepened when the state's format and content changed midway through the reform. This shift in content caused further disparities in alignment between the focus for the NSF program and the state test. Additionally, the new state test reflected a different purpose. Midwest High School resides in a state which adopted the ACT and WorkKeys as the official eleventh-grade high school assessment. These nationally normed assessments are strictly controlled. Whereas the former high school test (HST) used by the state prior to 2006 released specific test items to schools so that they could be used to analyze results in conjunction with instructional decision making, the ACT and WorkKeys did not. These national tests provide only general scores to districts with subcategory scores for the different strand of mathematics. The grades 3-8 MEAP test (and the formerly used grade 11 HST), provide important information to district which can be used to influence teaching practices and assist district leaders in making relevant programmatic decisions based on data. The ACT was not intended to provide formative data to districts. The change in the content and format of state tests plus the limited assessment information available about specific student results made it difficult for Midwest High School to monitor progress or make programmatic adjustments and refinements during the reform.

I also included two measures of high school course taking (number of mathematics courses taken and highest mathematics course level) to serve as additional proxies for mathematics achievement in my study. In general, both of these measures indicated that the program of reform had very little, if any, impact. When regression analyses controlled for demographic variables and other predictors, the highest-achieving

students showed a slight decline according to these measures. That is, highest-achieving students took slightly lower-level mathematics courses and slightly fewer courses. Conversely, the lowest-achieving students in Cohort 4 took nearly two-thirds of a level higher mathematics course and nearly two-thirds of a course more mathematics courses. Thus, the reform might have helped the lowest-achieving students.

As noted in Chapter 1, there is some research that is less positively inclined toward the use of heterogeneous grouping for all students. Although the results of this study showed only slightly negative but not significant findings for the highest achieving students, these findings would align with the studies that recommend the best and brightest be grouped with like peers for instruction (Gamoran & Berends, 1987).

The results were compromised by unpredicted complications which I discuss in the following section on research limitations.

Limitations of Study

This study was limited by several factors. These included limitations related to data collection (data entry), inconsistent course offerings (upper level math courses), changes in outcome variables (state assessments), legislative changes in high school mathematics requirements, and confounding variables.

Data Collection

The data for this study came from district and county level electronic databases which were developed during the timeframe of this study. The development of a county-wide database was initiated by new state information collection requirements. This new system required information typically stored through hard copy records to be transferred by office personnel into electronic databases. Several individual interpretations made by

individuals at the school or district level contributed to an end product that included significant inconsistencies. For example, during the process of transferring course title information from CA-60s (student information files) to new coding categories in the electronic system, at times the original title was incorrectly entered. In some cases, mathematics courses were coded with vastly different meanings. One such example was in the case of the Midwest High School course, *Discrete Mathematics/Statistics*, coded into the electronic system using the available option most closely associated with the title (in this case the option most closely related was simply, *Math*). This error meant that an advanced mathematics course was coded as one of the lowest level math courses. Other examples of miscoding done at the local level included entering information about *Intro Algebra* as *Algebra* and *College Preparatory Algebra* as *Algebra* even though at the time of data entry, these courses were taught using different program materials with differing expectations and differing pacing. These inconsistencies between the math course title, the intended course content and difficulty, and the final category or code entered for the course in the data warehouse contributed to unforeseen complexities in my research study. While I made every effort to examine and clarify any inconsistencies discovered across the study, I am still not certain whether I was able to identify and correct every original data entry error.

Mathematics Courses

Complexities associated with mathematics course offerings changing over time provided additional limitations for the study. Like many high schools, Midwest High School offered particular mathematics courses dependent upon the yearly budget, predicted enrollment figures, available certified staff, number of students interested

and/or qualified to enroll, number of course sections needed for a required course, number of special education students needing a co-teacher for a required course, and a host of other variables that most superintendents and high school principals must consider regularly.

Beginning in 2004 (impacting Cohort 2), several prior options for level 1 mathematics courses were eliminated. These changes in the lowest level mathematics course offerings remained constant from 2004 on. These changes would likely explain the slight downward movement in the number of courses taken by Cohort 2 students when compared to Cohort 1. To review, a typical pattern of course taking for Cohort 1 students included ninth grade enrollment in *Intro Algebra* followed by enrollment in *Algebra* in tenth grade. Many students with grades of “B” or better in Cohort 1 followed this traditional two course sequence by enrolling in two subsequent mathematics courses simultaneously to make sure they were able to enroll in the most advanced mathematics courses as juniors and seniors. By removing *Intro Algebra* as an option, the former two course sequence now took only one year to complete. Predictably, for many students, this change meant one less total mathematics course taken.

The inconsistencies from year to year for courses leveled in the 3-7 range posed additional challenges to my ability to make conclusions for the highest achieving students. Let me explain.

Pre-calculus/Trigonometry (course difficulty coded at level 5) was offered one year but was replaced by *Trigonometry* (level 6) the next. The next year *Calculus* was offered (level 6) along with *Discrete Mathematics/Statistics* (level 5), but the following year the *Discrete Mathematics* component of the title was removed, and presumably

removed for the course content as well, with purely *Statistics* being offered. *AP Calculus BC*, a full-year companion course to *AP Calculus AB*, was only offered for Cohort 4.

Important to note, is that for most courses at levels 3-6, the class composite was regularly made up of students from two or more grade levels (eleventh and twelfth graders with some few tenth graders). These realities made the mathematics course options for students from one cohort to another generally inconsistent. These inconsistencies made it very difficult to draw conclusions about the outcome variable of the average highest level math courses in which a student enrolled. In short, upper level mathematics course offerings changed from year to year at Midwest High School.

Changes in State Assessments

As I noted earlier, one of the most serious concerns I faced in discerning whether or not the treatment was successful resulted from the continued changes that occurred in the state assessment program after the first year of the study was completed. This is a major limitation for my study. Some of these changes are highlighted as follows. The Michigan Curriculum Frameworks were introduced in early 2000 and were formally assessed in the fall of the 2001-2002 academic year and yearly thereafter. Test results were reported using scale scores determined by state-level officials and were further refined by “cut scores” which included four levels: Exceeds Standards, Meets Standards, At Basic Level, and Apprentice. The state required that grades 4, 5, 7, and 8 participate in yearly testing. Mathematics was tested at the fourth and seventh grades at that time (at this time, all states were required to assess at least once in grades 4 and 7)

From 2003-2005, Michigan used the High School Test (HST) to assess the performance of student achievement on the core content of the Michigan Curriculum

Frameworks, administered yearly to students in Grade 11. (At this time, NCLB required states to assess at least once in Grade 11). The HST included several constructed response items (requiring short written answers), several extended response items (requiring longer written answers), and also included multiple choice items. Students wrote considerably long responses in essay form for reading, writing, and social studies. The math portion of the test included several problems requiring students to reason and problem solve and expected students to draw a picture or diagram and use words to demonstrate their thinking skills for several multi-step problems. These assessments aligned well with the process standards within the Michigan Curriculum Frameworks. These questions also aligned with the kind of mathematical tasks taught within the NSF-developed programs.

Beginning in 2005, the new Michigan Content Standards and Benchmarks were released, including specific grade level content standards for grades 3-9 and high school core subjects. Test content for all grade levels was changed to align with the newly released content expectations and the testing administration timeframe moved from fall to winter. New cut scores were set for all subjects and grade levels.

The content of the test changed dramatically from that of the former MEAP (HST). The new test, the Michigan Merit Exam (MME), included some questions from the former HST, but added two new mathematics components – a set of questions from both the ACT and the WorkKeys college and career readiness national examinations. Nearly all constructed and extended response items were eliminated in favor of a multiple choice format in mathematics. Both content difficulty changed (more advanced content was included than in the past) and content process changed (constructed response

questions asking students to explain and reason through words and pictures were removed). The items on the MME tended to be significantly harder than the items on the high school MEAP assessment (MME, 2011).

