The Influence of a Reform-Based Mathematic Program on Third, Fourth, and Fifth Grade Student Achievement

William M. Ward
Seton Hall University

Follow this and additional works at: https://scholarship.shu.edu/dissertations

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

Recommended Citation
Ward, William M., "The Influence of a Reform-Based Mathematic Program on Third, Fourth, and Fifth Grade Student Achievement" (2009). Seton Hall University Dissertations and Theses (ETDs). 1614.
https://scholarship.shu.edu/dissertations/1614
THE INFLUENCE OF A REFORM-BASED MATHEMATICS PROGRAM ON THIRD, FOURTH AND FIFTH GRADE STUDENT ACHIEVEMENT

BY

WILLIAM M. WARD

Dissertation Committee

Anthony Colella, Ph. D. Mentor
Barbara Strobert, Ed. D.
Christopher Tienken, Ed. D.
Paul Arilotta, Ed. D.

Submitted in partial fulfillment
of the requirements of the degree
Doctor of Education
Seton Hall University

2009
Abstract
The Influence of a Reform-based Mathematics Program on Third, Fourth and Fifth Grade Student Achievement

This nonexperimental, quantitative, cross-sectional, explanatory study (Johnson 2001, p. 10) investigated the differences between a reform-based elementary school mathematics program, traditional elementary school mathematics programs, and General Education student achievement in high-socioeconomic districts on the New Jersey Assessment of Skills and Knowledge (NJASK) at grades 3, 4, and 5. An analysis using correlational statistics of student scores from the 2007 and 2008 administration of NJASK3, NJASK4, and NJASK5, with a sample population of more than 200 New Jersey DFG-I and J public schools with over 60,000 third, fourth, and fifth grade students indicate mixed results for the existence of differences in mean scale scores between schools implementing a traditional mathematics program or a reform-based elementary mathematics program and student achievement on NJASK.

The study found statistically significant differences in mean school scale scores between fourth grade General Education students using the *Everyday Mathematics* program and those using a traditional program on the 2007 and 2008 tests. Statistically significant differences were also found between mean school scale scores for fifth grade students using the reform-based program and those using traditional programs on the 2008 NJASK5 test. All statistically significant differences favored the *Everyday Mathematics* schools. Differences on the 2007 and 2008 NJASK3, as well as the 2007 NJASK5 were not found to be significant.

Effect sizes, calculated using Cohen’s $d$, ranged from -0.55 to +2.13 for the *Everyday Mathematics* program.
Implications: The study results provide information to assist school leaders in selecting elementary mathematics programs. Recommendations for future research and practice are included.
Acknowledgements

I would like to thank a number of people for their assistance in helping me complete this dissertation.

I am thankful for the support of my dissertation committee. Dr. Anthony Colella was always supportive and encouraging of me and my ideas. Dr. Barbara Strobert provided the encouragement and push I needed to “get going and get it done.” Dr. Christopher Tienken helped me make the final document more comprehensive, consistent, and cohesive. Dr. Paul Arilotta provided a model for unfailing dedication to the students to whom we are entrusted. I would not have completed this journey without their support, encouragement, and assistance.

I extend thanks and acknowledgement to my colleagues in Cohort IX. Those who went before were a model for all of us. I am thankful for opportunity to know each of you and enriched by my association with such a fine group of educators. I am proud to be a member of Cohort IX.

I am appreciative of the assistance of the professionals at the New Jersey Department of Education: Joanne Tubman for her prompt reply to any question; Dr. Jo-Lin Liang, Dr. Donald White, and Dr. Timothy Peters for their guidance with state assessment data.

I am thankful to the graduate faculty at Seton Hall University for their understanding, support, and substance, particularly Dr. John Collins for his prompt responses to all inquiries regardless of how mundane or inconsequential.
I am most appreciative of my colleague Doug Barrett who tirelessly tutored me in the basics of *Excel* spreadsheets and provided daily encouragement. Thank you to Barbara Dean for the frequent suggestions of books and research articles.

The ready assistance of complete strangers in my pursuit of reports, journal articles, and research studies was of immeasurable help in completing the dissertation. Thank you to Kathy Wickert of WESTAT Education Studies, Jacqueline Allen of Mathematica Policy Research, Inc., John Tomicek of COMAP, Inc., and Alisha Strobel of PRES, Associates.

Similarly, special thanks to two noted researchers who took time from their own work to help me with my research: Dr. Janet Johnson, President of EDSTAR, Inc. and Dr. James Hiebert, University of Delaware.

Thank you to members of my family—Dad, Carol, Katie, Paul, Mimi, Shirley, Jens, and Merry, Margaret, Bob—for their support, encouragement, and sometimes not too gentle nagging. Each was helpful to me in his/her own way.

Finally, thank you to my wife, Annie. Your tolerance of late nights, weekends, and vacation time spent at the computer was instrumental in the completion of this dissertation. Thank you for your proofreading and editing suggestions; the mighty pen strikes again. Most importantly, thank you for your positive encouragement, patience, suggestions, support, and love throughout.
Dedication

This dissertation is dedicated to my parents, Mabel and Charles William Ward.

They always told me, "...work hard and do your best. The rest will take care of itself."

You were right. Thank you for your unwavering support.
TABLE OF CONTENTS

ABSTRACT..................................................................................................................ii

ACKNOWLEDGEMENTS.........................................................................................iv

DEDICATION..........................................................................................................vi

LIST OF TABLES......................................................................................................x

I INTRODUCTION AND BACKGROUND ..................................................................1

  Introduction to the Research Problem.................................................................1
  Statement of the Problem......................................................................................7
  Purpose of the Study............................................................................................7
  The Research Question.........................................................................................9
  Hypothesis.............................................................................................................10
  Significance of the Study.....................................................................................12
  Limitations of the Study......................................................................................14
  Delimitations of the Study..................................................................................15
  Definition of Key Terms......................................................................................15
  Summary.............................................................................................................20
  Organization of the Study...................................................................................20

II REVIEW OF THE LITERATURE ...........................................................................22

  Introduction..........................................................................................................22
  Literature Search Method....................................................................................22
  Criteria for Inclusion in the Literature Search....................................................23
  Review of Relevant Research, Theory, and Literature........................................24
    History of Mathematics Reform in the United States........................................24
    The Math Wars...................................................................................................27
    Mathematics Learning.......................................................................................34
    Research on Reform Mathematics vs. Traditional..............................................38
    Achievement of Diverse Groups of Students.....................................................50
  Summary and Synthesis.......................................................................................54

III RESEARCH DESIGN AND METHODOLOGY ..................................................57

  Introduction..........................................................................................................57
  Research Design..................................................................................................58
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>61</td>
</tr>
<tr>
<td>Setting for the Study</td>
<td>62</td>
</tr>
<tr>
<td>Treatment</td>
<td>63</td>
</tr>
<tr>
<td>Research Methods</td>
<td>63</td>
</tr>
<tr>
<td>Data Collection</td>
<td>63</td>
</tr>
<tr>
<td>Variables</td>
<td>64</td>
</tr>
<tr>
<td>Sampling</td>
<td>65</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>65</td>
</tr>
<tr>
<td>Procedures</td>
<td>67</td>
</tr>
<tr>
<td>Methods of Data Analysis</td>
<td>68</td>
</tr>
<tr>
<td>Discussion of Controls</td>
<td>69</td>
</tr>
<tr>
<td>Step-by-Step Procedures</td>
<td>72</td>
</tr>
<tr>
<td>Summary of Chapter III and Description of Chapter IV</td>
<td>74</td>
</tr>
<tr>
<td>DATA ANALYSIS</td>
<td>75</td>
</tr>
<tr>
<td>Introduction</td>
<td>75</td>
</tr>
<tr>
<td>Results</td>
<td>75</td>
</tr>
<tr>
<td>Research Question</td>
<td>77</td>
</tr>
<tr>
<td>Subsidiary Question 1</td>
<td>78</td>
</tr>
<tr>
<td>Hypothesis 1</td>
<td>78</td>
</tr>
<tr>
<td>Hypothesis 2</td>
<td>80</td>
</tr>
<tr>
<td>Subsidiary Question 2</td>
<td>82</td>
</tr>
<tr>
<td>Hypothesis 3</td>
<td>82</td>
</tr>
<tr>
<td>Hypothesis 4</td>
<td>84</td>
</tr>
<tr>
<td>Subsidiary Question 3</td>
<td>86</td>
</tr>
<tr>
<td>Hypothesis 5</td>
<td>86</td>
</tr>
<tr>
<td>Hypothesis 6</td>
<td>88</td>
</tr>
<tr>
<td>Subsidiary Question 4</td>
<td>90</td>
</tr>
<tr>
<td>Hypothesis 7</td>
<td>90</td>
</tr>
<tr>
<td>Hypothesis 8</td>
<td>92</td>
</tr>
<tr>
<td>Subsidiary Question 5</td>
<td>94</td>
</tr>
<tr>
<td>Hypothesis 9</td>
<td>94</td>
</tr>
<tr>
<td>Hypothesis 10</td>
<td>96</td>
</tr>
<tr>
<td>Subsidiary Question 6</td>
<td>98</td>
</tr>
<tr>
<td>Hypothesis 11</td>
<td>98</td>
</tr>
<tr>
<td>Hypothesis 12</td>
<td>100</td>
</tr>
<tr>
<td>Summary of Findings</td>
<td>102</td>
</tr>
<tr>
<td>Subsidiary Question 1</td>
<td>102</td>
</tr>
<tr>
<td>Hypothesis 1</td>
<td>102</td>
</tr>
<tr>
<td>Hypothesis 2</td>
<td>102</td>
</tr>
<tr>
<td>Subsidiary Question 2</td>
<td>103</td>
</tr>
<tr>
<td>Hypothesis 3</td>
<td>103</td>
</tr>
<tr>
<td>Hypothesis 4</td>
<td>103</td>
</tr>
<tr>
<td>Subsidiary Question 3</td>
<td>104</td>
</tr>
<tr>
<td>Hypothesis 5</td>
<td>104</td>
</tr>
</tbody>
</table>
V SUMMARY OF FINDINGS, CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS

Introduction and Research Questions

Summary of Findings

Discussion and Conclusions

Mean Scale Scores

Effect Sizes

Standard Deviations

Recommendations for Policy and Practice

Suggestions for Future Research

Summary

REFERENCES

APPENDICES

Appendix A: Questions to DFG-I and DFG-J Schools and Districts

Appendix B: Number of Schools and Students by DFG and Year
LIST OF TABLES

Table 1: Rule of Thumb for Interpreting Correlation.................................77
Table 2: 2007 NJASK3 DFG-I Mean Scores..............................................78
Table 3: T-Test Comparing DFG-I Everyday Math and Traditional Program Schools on 2007 NJASK3.................................................................79
Table 4: 2007 NJASK3 DFG-I Correlations................................................79
Table 5: 2008 NJASK3 DFG-I Mean Scores.................................................80
Table 6: T-Test Comparing DFG-I Everyday Math and Traditional Program Schools on 2008 NJASK3.................................................................81
Table 7: 2008 NJASK3 DFG-I Correlations................................................81
Table 8: 2007 NJASK3 DFG-J Mean Scores................................................82
Table 9: T-Test Comparing DFG-J Everyday Math and Traditional Program Schools on 2007 NJASK3.................................................................83
Table 10: 2007 NJASK3 DFG-J Correlations...............................................83
Table 11: 2008 NJASK3 DFG-J Mean Scores.............................................84
Table 12: T-Test Comparing DFG-J Everyday Math and Traditional Program Schools on 2008 NJASK3.................................................................85
Table 13: 2008 NJASK3 DFG-J Correlations...............................................85
Table 14: 2007 NJASK4 DFG-I Mean Scores.............................................86
Table 15: T-Test Comparing DFG-I Everyday Math and Traditional Program Schools on 2007 NJASK4.................................................................87
Table 16: 2007 NJASK4 DFG-I Correlations...............................................87
Table 17: 2008 NJASK4 DFG-I Mean Scores.............................................88
Table 18: T-Test Comparing DFG-I Everyday Math and Traditional Program Schools on 2008 NJASK4.................................................................89
Table 19: 2008 NJASK4 DFG-I Correlations.................................89
Table 20: 2007 NJASK4 DFG-J Mean Scores.................................90
Table 21: T-Test Comparing DFG-J Everyday Math and Traditional Program Schools on 2007 NJASK4................................................91
Table 22: 2007 NJASK4 DFG-J Correlations..................................................91
Table 23: 2008 NJASK4 DFG-J Mean Scores.................................92
Table 24: T-Test Comparing DFG-J Everyday Math and Traditional Program Schools on 2008 NJASK4................................................93
Table 25: 2008 NJASK4 DFG-J Correlations..................................................93
Table 26: 2007 NJASK5 DFG-I Mean Scores.................................94
Table 27: T-Test Comparing DFG-I Everyday Math and Traditional Program Schools on 2007 NJASK5................................................95
Table 28: 2007 NJASK5 DFG-I Correlations..................................................95
Table 29: 2008 NJASK5 DFG-I Mean Scores.................................96
Table 30: T-Test Comparing DFG-I Everyday Math and Traditional Program Schools on 2008 NJASK5................................................97
Table 31: 2008 NJASK5 DFG-I Correlations..................................................97
Table 32: 2007 NJASK5 DFG-J Mean Scores.................................98
Table 33: T-Test Comparing DFG-J Everyday Math and Traditional Program Schools on 2007 NJASK5................................................99
Table 34: 2007 NJASK5 DFG-J Correlations..................................................99
Table 35: 2008 NJASK5 DFG-J Mean Scores.................................100
Table 36: T-Test Comparing DFG-J Everyday Math and Traditional Program Schools on 2008 NJASK5................................................101
Table 37: 2008 NJASK5 DFG-J Correlations..................................................101
Table 38: Standard Deviations by Test and Program............................124
INTRODUCTION AND BACKGROUND