The changes in content and process in the state assessment had direct impact on the validity of these assessments with respect to my study. When the state adopted the inclusion of the ACT and WorkKeys subtests, local districts no longer had the benefit of examining released items from the assessment. The state no longer released any assessment items for local districts to use in conjunction with actual student scores to impact instructional decisions. In short, district personnel no longer had the ability to examine specific test items to determine if the assessment aligned with key elements of the programs being used in their schools, including Midwest. When released items were provided to local districts from the MEAP (HST), it was much easier to determine if secondary classroom instruction was aligned with assessments and if larger, programmatic decisions were, in fact, on track. The inclusion of the ACT and WorkKeys college readiness indicators shifted the high school test from a tool designed to inform secondary instruction to a tool designed to screen for college readiness. I will return to this issue when I describe recommendations for policy in the final portion of this chapter.

Dependency on Standardized Assessment

As noted in Chapter 1, studies examining the impact of NSF curricula on student achievement have found that students typically fared better than their counterparts taught with traditional curricula when assessed with tests emphasizing higher-level thinking and problem solving skills (Senk, 2003). I wanted to include a measure to test this claim. As already described, however, irregularities during the test administration of the problem

based math assessment at Midwest created issues with reliability and validity of these data which ultimately precluded me from using them. By default, I had to rely on student performance measured by state assessments only. My findings indicate that across nearly all ways in which I examined the standardized test results (MME, ACT, WorkKeys), none were positively correlated with the treatment.

Ironically, the HST used for Cohort 1, which included complex, real-life math problems requiring constructed responses aligned with the kind of tasks central to the NSF-developed high school program, was replaced by the MME by the time students in Cohort 2 were in eleventh grade. It appears that the control group was the only group that included eighth and eleventh grade consistent measures for growth which were able to be equated across time. This may, perhaps, somewhat explain the downward trend in achievement scores beyond Cohort 1.

I include the history of changes in the MEAP assessment program because these changes impacted the validity of findings, which depend on comparisons in the data collected during this study from one year to the next. To summarize, the eighth-grade MEAP test remained constant for Cohorts 1 and 2, but changed significantly for Cohorts 3 and 4. The HST was implemented for the last time for Cohort 1, and the MME (including subcomponents of ACT and WorkKeys) replaced the HST and was administered for Cohorts 2, 3, and 4. For each derivative of these state tests, new scale scores were developed, many of which were not able to be equated. My use of the performance levels (1-4) as set by the state limited my ability to perform finer grain analyses that might have been possible had scale scores been able to be equated.

Furthermore, even the use of performance levels can be criticized, to some extent, as content and methods used in the assessments differed between particular cohorts.

Legislative Changes in State Mathematics Requirements

Further limitations of the study resulted from legislation passed during the timeframe of the study which raised the state graduation requirements for high school mathematics from two to four credits. The additional mathematics requirements were in place for Cohort 4 students at the onset of their ninth grade year. Therefore, students in Cohort 4 needed to take more mathematics and specifically were required to take Algebra, Geometry, and Algebra II plus enroll in a mathematics course during their senior year to earn their high school diploma. This new legislation complicated the outcome variables established at the onset of my study. For this final cohort, both the number of math courses and the highest level of math course in which students enrolled were likely influenced by these new requirements.

My findings suggest that taking more math courses does not equate to stronger indicators of learning. Although Cohort 4 had a slightly positive coefficient (.22), indicating that students in Cohort 4 took more math courses than students in Cohort 1, the coefficient was not significant. Again, this slight upward trend for Cohort 4 would tend to be predicted with the legislative requirement that all students enroll in a mathematics course during their senior year. This finding cannot be attributed solely to the reform treatment since the legislation requiring additional math credits for graduation was in effect. Perhaps this finding provides empirical evidence of support for this new legislation, but perhaps it is at least in part a result of the reform. With the data I presently have, it is impossible to disentangle the causes of this outcome. Even the fact that Cohort

4 is net of the effect of the earlier Cohorts 2 and 3 does not argue for definitive evidence one way or the other.

Confounding Variables

Although the decision to change both student grouping practices and a new curriculum during the secondary mathematics reform at Midwest was based on evidence, nonetheless, it confounded my ability to link my conclusions to one treatment or the other. It is difficult if not impossible to determine whether my results are related to de-tracking, to the NSF developed curriculum, to the combination of both, or that, perhaps, the treatments cancelled one another out. My findings are limited because I examined two important changes that occurred at the same time.

Summary of Limitations

Nearly every outcome variable used in this study was negatively impacted by confounding circumstances that often accompany studies conducted in naturalistic settings like schools. Even though my research study involved several limitations, there are still implications that can be learned from this study which offer future policy makers and school leaders potentially helpful insights.

Recommendations for Policy

- State accountability plans may be strengthened by including ways for districts to provide data about student growth using alternative assessments as supplemental information to state testing results.
- The legislation of a policy requiring that all high school students pass Algebra, Geometry, and Algebra II, as well as requiring all seniors to enroll in a mathematics course, should be evaluated using early evidence

(data) from local districts. The evidence from this study revealed many complexities, especially with regard to inconsistencies in mathematics course titles and multiple iterations of course offerings within Midwest High School. This study underscores the likelihood that data collected for purposes of evaluating the effect of this policy will represent information that will be interpreted at the local level using very inconsistent methods. This study supports the existing literature which emphasizes the lack of curriculum coherency in high schools across the United States in mathematics course offerings, titles, content, and pathways (Schmidt, 2002). While it can be argued that Midwest High School improved curriculum coherency for entry-level mathematics course offerings such as Algebra and Geometry, it still included several variations of upper-level mathematics courses. For example, Trig/PreCalculus, Trigonometry, Discrete Mathematics/Statistics, and Statistics were inconsistently offered to students from one year to the next, even after the reform. Given this evidence, and assuming most schools will shape the courses they provide according to the unique dynamics existing within their own landscape, evaluating the aforementioned policy with real data will be very difficult. I suggest that evaluation tools include carefully constructed rubrics for districts to use to place local course titles into classifications with specific and consistent meanings. This would require district-level mathematics experts to interpret these rubrics while examining their own local course content and materials. The goal would be to improve the likelihood that

data collected would reflect the real meaning and content associated with locally developed mathematics courses, which would ultimately improve the reliability of the data gathered during this evaluation process.

The current legislation requiring students to enroll in a senior level mathematics course should also be evaluated using data that reflect the outcomes that have occurred in local districts. This study demonstrates that the relative percentage of students enrolled in higher-level mathematics courses as seniors remained steady across the eight years of this research. It appears as though Cohort 4 students enrolled in no more higher-level mathematics courses as seniors than their predecessors and that at least 25% of students enrolled in level 1 courses to fulfill this state requirement. The regression analyses I conducted confirmed that more mathematics courses do not necessarily lead to stronger mathematics test scores. This study provides some evidence suggesting that once students reach a particular level, the benefits observed on standardized achievement tests level off.

The opportunity costs associated with requiring senior-level high school students to enroll in a mathematics course should be carefully examined. It may be that other options might provide education experiences that better align with specific student goals for post high school continued development than the required fourth mathematics credit. My study provides early evidence that the number of students enrolling in the highest-level mathematics courses as seniors remained fairly constant over

time, even after legislative course requirements changed. This suggests that for Midwest High School students, the new legislation did not increase the overall number of seniors taking higher level mathematics. Lack of attention to the complexities surrounding secondary mathematics course offerings as described by my research, and lack of attention to how this policy has affected the potential opportunity costs for students by way of limiting accessibility to courses outside of mathematics, could fundamentally flaw evaluation of policies mandating course taking.