Introduction to the Research Problem

The publication in 1983 by the National Commission on Excellence in Education (NCEE) of *A Nation at Risk* identified concerns about the state of U.S. public education. Our Nation is at risk. Our once unchallenged preeminence in commerce, industry, science, and technological innovation is being overtaken by competitors throughout the world. This report is concerned with only one of the many causes and dimensions of the problem, but it is the one that undergirds American prosperity, security, and civility. We report to the American people that while we can take justifiable pride in what our schools and colleges have historically accomplished and contributed to the United States and the wellbeing of its people, the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people. What was unimaginable a generation ago has begun to occur with others matching and surpassing our educational attainments.

If an unfriendly power had attempted to impose on America the mediocre educational performance that exists today, we might well have viewed it as an act of war. As it stands, we have allowed this to happen to ourselves. We have even squandered the gains in student achievement made in the wake of the Sputnik challenge. Moreover, we have dismantled essential support systems which helped to make those gains possible. We have, in effect, been committing an act of
unthinking, unilateral educational disarmament. (National Commission on Excellence in Education, 1983, p. 5)


Ironically, the original report, *A Nation at Risk* (NCEE, 1983), which served as the stimulus for many critiques of U. S. education and the attempts at reform that followed,
was not supported by evidence and the corresponding citations of research data. Statements were unsupported in the document and went unchallenged due to the prestige of the panel (A. Bartlett Giamatti, President, Yale University; Glenn Seaborg, Nobel Laureate; Jay Sommer, National Teacher of the Year) and the claims of government officials like Secretary of Education Terrell Bell, President Ronald Reagan, and later Secretary of Education William Bennett. Follow-up analysis of the claims made in *A Nation at Risk* (1983) challenged the conclusions of the report and launched a series of accusations and counter-assertions.

Among the most prominent was the *The Sandia Report* (Sandia National Laboratories, 1993). Originally drafted in 1990, the report was a review of the data documenting the performance of U.S. students and the claims made in *A Nation at Risk* (NCEE, 1983) and other reports. Produced by the Sandia National Laboratories, a branch of the U.S. Energy Department, the report provided data contradicting many of the assertions made by *A Nation at Risk* (1983) and the Reagan and Bush administrations. *The Sandia Report*, not published until after President George H. W. Bush had been defeated in the 1992 Presidential Election, was finally released in 1993 following charges that it had been suppressed by the Bush administration (Miller, 1991).

As an example of the disparate views, *A Nation at Risk* (NCEE, 1983) asserted that, "Average achievement of high school students on most standardized tests is now lower than 26 years ago when Sputnik was launched" (p. 8). Analysis of NAEP data in *The Sandia Report* indicated consistent overall levels of student achievement with some gains among minority groups and disadvantaged students (*The Sandia Report*, as cited in Berliner & Biddle, 1995, pp. 27-28). In contrast to the portrayal in *A Nation at Risk*
(1983), the authors of *The Sandia Report* (SNL, 1993) wrote, "To our surprise, on nearly every measure we found steady or slightly improving trends" (Carson, Huelskamp, & Woodall, 1993, p. 259).

*A Nation at Risk* (NCEE, 1983) has also been seen to support a broader political agenda of educational vouchers, school prayer, and tougher standards (Berliner & Biddle, 1995, pp. 136, 140, 149). Along with greater criticism of public schools, and the student achievement results they provided, came a stronger push for educational vouchers. "Both the Reagan and Bush administrations favored vouchers, and in the spring of 1991 the latter brought forth America 2000, a proposed educational policy that would have provided, among other things, tax-supported vouchers that could be used in private schools" (Berliner & Biddle, 1995, p. 149).

While the assertions of *A Nation at Risk* (NCEE, 1983) may have been suspect, elements of *The Sandia Report* (SNL, 1993) and comparisons of international student performance highlighted concerns about the levels of U.S. students' performance. "We are not implying that these performance levels are adequate for today's or tomorrow's society" (Carson et al., 1993, p. 270). An earlier review of NAEP results indicated students "…exhibit serious gaps in their knowledge and are learning a number of concepts and skills at a superficial level" (Carpenter et al. 1988, pp. 40-41), concluding, "…students’ achievement at all age levels shows major deficiencies" (p.41). While not using international assessment data because they found "little credible data regarding international comparisons," beyond the International Assessment of Educational Progress (Huelskamp, 1993, p. 720), the Sandia authors concluded, "…U.S.
student performance in both math and science continues to be low compared to that of other participating nations” (p. 720).

Following the release of *A Nation at Risk* (NCEE, 1983), the National Council of Teachers of Mathematics (NCTM) was one of the first professional organizations to respond to the NCEE report with the publication of *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989). The Standards were a reaction to the NCEE recommendation that, “In addition to the traditional sequence of studies available for college-bound students, new, equally demanding mathematics curricula need to be developed for those who do not plan to continue their formal education immediately” (NCTM, 1989, p.32). Along with *A Nation at Risk* (NCEE, 1983) and *Educating Americans for the 21st Century* (NSBCPEMST, 1983), the NCTM Standards were designed to reform the manner in which mathematics is taught in the United States.

The Standards’ call for increased emphasis on mathematical reasoning, understanding, and problem solving led to a number of National Science Foundation (NSF) funded curriculum projects. Three NSF-supported elementary mathematics curriculum projects were developed to promote widespread implementation of mathematics curricula reflective of the NCTM Standards. These elementary school, reform-based curriculum programs, *Investigations in Number, Data and Space* (Technical Education Research Centers, 1998), *Math Trailblazers: A Mathematical Journey Using Science and Language Arts* (Institute for Mathematics and Science Education, 1999) , and *Everyday Mathematics* (Bell et al., 1988-1996), were designed to increase students’ mathematics conceptual understanding and problem solving abilities.
Soon after the field testing and implementation of the NSF funded, reform-based mathematics curriculum materials, critics began to challenge the stated goals, teaching principles, and assessment practices of the proposed reforms. The debate surrounding the mathematics reform movement became so intense; it came to be known as the math wars (Schoen, Fey, Hirsch, & Coxford, 1999).

The formation of the National Mathematics Advisory Panel by President George W. Bush in April 2006 marked the continuing debate concerning K-12 mathematics instruction in the U.S. In his charge to the Panel, President Bush specified several areas required in the Final Report:

*Report to the President on Strengthening Mathematics Education*…the Panel shall submit to the President…recommendations, based on the best available scientific evidence on the following: (a) the critical skills and skill progressions for students to acquire competence in algebra and readiness for higher levels of mathematics; (b) the role and appropriate design of standards and assessment in promoting mathematical competence; (c) the processes by which students of various abilities and backgrounds learn mathematics; (d) instructional practices, programs, and materials that are effective for improving mathematics learning;…”

(Executive Order 13398, p. 1)

The requirement that the Panel address “instructional practices, programs, and materials that are effective for improving mathematics learning” (Executive Order 13398, p. 1) in its final report is indicative of the ongoing math wars (Colvin, 1999; O’Brien, 2007) throughout the US. With the debate about mathematics methods, materials, programs, and practices created by the NCTM Standards continuing 25 years after their publication,
it is not surprising that the President looked to the National Mathematics Advisory Panel to help resolve the issues surrounding reform-based mathematics.

Statement of the Problem

Concerns about U.S. student achievement in mathematics have been prominent over the last 25 years. The publication of the NCTM Standards documents led to NSF funding for the development of K-12 curricula reflective of the Standards. The new curricula emphasized problem solving and higher order thinking skills such as synthesis, generalization and evaluation. These NSF funded programs were built from a constructivist approach with the underlying idea that children construct mathematical knowledge based on their experiences. An associated underpinning was what students learn is dependent on how they learn it. The dichotomy created by Standards-based and traditional programs, and the ensuing math wars, led to a problem for school leaders in the selection of instructional materials. What is the influence of different types of K-5 mathematics programs on student achievement? The current body of research evidence surrounding reform-based and traditional mathematics programs fails to give educational leaders clear directions for program selection.