- Educational data collection policies should be evaluated in light of the problems raised in this study, which highlight the potential for using inaccurate information within high-stakes contexts. This recommendation is somewhat related to the previous recommendation, but expands it to include all education data which are collected and reported in various tools for public use. The state in which Midwest is located uses a dashboard of indicators for education progress which includes information used to rate the effectiveness of schools. The same complexities described above regarding the differences determined at the local level surrounding mathematics course offerings might also apply to how local districts define other core courses. Further, local iterations of prerequisite requirements, course grading, dual enrollment offerings, and other critical education components cause differences across districts.

As we look ahead to the implementation of National Core State Standards and the Smarter Balanced Assessment program scheduled to begin in

2014, the potential for attaining much-needed interstate coherence in core content curriculum and assessment exists. While we wait for the official implementation of these resources to occur, it would be wise for state education officials to utilize the time during the transition years to better define the existing parameters used for collecting and interpreting education data. Such definitions might increase uniformity of data across local districts. In this way, we may be able to improve the reliability and validity indicators of the impact of these new standards and assessments.

- The current State assessment policy to which Midwest adheres currently does not require the assessment of mathematics at the ninth- and tenth-grade levels. Given the state and national trends for a decline in high school mathematics achievement when compared to elementary and middle school results, a case can be made for secondary assessment policies needing to include grades nine and ten rather than waiting to assess secondary performance at the conclusion of eleventh grade. This might lead to positive implications if these ninth and tenth grade assessments included specific information that could be provided to local districts to inform practice and ultimately improve conditions for students before taking the high-stakes eleventh-grade assessment. I maintain that the window of time is currently too long between the eighth and eleventh grade assessments without data to guide local and state decisions about instruction.

Suggestions for Future Research

The findings from my study could be further enhanced by research conducted to examine some additional aspects of the effects of the mathematics curriculum reform. First, a study focusing on the qualitative components of de-tracking could provide a stronger understanding of how students see themselves in relation to others when grouped heterogeneously. How are self-perceptions among students impacted by reform that includes different grouping practices? By carefully examining interactions among students and between students and teachers that occur in detracked mathematics classrooms, we might better understand important elements that are impossible to discover using only assessment data.

Second, it would be interesting to study the various ways in which schools are implementing state policies requiring specific mathematics coursework as well as mathematics course taking at the senior level. Has this policy resulted in improvements in student achievement? Has it resulted in more or fewer mathematics course options at the local level? Have mathematics course failures increased? How are districts providing support for students in meeting these new requirements? How is the policy mediating positive or negative outcomes for students?

Third, how might research capture the deeper, important instructional changes that were occurring at Midwest High School which, as a result of my research design, I was unable to examine? The NSF programs implemented at Midwest required that most teachers fundamentally change the manner in which they understood mathematics content and process. I provided ample evidence that teachers participated in several ongoing opportunities to learn new ways to engage students in learning important mathematics

concepts. Yet, I can't be sure that once the classroom door was closed, pedagogy actually changed. Qualitative research methods designed to capture how teachers actually think and teach, based on new understanding of content and pedagogy such as that associated with NSF-developed mathematics curricula, would contribute to the story provided by my research. I plan to pursue this line of inquiry in the future.

APPENDICES

Appendix A: Compiled Teacher Feedback After Year 1 Implementation

Math Connections Feedback Comments-October 2005

1. Describe any positive changes you have observed in students' mathematics.

More willing to take chances

More advanced students helping slower students.

From the warm-ups we do, I have observed that students' ability to process fractions and decimals has improved.

Once they learn the material they seem to remember it.

They are more willing to explain the reasoning behind a problem instead of simply doing the process.

I was pleased with how fast they learned about the various plots.

2. Describe, specifically, how student learning in mathematics is different this year, as compared with previous years.

I am not in that class enough right now to answer this.

N/A

It seems much more difficult for the students to understand the main concepts, especially when the homework is not like the examples in the book.

They see how they would use the material and apply it to math.

Much slower and more in depth- They get tired of doing so many thinking problems.

First year teaching

3. Please describe instructional challenges you have experienced thus far.

ATTENTION and WILLINGNESS TO LEARN, also being prepared.

Getting all students to work while in groups and keeping groups on task.

Chapter 6 is too long and could be shortened. The book is too wordy. The assignments are different from the examples and the students can't adapt to the more difficult problems.

Presenting material at a slower rate- We teach them one problem and sometimes they need more practice.

The students don't always remember the different processes.

Students have struggled with deviation.

4. Describe organizational tips you might suggest to others.

Keep them busy. Idle time is a nightmare.

Have everything planned out. Make sure you demonstrate in many different ways.

Restructure and shorten parts of Chapter 6. Make worksheets that are more like the examples in the book. Make additional lessons and present them to the students so they are ready to do the more difficult problems.

Write the week's assignments on board.

Have a binder and check it frequently.

Write assignments on the board.

Check homework frequently.

They seem to like data that applies to them more than book data.

5. Describe any “aha” moment(s) you might have experienced thus far.

Happens with the calculators, when they see rather than just compute.

Letting slower students use the calculator in front of the class.

Sometimes when using the graphing calculators the students are amazed at how easily a long tedious problem becomes simple.

Problems and pages I am not used to referring back to pages for graphs.

Allow the “trouble” students to work the calculator.

6. Describe the support that you need right now to continue to be successful (release time, team-planning, be specific).

Team planning would be nice, after Brian Twitchel trains the new teachers.

I think that occasional “Team planning” to discuss what does/doesn’t work would be awesome. (1/2 day per MP maybe)

Class sizes have to be reduced before we will be successful.

Release time to team plan. To find out ways that are working and ways that are not working.

Smaller class sizes – less than 20.

See/discuss how others are teaching certain topics, how they keep the students motivated and how taking notes and group work run in their classroom.

7. Please list specific questions you would like to have addressed by Brian Twitchel and the Math Connections trainer scheduled to be here in November.

Prep for ACT-How would he handle 30 students given our class map? What would he do with a student who fails a semester?

Group work tips

What parts of the book does he skip and why.

Is there any way to break up Chapter 1 differently? –lots to test over

How can we get through a chapter quicker? How can we quiz less frequently?

8. Please describe how the Brian Twitchel could most effectively help you right now.

Show new teachers how he presents the material, and how he runs the class.

Tips on what does/doesn't work in upcoming sections.

Suggestions on how he would teach the connection series to classes of 27 to 30 students.

How can we get through a chapter quicker? How can we quiz less frequently?

By giving up different activities that work well in groups.

Ways to motivate students to do their work.


9. Other insights

We have some great new teachers, so everything we can do to support them will go a long way.

Appendix B: Slides from Presentation to Board of Education by the Researcher

April 26, 2004


Slide 1




Ultimate Goal of Mathematics Education

- To help our students confidently use their mathematical knowledge to think, reason about, and solve new kinds of problems, to develop new procedures and algorithms, and to understand new concepts.

1



Slide 2



Need For K-12 Curricular Changes In Mathematics


Internal data analysis

Comprehensive analysis of existing structure, teaching methods, curricular materials conducted beginning 2003-04

Curriculum maps completed

- K-5
- 6-8
- 9-12

2




Slide 3

Internal data analysis. . .			
MEAP data reviewed			
MATH	Elementary 4 th Grade	Middle School 8 th Grade	High School 11 th Grade
2003	76.0 65.0 (State)	56.0 52.0 (State)	51.0 60.0 (State)
2002	75.9 64.5 (State)	64.8 53.8 (State)	68.5 67.0 (State)
2001	Final year of non-reformed assessment	8 th grade not tested this year	60.8 68.4 (State)

Slide 4

Internal Data Analysis . . .	
Present sequence of instruction analyzed	
A. Do all students have equal access to mathematics concepts?	
General Math	College Prep Geometry
Technical Math I	Trigonometry/Pre-Calculus
Technical Math II	Advanced Placement Calculus
Algebra	Business Math
College Prep Algebra I	
College Prep Algebra II	
Geometry	


Slide 5




External data analysis . . .

- University experts indicate weakness among college freshmen in mathematical thinking (need for remediation)

5




Slide 6



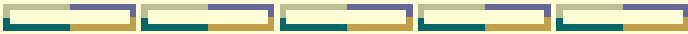
External data analysis . . .

- Workforce leaders indicate that graduates are not prepared to think and reason using mathematics which is relevant to the demands of the real world.

6




Slide 7




External data summary . . .

- Today being basically literate is not enough; all citizens must be fundamentally literate. It is not enough for individuals to be able to do arithmetic problems; they must use arithmetic to solve problems . . . In a society where the ability to work with information and knowledge is the key to employability in well-paying jobs and essential to effective citizenship; it is no longer enough to have a relatively few who are well educated. Today, most must be well educated.

Phillip C. Schlechty 1997 7



Slide 8




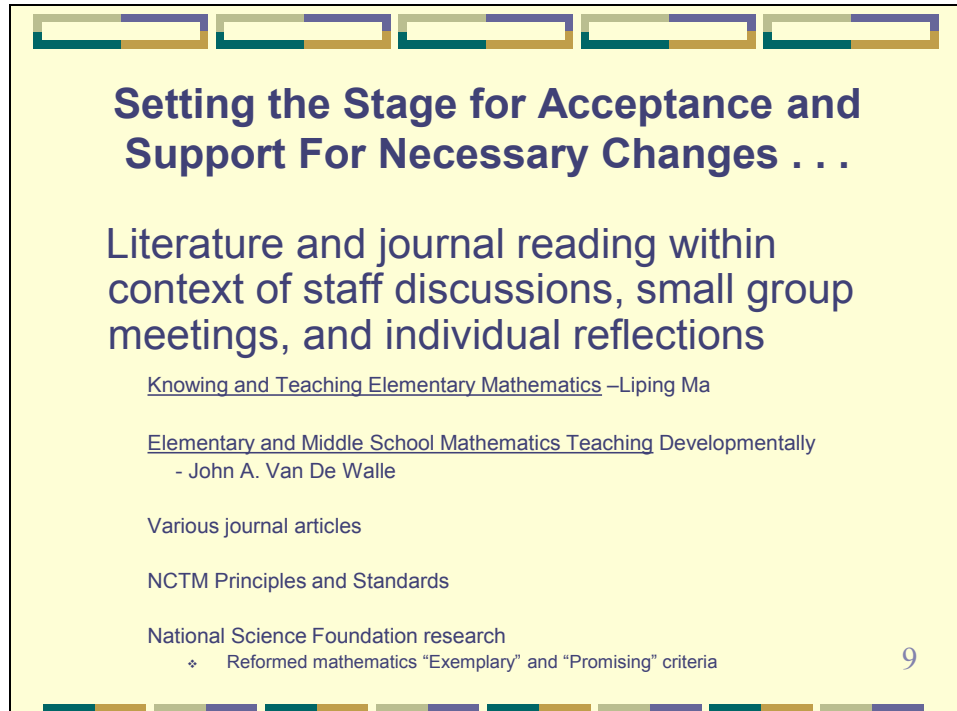
How Will the Changes Occur Most Successfully?

- Through continued professional development that allows teachers to . . .
 - Learn by acquiring new knowledge
 - Learn from professional interaction
 - Learn in and from practice

5-7 years to experience true reform

8





Setting the Stage for Acceptance and Support For Necessary Changes . . .

Literature and journal reading within context of staff discussions, small group meetings, and individual reflections

Knowing and Teaching Elementary Mathematics –Liping Ma

Elementary and Middle School Mathematics Teaching Developmentally
- John A. Van De Walle

Various journal articles

NCTM Principles and Standards

National Science Foundation research
✦ Reformed mathematics "Exemplary" and "Promising" criteria

9

Slide 10



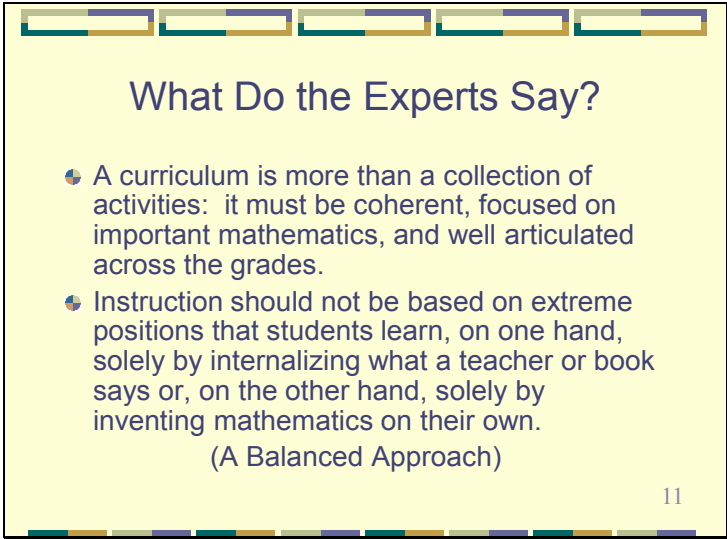
Various mathematics expert consultations and presentations

Including . . .

- ❖ Dr. Charles Van Embes (C.M.U.)
- ❖ Dr. Glenda Lappan (M.S.U.)
- ❖ Dr. Michael Goldenburg (U.of M)
- ❖ Dr. Deborah Ball (U of M)
- ❖ Dr. Matt Bachelis (Wayne State)

10

Slide 11



What Do the Experts Say?

- 🌐 A curriculum is more than a collection of activities: it must be coherent, focused on important mathematics, and well articulated across the grades.
- 🌐 Instruction should not be based on extreme positions that students learn, on one hand, solely by internalizing what a teacher or book says or, on the other hand, solely by inventing mathematics on their own.
(A Balanced Approach)

11

Slide 12




What Does the Research Say?

- **Quality programs intertwine strands of mathematical proficiency**
 - Conceptual understanding:** comprehension of mathematical concepts, operations, and relations
 - Procedural fluency:** skill in carrying out procedures flexibly, accurately, efficiently, and appropriately
 - Strategic competence:** ability to formulate, represent, and solve mathematical problems
 - Adoptive reasoning:** capacity for logical thought, reflection, explanation, and justification
 - Productive disposition:** habitual inclination to see mathematics as sensible, useful, and worthwhile, coupled with a belief in diligence and one's own efficacy

12

Slide 13




What is the emphasis in an NSF “exemplary” mathematics curriculum?

- ❖ Depth of understanding is paramount
- ❖ Multiple ways to solve problems
- ❖ Emphasis on mathematical thinking
- ❖ Technology plays a greater role in accessing higher levels of mathematics for all learners
- ❖ Front loaded time to problem solve before algorithms are introduced and memorized

13

Slide 14



Present Recommendations

Elementary


(2004-2006 Plan)

- ❖ Implement NSF “exemplary” elementary math program beginning 04-05.

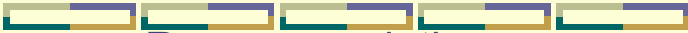
Math Trailblazers, Kendall Hunt Publishing, 2003,
“Exemplary” program status from NSF.

- ❖ Provide professional development, release time for planning, and administrative support during implementation
- ❖ Continue implementation and development during 05-06.

14



Slide 15




Recommendations –

Middle School


(2004-2006 Plan)

- ❖ Pilot specific units of study from NSF “exemplary” programs in grades 6-8 during 04-05 school year.
- ❖ Implementation of NSF “exemplary” mathematics program 05-07.
- ❖ Continue to provide staff development in the area of differentiated instruction to all teachers to feel better equipped to meet the needs of all students (especially those taking lowest track math) within an equitable curricular sequence.