Purpose of the Study

The key measure of success for any curriculum or program of instruction is the level of student achievement. With calls for U.S. mathematics achievement to be the highest in the world, it was clear that new methods, materials, programs, and practices are required. The purpose of this study is to better understand the role
instructional/curricular materials play in the development of elementary students’ mathematics skills. This investigation compares the mean Mathematics scores for sample populations on third, fourth and fifth grade New Jersey state tests. Attention is given to various populations of General Education students within DFG-I and J districts.

This research proposes to determine how implementation of a reform-based, elementary school mathematics curriculum is related to the acquisition of students’ mathematics skills as assessed by the New Jersey Assessment of Skills and Knowledge (NJASK). The study examines the differences between traditional mathematics programs and a reform-based mathematics program on mathematics achievement as measured by the NJASK tests for the General Education student population in New Jersey DFG-I and J districts. For this study, mathematics programs were categorized as reform-based or traditional mathematics programs. The study will investigate the differences, if any, between use of the reform-based elementary school mathematics program *Everyday Math* and student achievement. Traditional programs will include those published by Harcourt Brace, Scott Foresman, Silver-Burdett-Ginn, Addison Wesley, and Houghton Mifflin, McGraw Hill.

Third, fourth, and fifth grade General Education students in suburban New Jersey public school districts with similar SES, ethnicity, special education classification rates, and linguistic diversity comprise the sample populations. Mathematics performance refers to student achievement on the Mathematics portion of the NJASK3, NJASK4 and NJASK5. The mathematics performance data was collected from publicly available sources at the New Jersey Department of Education.
The Research Question

The proposed research attempts to answer one overarching question related to student achievement and the implementation of a reform-based or a traditional elementary school mathematics program. Research Question: What is the influence of implementing a reform-based mathematics program on the mathematics achievement, as measured by NJASK tests, of General Education students in grades 3 through 5 who attend school in New Jersey school districts classified as DFG-I and DFG-J? This research will attempt to determine if implementation of a K-5 reform-based mathematics program is related to differences in performance on New Jersey state tests (NJASK) for General Education students in comparison to students using a traditional program.

Research Question: What is the influence of implementing a reform-based mathematics program on the mathematics achievement, as measured by NJASK tests, of General Education students in grades 3 through 5 who attend school in New Jersey school districts classified as DFG-I and DFG-J?

Subsidiary Question 1. What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-I General Education students on NJASK3 mathematics tests?

Subsidiary Question 2. What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-J General Education students on NJASK3 mathematics tests?

Subsidiary Question 3. What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-I General Education students on NJASK4 mathematics tests?
Subsidiary Question 4. What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-J General Education students on NJASK4 mathematics tests?

Subsidiary Question 5. What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-I General Education students on NJASK5 mathematics tests?

Subsidiary Question 6. What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-J General Education students on NJASK5 mathematics tests?

Hypothesis
The null hypothesis for this nonexperimental quantitative study, $H_0: \mu = 0$, indicates that the results of this study are due to chance such that there is no patterned influence of the use of the reform-based elementary mathematics program, *Everyday Mathematics*, on student achievement as measured by the NJASK3, NJASK4, or NJASK5. If the null hypothesis is proved, no statistically significant difference exists between the use of *Everyday Mathematics* and a traditional mathematics program on student achievement. The alternative hypothesis is $H_A: \mu \neq 0$, indicates that the results are not due to chance and a statistically significant difference exists between use of the reform-based mathematics program, *Everyday Mathematics*, and student achievement. Regardless of the outcome, the information obtained may be used as one factor in determining the selection of an elementary mathematics program. This information will
be useful in determining the allocation of scarce resources (money, time, professional
development initiatives).

The research question and six subsidiary questions are supported by 12 hypotheses:

Hypothesis 1. There is no statistically significant difference in 2007 NJASK3
mathematics scores between DFG-I General Education students using the *Everyday Mathematics* program and those using a traditional mathematics program.

Hypothesis 2. There is no statistically significant difference in 2008 NJASK3
mathematics scores between DFG-I General Education students using the *Everyday Mathematics* program and those using a traditional mathematics program.

Hypothesis 3. There is no statistically significant difference in 2007 NJASK3
mathematics scores between DFG-J General Education students using the *Everyday Mathematics* program and those using a traditional mathematics program.

Hypothesis 4. There is no statistically significant difference in 2008 NJASK3
mathematics scores between DFG-J General Education students using the *Everyday Mathematics* program and those using a traditional mathematics program.

Hypothesis 5. There is no statistically significant difference in 2007 NJASK4
mathematics scores between DFG-I General Education students using the *Everyday Mathematics* program and those using a traditional mathematics program.