15




Slide 16




Recommendations High School (2004-2006 Plan)

- ❖ Two (2) years of mathematics must include one year of algebra and one year of geometry or the equivalency in a department approved integrated math program.
- ❖ Begin to implement new sequence using reformed materials for Algebra teaching during 2004-2005. Replace General Math course title with new course name that would more adequately reflect curriculum to be taught.
- ❖ Replace Algebra and Tech Math classes with an integrated math program. Adjust the curriculum for College Prep Algebra I and II, College Prep Geometry, Trig/Pre-Calc, and A.P. Calculus to include greater use of technology, statistics, probability, and provide for more real life applications. Continue to merge specific courses in 05-06 to develop greater equity for curriculum (general math).
- ❖ Continue to provide staff development in the area of differentiated instruction to allow teachers to feel better equipped to meet the needs of students within an equitable curricular sequence.

16




Slide 17



Questions from the Board?

Thank you for your continued support!

17



Appendix C: Memo to John M. Smith, Superintendent of Midwest School District

To: Mr. John M. Smith

From: Kari S. Selleck

Re: High school mathematics textbook recommendations

Date: May 19, 2004

As I stated to you during my Board presentation on Monday, May 17, we have worked extremely hard during the last year to examine our current K-12 mathematics program. We have reviewed course offerings, student achievement data, teacher perceptions, current research, and many other sources of information to allow us to move forward with significant program changes across all levels of our district. In short summary, the following will be implemented during the 2004-2005 school year:

1. Full implementation of new elementary mathematics program entitled, *Math Trail Blazers* (received National Science Foundation “exemplary” status). Comprehensive professional development and planning time to accompany and support implementation will be provided throughout the year.
2. Middle school professional development and review of reformed math programs will be provided to allow for successful implementation of new curriculum during the 2005-2006 school year.
3. High school professional development will occur across the year to support the implementation of *Math Connections* (received NSF “exemplary” statue) in the newly

reformed Algebra class. Initial training will be held during consecutive training days prior to the beginning of the school year (August 17-18-19).

****Middle school and high school math teachers, including special education teachers will be combined for all training.**

4. High school professional development will occur throughout the year to support the implementation of the new *Glencoe* materials for College Preparatory Algebra and Geometry.
5. Professional development associated with *Math Connections* and the *Glencoe* programs emphasizes the use of technology to increase student achievement in the area of mathematics and science.
6. Midwest Middle School math and science teachers have been selected to be part of the DART grant. This grant will provide teachers with additional opportunities to learn how technology can enhance math and science learning. Midwest's middle school teachers will have increased access to technology advancements provided by the research institute during this grant project.

Recommended purchases:

Algebra

Math Connections: A Secondary Core Curriculum

IT'S ABOUT TIME, Inc.

Herff Jones Education Division

2004

College Prep Algebra

Algebra I

Glencoe Mathematics

2004

College Prep Geometry

Geometry

Glencoe Mathematics

2004

All materials selected were evaluated by the entire Mathematics Department using a thorough review process. Evaluative criteria were used to examine all components of the programs to insure alignment with national and state math standards.

Estimated costs for all above materials (including student textbooks, teacher resources, accompanying software): \$28,000

All funds have been budgeted within the high school textbook account for expenditure for the 2003-2004 fiscal year.

Outcomes of above work:

Math achievement will improve as evidenced by the following:

1. Improved attitudes and self perceptions about mathematics by all students
2. Improved MEAP scores at all levels
3. Increased retention and achievement in existing mathematics courses
4. Increased enrollment, retention, and achievement in advanced mathematics courses
5. Improved SAT scores

Appendix D: 2004 News Article Included in Midwest Weekly Column in Local Newspaper

Raising Higher Levels Of Excellence In Mathematics

Although mathematics achievement has always been strong at Midwest School District, in the spring of 2004, K-12 teachers came together as a community of learners to analyze current research in mathematics, to review longitudinal data, assess existing instructional practices, and examine student attitudes and perceptions about math. As a result of this yearlong examination, K-12 teachers and leaders concluded that to build even stronger foundations in mathematical understanding for all students in Midwest School District, an intensive effort of curriculum realignment and professional development would need to be planned.

At the conclusion of the 2003-2004 school year, elementary teachers reached consensus to implement the K-5 program, *Math Trailblazers*. High school teachers unanimously selected *Math Connections* as their program of choice to begin in the ninth grade for 2004-2005 and to continue for 10th graders in 2005-2006. Both national programs were developed as a result of recommendations made by many international research institutes, following years of study of students' mathematic achievement across the world.

Curriculum Director, Kari Selleck, explains:

After many months of research and discussion among all district stakeholders, we established criteria that aligned with national and state standards for exemplary mathematics programs for the 21st century. We knew that we would only review

programs that had received the coveted 'Exemplary' status through the National Science Foundation. This way, we were assured that we would adopt programs which would lead us to higher levels of mathematics achievement, rather than adopting programs that look much like the ones we already have.

After program decisions were made, the difficult work began. Several release days were provided during last year's school year and during the summer months for teachers to learn from program trainers to insure implementation success. In addition, time was built into this year's schedule for teams of teachers to meet periodically to continue professional dialogue about instructional practices and student work. Selleck comments on the success of the program implementation already being observed in the district.

I am amazed at the improved mathematical reasoning already happening across all grade levels. I applaud the teachers, support staff, and administrators for their 100% commitment to improving their existing instruction practices in mathematics. Teachers are spending hours and hours above and beyond the school day to effectively plan for mathematics instruction. My hat goes off to this amazing staff! Students are not only mastering basic facts and skills at an earlier age, but are demonstrating their expanded abilities to reason mathematically, solve problems using a variety of strategies, think independently, work alone or in groups, and effectively communicate their mathematical thinking like never before.

Additional professional development will occur across the 2004-2005 school year. After the winter break, parents will be invited to participate in an activity night, designed to

allow parents to actually participate in some of the higher-level math processes that students are learning on a daily basis. Selleck explains that parent education is an element that cannot be forgotten.

One of the biggest challenges is to provide parents with the necessary background knowledge so they can be able to help their children at home and support classroom practice. I hope that parents continue to ask questions and take the opportunities provided by our district over the next few years to learn about our exciting mathematics program. Parents should patiently encourage children to try to explain the how's and why's of doing mathematics, rather than focusing on only one way to arrive at mathematical solutions.

Selleck asks everyone to remember, “We are ALL learning how to prepare our students to be successful in a world that will place demands on our graduates that are much different than when we were educated! ... Our goal in Midwest is to reach higher levels of excellence in mathematics as we prepare our students for life in the 21st century!”

Item of note: Middle school teachers will select a 6-8 grade program from NSF “Exemplary” choices during the 2004-2005 school year, for full implementation scheduled to begin in 2005. The middle school decision will align with the elementary and high school’s new curriculum.

Below you will see evidence of the many components of the new mathematics program across Midwest School District.

[photos]

#1 Teachers hear from Kari Selleck, Curriculum Director, regarding the collection of baseline achievement data in elementary mathematics.

#2 *Math Trailblazers* national trainer, Lynnette Moyer, works with teachers in implementation of the new curriculum.

#3 Ninth-grade mathematics instructor, Ned Dier, helps facilitate a discussion with high school students, who are working together on real-life situations where the distributive property is needed.

#4 Ned Dier demonstrates the use of the storage case for graphing calculators. Graphing calculators are a vital component in all ninth-grade classrooms.

Appendix E: Letter from High School Principal to Staff, 2005

REFLECTION ON CHANGE – AS IT RELATES TO THE MATH DEPARTMENT AT MHS

1. Three years ago our curriculum director realized through the analysis of data and a review of the most current research that math reform was needed at Midwest School District. Although a portion of students were finding success in the traditional teaching of math, the majority were finding themselves less successful in their ability to do well in the classroom, to apply math they learned to the real world, and to maintain basic mathematical concepts from one level to the next. Furthermore, institutions for higher education continued to claim that students entering as freshmen were ill prepared for the rigor in mathematics classes. Employers across our county echoed the sentiments that students lacked problem solving and estimation skills within the world of work.
2. The school year 2003/04 found work being focused at the secondary level, primarily at the high school where test scores were the lowest. (Include a look at mathematical scores from 2003/04 at the district level; also how we compared to the state average.) Much time, research into reform, and discussion was held.
3. Following recommendations set forth by the National Council of Teachers of Mathematics, the department looked at programs that had been awarded an “Exemplary” rating by this nationally recognized council. Representatives of this department looked at these programs, including *Math Connections*, *Core-Plus*, *Mathematics: Modeling Our World*, and *SIMMS* (which were deeply rooted in

research by a team of mathematic teachers in the nation) and decided that they felt *Math Connections* aligned best with the national standards and our goal to improve mathematical understanding for all students. It was decided that *Math Connections* would be used at the high school level in the school year 2004/05. Several formal and informal discussions continued to take place centering on the difficult, ethical question: Who should be served by the Math Connections program? How should we structure our courses? After much heated debate, particularly about tracking, an *interim* compromise was reached so that the most advanced students would continue in the traditional program with *Math Connections* materials to be used for all other, heterogeneously grouped math classes. To the dismay of district leaders, it was later determined that only those students representing the lower ability population in mathematics were placed in the classes using *Math Connections*. These are currently our Connections II kids (75 students). Randy Sanga and Ned Dier were the only teachers who taught *Math Connections* in the 2004/05 school year. At the conclusion of this first year, in spite of the lopsided grouping of these courses, both instructors were pleased with the quality of instruction and the achievement levels experienced among the students using *Math Connections* materials. In fact, Ned Dier explicitly stated his belief that “this program would be good for ALL kids.”

4. Much discussion was held within the math department throughout the 2004-05 school year dealing with the second year of *Math Connections*. Much of this discussion surrounded change for the upper middle and higher students. The basic

premise assumed by some veteran teachers was that top kids were doing all right in the traditional programs based on individual MEAP data and that there was simply no proof that the new program would be beneficial to this select group of students. Several times during this discussion, both Ned and Randy stated this type of program would benefit every student and that they used some of the same techniques they were using in *Connections* within their traditional classrooms. Research continued to be provided to demonstrate the need for use of a standards-based math program for ALL students. Even so, the majority of veteran teachers stated that they really did not oppose the series; however, they were not comfortable teaching it. It was stated by Kelly Smith to them that they seemed to loudly say, “The program is fine with me as long as you do not make me teach it.”

5. Recognizing that the grouping process used for the 2004-2005 school year provided unfair hurdles for instructors using *Math Connections* (and did not align with the original intent of the decisions made earlier), Ned, Kelly Smith, and Terri Varce took the data supplied by the middle school teachers and counselors to determine which students should go to the College Prep Algebra and who should go to the *Math Connections*. This need to sift and sort our students is a result of the now two-tiered structure we have perpetuated as we continue to move to using research-based materials for ALL students. The result for this second year is that more of the upper middle were included within these Connections I classes for 2005-06 and additional numbers of students chose to take a second year of math,

using standards-based materials. However, there still remains a perspective that the low or “dummy” students are in the *Connection* series.

As we look to 2006-07, this is some background that led to our recent discussions in the math training on 11-10 and 11-11.

6. As your leader, I hope you see that I am working tirelessly for what is right for all students. I have a passion for public education which means that as educators we believe that all students can learn given the right environment. That is the premise I use as I talk with you about math today and math as it will look in the future. So I ask you to think about the questions that I have spent the last two years reflecting on and the four years prior to that as I worked with an elementary staff who saw that all of their hard work in the area of mathematics was not getting the results they hoped to get for all kids.

- A. Why can't all students be successful using a standards-based mathematic program that aligns with broad goals for reaching higher levels of mathematics understanding?
- B. How do we truly define mathematical success for our students? MEAP scores? ACT scores? College entrance percentages? Success in the work force?
- C. Do we want our program to be viewed as successful by the whole or by only those students at the top?

- D. Can the lower/middle students be more successful in *any* program if the class is homogeneous?
- E. What do we want our students to be able to do with their math skills once they leave us?
 - 1. Be mathematicians?
 - 2. Be able to apply math to their own world?
 - 3. Have a deeper understanding of math concepts?
- F. How do we feel about the fact that K-8 colleagues in our district have embraced a total standards-based program and we have not?
- G. What will our students expect from us as math teachers over the course of the next 2-8 years?
- H. What is your own philosophy about public education and our role as educators to ensure that all students can learn given the right environment?
- I. What is your current level of understanding of recommendations being made by experts across our nation in the area of secondary mathematics instruction?

Just two more ideas to reflect on:

- 1. As an educational leader, I believe that it is my role to be a visionary and to look at the whole, broad picture. As a result of that ideal and my current knowledge of education, classroom methods, student learning, teaching, and global perspectives, I believe that our methods as secondary teachers need to change.

2. One of the reasons that education is always being mandated to do things is that we seem to historically have a problem with being proactive. Our history shows that we fight to maintain the status quo for a variety of reasons. For this reason, there were few secondary teachers or administrators who were even asked to sit on Michigan's high school reform committee, which will be handing down mandates yet this year to us, to which we will then be forced to react!

Appendix F: High School Mathematics Materials Review Rubric, 2003

Program Title: _____

Reviewed by: _____

Primary considerations

_____ Aligns with district goals for students

1. Learn to value mathematics
2. Become confident in their ability to do mathematics
3. Become mathematical problem solvers
4. Learn to communicate mathematics
5. Learn to reason mathematically

_____ Aligns with K-12 mathematics philosophy

_____ Aligns with Professional Standards for Teaching Mathematics

_____ Aligns with NCTM Principals and Standards for School Mathematics

_____ Emphasis on sense-making and mathematical understanding

_____ Promotes learners to raise their own questions, generate their own hypotheses and models as possibilities, and test them for viability

_____ Promotes mental engagement in their learning

_____ Promotes a problem-based atmosphere

_____ Promotes depth and understanding of concepts and skills

_____ Assessment and instruction are intertwined

_____ Promotes cooperative learning

Secondary considerations

_____ Teacher materials organized in manner that is accessible

_____ Supportive materials (manipulatives) available which align with program

_____ Professional development and training is available to teachers through program representatives

Reflective comments:

Appendix G: Correspondence Concerning and Minutes of Meeting at MSU

Correspondence

December 2007

Correspondence Between Curriculum Director and Math Teachers About Need for Visitation by Nearby University Professor (Author of one of the NSF Programs being used at Midwest)

Hi everyone:

After talking with John and Kelly this morning, we thought it might be a good idea to meet for $\frac{1}{2}$ day sometime next week or the following week to talk “math” . . . where we are, where we are going, etc. I would like to invite a guest to join us from Michigan State for a portion of our time together. Dr. Glenda Lappan is a one of the original authors of *Connected Math* and has served as a key force in the standards-based math movement at the national and international level. I would like to ask her to meet with us to answer some of the questions we might have right now. For example:

1. Should we “stay on course” implementing purely from *Connected Math* at the middle school rather than filling in gaps to meet State grade-level content expectations?
2. What should we consider, given our reform efforts, as we look to develop end-of-course testing options to meet State expectations?
3. How do we continue to assure parents and students that their individual needs are/will be met through our program?

These are just some of the questions I would like her to address. Of course, you would bring your questions to the table.

I feel better having this dialogue as a leadership team before we bring both teaching groups together for a work session. I will contact Dr. Lappan and then let you know if this is an option.

Let me know your thoughts, please.

Thanks,

Kari

Dr. Lappan:

I am writing for two purposes:

1. I want to share with you progress we are making toward reforming our K-12 mathematics program at Midwest. I believe that we have done what many have not been able to do --which is to adopt and truly implement (in pure form) three NSF programs (*Math Trailblazers*, *Connected Math*, and *Math Connections*). We have moved to an all inclusive, heterogeneous philosophy through the ninth grade. Over the last four years, we have worked very hard to eliminate the four-track system that had been instituted for many years prior. By far, this has been the most challenging, yet the most pride-evoking project of my professional leadership career.