Hypothesis 6. There is no statistically significant difference in 2008 NJASK4
mathematics scores between DFG-I General Education students using the *Everyday Mathematics* program and those using a traditional mathematics program.
Hypothesis 7. There is no statistically significant difference in 2007 NJASK4 mathematics scores between DFG-J General Education students using the *Everyday Mathematics* program and those using a traditional mathematics program.

Hypothesis 8. There is no statistically significant difference in 2008 NJASK4 mathematics scores between DFG-J General Education students using the *Everyday Mathematics* program and those using a traditional mathematics program.

Hypothesis 9. There is no statistically significant difference in 2007 NJASK5 mathematics scores between DFG-I General Education students using the *Everyday Mathematics* program and those using a traditional mathematics program.

Hypothesis 10. There is no statistically significant difference in 2008 NJASK5 mathematics scores between DFG-I General Education students using the *Everyday Mathematics* program and those using a traditional mathematics program.

Hypothesis 11. There is no statistically significant difference in 2007 NJASK5 mathematics scores between DFG-J General Education students using the *Everyday Mathematics* program and those using a traditional mathematics program.

Hypothesis 12. There is no statistically significant difference in 2008 NJASK5 mathematics scores between DFG-J General Education students using the *Everyday Mathematics* program and those using a traditional mathematics program.

Significance of the Study

The challenge of how to create and sustain improvement in our nation’s precollege mathematics achievement continues as a point of national debate among parents, mathematics educators, and researchers. This study will help educators in their
search for effective mathematics curricula and instructional materials. In 2005, a report from the National Center for Educational Statistics (NCES) once again demonstrated concerns about the performance of U.S. students (NCES, 2005) on NAEP tests. U.S. performance on international Programme for International Student Assessment (PISA) and TIMSS tests continues to raise public concerns (Bybee & Stage, 2005; Steen, 2003). While scores for U.S. fourth- and eighth-graders were above international averages, our Nation’s high school seniors scored below the international average for 21 countries on a test of mathematics and science achievement (Mullis, Martin, Gonzalez, & Chrostowski, 2004). In testimony before a Senate committee it was stated that the economic future of our country depends on its ability to make greater strides in K-12 mathematics and science (Augustine, 2006). Determining the difference, if any, of reform-based mathematics instructional programs and traditional programs on student achievement is important in guiding the decisions of education leaders.

Views of how students learn mathematics and the manner in which instruction should take place have evolved over time based on a combination of experience, theory, and research. Advances in cognitive research have led to greater understanding of the way in which students learn. These developments have shifted the instructional emphasis away from rote learning to the need for students to connect new skills and concepts with past learning and to develop habits of mind that “involve exploring, inventing, conjecturing, proving, and problem solving” (Schoen, Fey, Hirsch, & Cosford, 1999, p. 446). With increasing knowledge about student learning, the NCTM developed *An Agenda for Action* (NCTM, 1980) calling for “deceased emphasis” on “performing paper and pencil calculations” along with greater emphasis on “operation sense” and
“collection and organization of data,” (NCTM, 1980, as cited in Klein, 2007, p. 22). The Agenda, along with future Standards documents (NCTM, 1989, 1991, 1995) reflected the theory that active involvement in problem solving, incorporation of manipulative materials, opportunities for students to develop their own procedures and actively construct their own knowledge would allow students to develop important mathematics skills and deeper conceptual understanding (Carpenter et al., 1989; Cobb et al., 1991; Hiebert, 1999; Hiebert & Wearne, 1993; Hiebert & Wearne, 1996; Kamii & DeClark, 1985; Mack, 1990; Wearne & Hiebert, 1989). “...the primary theory of action behind this set of reforms is that higher level objectives, including a focus on developing critical mathematics concepts and skills and pedagogical aids such as the use of manipulatives and improved sequencing of objectives, and other features of textbooks will improve student outcomes” (Slavin & Lake, 2008, p. 429). Supported by the NSF, reform-based curriculum were developed using these theories. This study examines the influence of these theories, through reform-based curriculum, on student achievement.