But, we are at a point in our progress where we need to make some decisions that I don't feel capable in guiding.

2. I was hoping to make an appointment with you to ask you some questions. I would like to bring my middle and high school department chairs and their principals with me. I expect that we would need no more than 30 minutes of your time.

Minutes of Meeting

Date: December 12, 2007

Present: Dr. Glenda Lappan, Kari Selleck, Midwest Middle School Principal, Midwest High School Principal, Midwest High School Math Department Chair, Midwest Middle School Math Department Chair

During this meeting, several important questions were asked of Dr. Lappan.

Background:

We are currently implementing three standards-based math reform programs in the purest sense: elementary – *Math Trailblazers*; middle school – *Connected Math*; high school – *Math Connections*. All programs were developed by the National Science Foundation and were ranked “exemplary programs.” Although we are trying to implement these programs as intended by program developers, State grade-level content expectations that have been released in the last few years do not necessarily align with the programs. We run the risk

of losing the momentum to implement the programs, as intended, if we begin to develop curriculum plans to address specific grade-level content expectations. We have made a strong commitment to refrain from using individual teacher autonomy or grade-level autonomy to stray from the intended NSF programs' written curriculum. Yet, we are facing increasing pressure to comply with State expectations.

Question: Should we “stay the course” to continue implementing these standards-based programs in the truest sense or should we alter the intended curriculum and begin to implement lessons that specifically address individual GLCE’s that may or may not align with the sequence of the NSF programs?

Dr. Lappan responded to our first question by complementing our district’s efforts. Dr. Lappan stated that it is extremely rare to have a district in the position that we are in. She was extremely impressed that we had made the decisions that we made regarding use of three reform (standards-based) programs. She extended great support for the fact that we were so far into the process of reform and that we had managed to be in years three, four, and five of implementation at the various levels.

She emphasized that the GLCE’s are “moving targets,” and having worked directly with the organizations and individuals who have led the efforts to develop these content expectations, she is well aware of the political influence associated with these. Dr.

Lappan shared her belief that the GLCE’s will most likely shift, change, and be rewritten in the near future. Because of this, Dr. Lappan urged us to “stay the course” with our authentic implementation of NSF math programs and continue to operate with the belief that the thinking skills and problem-solving strategies emphasized across all NSF programs will allow our students to do well with the limited “unfamiliar content” that

students might encounter on State assessments. Dr. Lappan cited studies that showed that students taught using NSF programs as compared with traditional programs have fared “equally well” when tested on traditionally represented content. However, she explained that the studies show that when tested on mathematical content that represents higher-level mathematical thinking, students taught through NSF programs fare far better than those taught using traditional programs. Therefore, Dr. Lappan urged us to “stay the course” with what we are presently doing.

Question: Given the present schedule and 43 minute class time within which our middle school math courses operate, what suggestions can you give us as we struggle to complete all the units of instruction within the Connected Math Program?

Dr. Lappan, as one of the original program authors for Connected Math Series, indicated that a 50- or 60-minute math period would most certainly be more optimal for completing more work. She talked about how even moving to 50 minutes would add another class period of time to each week. So, her first preference would be to advise us to extend our math course time periods to 50-60 minutes per day for instruction. Next, Dr. Lappan discussed the design of the program and provided several examples of how particular units included several parts to each problem. By design, each problem was written to include the “heart of the problem” and then provide successive and increasingly sophisticated extensions to the problem. She emphasized that we should strive to get through as many units as possible across the program, but if time continues to be a challenge, she would advise us to back off on some of the extra investigations provided in each unit. She used several examples to illustrate her point. These examples included “Shapes and Designs,” “Factors and Multiples,” “Bits I, II, III,” etc. She talked about

reducing the volume within each unit, rather than ending the year failing to implement all units.

Background provided: We have moved to heterogeneous grouping of all math classes for required courses. We have not faced criticism from parents at the middle school level, even though several levels of courses traditionally had been offered. (Students entering as sixth-graders were placed in low/co-taught math, regular math, or advanced math). We now have students moving through 6th, 7th, and eighth-grade math in heterogeneous groupings. We have worked to phase out the prior tracking system for high school as well. Over the last four years, we have eliminated tracking first at the 9th grade, then the 10th, and now the eleventh for required math courses. This year, as the high school implements a 4 x 4 block schedule, students have multiple opportunities to take additional math electives at any time during their schedule. In summary, the internal and external pressures that may have existed within our traditional program to provide early opportunities for advanced track math options have been eliminated due to the new schedule.

Question: *Since we realize that one of the principles of NSF programs is that classroom activities depend on heterogeneously grouped student arrangements, how should we address State requirements that warrant districts offering an Algebra I exit exam so students can “opt out” of this course. Won’t we be setting up our teachers and students to fail if we remove the top 1/3 of students from the heterogeneous mix? Additional questions we have include: Should we develop an Algebra I exit exam that aligns with our program content or the State-level GLCE’s? Further, should we have all eighth-*

graders take this exit exam as a matter of protocol at the conclusion of the eighth-grade year?

Dr. Lappan talked about the range of learners within heterogeneous groups and encouraged us to stick with our present plan, yet she made a significant clarification regarding the atypical student at the upper-most levels of natural ability, interest, and aptitude. She used the example of one of her own daughters who, as a middle and high school student, would rather spend all day in the summer at a math camp at a local university than doing other, typical adolescent things. Dr. Lappan said that we need to remain sensitive to and aware of those few students who could truly benefit from additional enriching opportunities. She cited examples of doubling up in courses, dual enrollment, summer camps, on-line experiences, etc. She emphasized that an Algebra I exit exam should be “ready for the asking” to provide additional options for those few students who seek out extra opportunities.

Dr. Lappan also encouraged us to develop the Algebra I exit exam based on the NSF program concepts, rather than on the State GLCE’s.

Background: A recommendation from the high school math department to institute a prerequisite “minimum grade allowance” to be able to earn credit and move forward in our math sequence was recently adopted. While the rationale for this criterion made sense at the time of adoption, now that it is being implemented, several unanticipated outcomes have resulted.

Question: *Are we serving our students in the best fashion by imposing a “minimum grade expectation of a C- or better per semester” to move forward to the next semester of math? Is it possible for a student to receive less than a C- during the first semester and be*

successful the second semester so that one would that would question the necessity of retaking the first half? Why should mathematics be uniquely defined to require this minimum grade to move forward over other subject areas?

Dr. Lappan indicated that prerequisites of this nature can, in fact, impede the path for students in unjust ways. She encouraged us to establish a means to involve teachers in reviewing individual progress and to carefully consider decisions that accompany prerequisite guidelines on a case-by-case basis. She felt that there are no guidelines that fit all students and encouraged us to adopt procedures that included multiple measures (including administrator/teacher input) to understand the needs of each individual student before making decisions.

Dr. Lappan agreed that mathematics is far more “ill-structured” than traditionally considered. She appreciated the question being asked about the relationship between Algebra I and Geometry. She shared that the algebraic and geometric strands of mathematics actually tap into distinctly different thinking skills for students. Algebra taps into representational, abstract thinking skills while geometry taps into spatial skills. She iterated that it is not uncommon to see students struggle with Algebra, but then do very well in Geometry. In fact, Dr. Lappan shared her belief that students might fare better, developmentally, if introduced to geometry prior to algebra. However, this would require a complete restructuring of institutional arrangements that have long and deep roots. For another day. . .

Conclusions from the meeting:

1. Stay the course!
2. Trust in the content within the NSF programs and continue to implement in authentic and pure means.

3. Examine the possibility of adding time to middle school mathematics classes for the 2008-2009 school year.
4. Consider looking at the “heart of the problem” for each unit in *Connected Math* and begin to scale back 1 or 2 of the investigations for each unit. Potentially use Dr. Lappan’s guidance to help accomplish this curricular task.
5. Develop an exit exam for Algebra I based on the content which aligns with the NSF program, *Math Connections*.
6. Reconsider the minimum requirements recently adopted for passing and receiving credits in high school mathematics. Utilize case-by-case decisions based on multiple measures and teacher/administration feedback.
7. Consider the potential curricular implications of distinctly different skill sets (background knowledge) required for Algebra and Geometry.
8. Contact *Math Trailblazer* authors and *Math Connections* authors to determine if the same logic of “extra investigations” embedded within each unit in *Connected Math* might also apply to these programs. Struggles with completing all program components continue to exist in the elementary and high school math courses. This might help us solve these challenges.
9. Invite Dr. Lappan to speak to the middle and high school groups as a *supporter* and *champion* of their work.

REFERENCES

REFERENCES

- Alexander, K., & Cook, M. (1982). Curricula and coursework: a surprise ending to a familiar story. *American Sociological Review*, 47(5), 626-40.
- Alvarez, D., & Meehan, H. (2006). Whole school detracking: A strategy for equity and excellence. *Theory Into Practice*, 45(1), 45-46.
- Batista, M. T. (1999). The mathematical miseducation of America's youth: Ignoring research and scientific study in education. *Phi Delta Kappan*, 81, 425-433.
- Berlinghoff, W. R., Slover, C., et al. (2000). *Math Connections. A Secondary Mathematics Core Curriculum*. Armonk, NY: It's About Time.
- Bieler, J. (1997). *Math Trailblazers*. Dubuque, IA: Kendall Hunt.
- Burris, C. C., Heubert, J., & Levin, H. (2006). Accelerating mathematics achievement using heterogeneous grouping. *American Educational Research Journal*, 43(1), 103-134.
- Burris, C. C., Welner, K. G., & Bezoza, J. W. (2009). *Universal access to a quality education: Research and recommendations for the elimination of curricular stratification*. Education and the Public Interest Center & Education Policy Research Unit, National Education Policy Center, University of Colorado at Boulder. Retrieved from <http://epicpolicy.org/publication/universal-access>
- Cogan, L., Schmidt, W., & Wiley, D. (2001). Who takes what math and in which track? Using TIMSS to characterize U.S. students' eighth-grade mathematics learning opportunities. *Educational Evaluation and Policy Analysis*, 23(4), 323-324.
- Finley, M. K. (1984). Teachers and tracking in a comprehensive high school. *Sociology of Education*, 57(4), 233-243.
- Gamoran, A., & Berends, M. (1987). The effects of stratification in secondary schools: Synthesis of survey and ethnographic research. *Review of Educational Research*, 57(4), 415-434.
- Hallinan, M. T. (2004). Whatever happened to the detracking movement? *Education Next*, 4(4). Retrieved from <http://www.hoover.org/publications/ednext/3260116.html>
- Hamilton, L. S., McCaffrey, D. F., Stecher, B. M., Klein, S. P., Robyn, A., & Bugliari, D. (2003). Studying large-scale reforms of instructional practice: An example from mathematics and science. *Educational Evaluation and Policy Analysis*, 25(1), 1-29.
- Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, UK: Routledge.

- Horn, I. S. (2005). Lessons learned from detracked mathematics departments. *Theory Into Practice*, 45(1), 72-81.
- Kennedy, M. (1998). *Form and substance in inservice teacher education*. Research monograph. National Institute for Science Education, University of Wisconsin-Madison. Retrieved from <http://www.eric.ed.gov/PDFS/ED472719.pdf>
- Le, V. N., Stecher, B. M., et al. (2006). *Improving mathematics and science education: A longitudinal investigation of the relationship between reform-oriented instruction and student achievement*. RAND Corporation. Santa Monica, CA: National Science Foundation.
- Lortie, D. C. (1975). *Schoolteacher*. Chicago, IL: The University of Chicago Press.
- Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Lawrence Erlbaum.
- Mayer, D. P. (1998). Do new teaching standards undermine performance on old tests? *Educational Evaluation and Policy Analysis*, 20(2), 53-73.
- McKnight, C. C., Crosswhite, F. J., et al. (1987). *The underachieving curriculum: Assessing U.S. school mathematics from an international perspective*. Champaign, IL: Stipes.
- Michigan Department of Education (2011). *MME (Michigan Merit Examination) guide to reports*. Retrieved from http://www.michigan.gov/documents/mde/Guide_to_Reports_-_Spring_2012_MME_390749_7.pdf
- National Assessment of Educational Progress (1972). *Questions and answers about the National Assessment of Educational Progress*. Ann Arbor, MI: National Assessment of Educational Progress.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: Government Printing Office. Available from <http://www.ed.gov/pubs/NatAtRisk/risk.html>
- National Commission on Mathematics and Science Teaching for the 21st Century. (2000). *Before it's too late: A report to the nation from the National Commission on Mathematics and Science Teaching for the 21st Century*. Washington, DC: U.S. Department of Education.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- National Education Association. (1990). *Academic tracking: Report of the NEA executive committee subcommittee on academic tracking*, pp. 1-42. Washington, DC: Author.
- National Research Council. (1989). *Everybody counts: A report to the nation on the future of mathematics education*. Washington, DC: National Academies Press.
- No Child Left Behind Act of 2001, Pub. L. No. 017-110. Available from <http://www.ed.gov/policy/elsec/leg/esea02/index.html>
- Oakes, J. (1992). Can tracking research inform practice? Technical, normative, and political considerations. *Education Researcher*, 21(4), 12-21.
- Oakes, J., & Guiton, G. (1995). Matchmaking: The dynamics of high school tracking decisions. *American Educational Research Journal*, 32(1), 3-33.
- Riley, R. W. (1988). The state of mathematics education: Building a strong foundation for the 21st century. *Notices of the American Mathematical Society*, 45(4), 487-491.
- Schmidt, W. (2008, May). *Dividing opportunities: Tracking in high school mathematics* (PROMISE Research Report, Vol. 3). East Lansing, MI: Michigan State University.
- Schmidt, W., Houang, R., & Cogan, L. (2002). A coherent curriculum: The case of mathematics. *American Educator*, 26(2), 1-17.
- Schoenfeld, A. H. (2004). The math wars. *Educational Policy*, 18(1), 253-286.
- Senk, S. L., & Thompson, D. R. (Eds.). (2003). *Standards-based school mathematics curricula: What are they? What do students learn?* (Studies in Mathematical Thinking and Learning Series). Mahwah, NJ: Lawrence Erlbaum.
- Spillane, J. P., & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. *Educational Evaluation & Policy Analysis*, 21(1), 1-27.
- Stein, M. K., Smith, M. S., Henningsen, M. A., & Silver, E. A. (2000). *Implementing standards-based mathematics instruction: A casebook for professional development*. New York, NY: Teachers College Press.
- Supovitz, J. (2006). *The case for district-based reform: Leading, building, and sustaining school improvement*. Cambridge, MA: Harvard University Press.

- Useem, E. (1992). Getting on the fast track in mathematics: School organizational influences on math track assignment. *American Journal of Education*, 100(3), 325-353.
- Van De Walle, J. A. (2004). *Elementary and middle school mathematics: Teaching developmentally*. Boston, MA: Pearson Education.
- Welner, K. G. & Burris, C. C. (2006). Alternative approaches to the politics of detracking. *Theory Into Practice*, 45(1), 90-99.
- Westbury, I. & Travers, K. (1990). *Second International Mathematics Study*. Urbana, IL: Urbana College of Education. ERIC # ED325360
- Wilson, S. M. (2003). *California dreaming: Reforming mathematics education*. New Haven, CT: Yale University Press.