Limitations of the Study

This study is designed to compare the achievement of General Education students using a reform-based mathematics instructional program with the achievement of General Education students using traditional programs in high-socioeconomic New Jersey school districts (i.e. DFG-I and J) on NJASK tests at grades 3, 4, and 5. This study does not control for the possible rival explanations posed by the impact of student intelligence and prior mathematics achievement; varying levels of professional development related to mathematics instructional topics; the varying amounts of time on mathematics tasks
related to classroom instruction time and homework; teacher quality and teacher knowledge of mathematics. There are no observations of classroom instruction related to the level of implementation or types of instructional activities for either the traditional programs or the reform-based program. Measures of mathematics achievement are limited to the third, fourth and fifth grade NJASK scores for 2007 and 2008. This study will use school-level student achievement data from selected New Jersey school districts, and as such, it may not be possible to make generalizations about the results to other student populations in other regions of the state or country, or other types of districts.

Delimitations of the Study

The scope of this study is the comparison of a reform-based elementary mathematics instructional program with traditional elementary mathematics instructional programs and the differences among NJASK scores for General Education third, fourth, and fifth grade students in regular education classroom settings. The study delimited the population to students with more than one year in the school district. The study population was further delimited to students who are not classified as eligible for special education services and/or not classified as Limited English Proficient (LEP).

Definition of Key Terms

Constructivist learning- A theory explaining how individuals learn, based on ideas attributed to Jean Piaget, suggesting that learners construct knowledge from their experiences through assimilation (fitting new ideas into an existing mental framework)
and accommodation (creation of a new or adjusted mental framework). Constructivism implies active learning whether at the physical or mental level.

Correlation- A quantitative measure indicating the degree or relationship between two variables (Gay, Mills, & Airasian, 2006).

Disaggregated data- Data from a larger set that has been grouped by a particular factor. Sub-groups for disaggregation include gender, economic status, mobility, ethnicity, special education/disabled, Limited English Proficient (LEP), and General Education.

District Factor Grouping (DFG)- A system for ranking New Jersey school districts by their socio-economic status (SES). Introduced by the New Jersey Department of Education in 1975 based on 1970 Census data, DFG is periodically updated taking into account new Census data. The 1984 revision slightly changed the theoretical model for determining SES. The most recent revision took place in 2004 using the 2000 Census. DFG ranks school districts from A to J, with J districts having the highest SES (New Jersey Department of Education, 2004).

Economic Status-Defined as Economically Disadvantaged or Non-Economically Disadvantaged under New Jersey Department of Education guidelines. Economically Disadvantaged is the status attributed to a student qualifying for free or reduced-price lunch as defined by the U.S. government under the National School Lunch Program (NLSP).

Effect size- The statistical measure of the strength of the relationship between two variables. Something brought about by another factor. An occurrence caused by another action or event. In this research study, effect refers to the results of a statistical treatment.
Ethnicity- A student’s racial designation as reported to the State of New Jersey based on information gathered upon student registration in a school district. On the New Jersey school report cards these categories include White, Black, Asian/Pacific Islander, American Indian/Alaska Native, Hispanic, and Other.

*Everyday Mathematics*- A kindergarten through sixth grade mathematics instructional curriculum developed by the University of Chicago School Mathematics Project (USCSMP) under a grant from the National Science Foundation (NSF), including teacher, student, and resource materials. *Everyday Mathematics* is based on the NCTM *Standards* documents and is considered a reform-based mathematics instructional program.

General Education- Students not included as LEP or special education in the reporting of NJASK assessment data.

Limited English Proficient (LEP)- In New Jersey, Limited English Proficient students are those whose performance on an approved test of listening, speaking, reading, and writing of English identifies them as needing additional, specialized English instruction from an appropriately certificated teacher. New Jersey state regulations mandate the use of multiple criteria for identifying and exiting students from a language assistance program.

*Math Wars*- The term given to disagreements about the manner in which mathematics instruction takes place. Following the publication by NCTM of *Curriculum and Evaluation Standards* (NCTM, 1989), extended and heated debate occurred about the textbooks, curricula, philosophies, and instructional methodologies being implemented in U.S. classrooms. The disagreements over traditional and reform-based mathematics.
National Council of Teachers of Mathematics (NCTM)- The mission of the National Council of Teachers of Mathematics is to provide a vision for mathematics education, as well as leadership and professional development to support teachers of mathematics in ensuring equitable mathematics learning of the highest quality for all students (http://www.nctm.org/about/default.aspx?id=166).

New Jersey Assessment of Skills and Knowledge (NJASK)- The NJASK tests are a series of state assessments administered to New Jersey public school students in grades 3-8 to determine the level of student achievement in language arts, mathematics, and science. The NJASK tests were implemented in 2003 in response to the requirements of NCLB legislation.

New Jersey Core Curriculum Content Standards (NJCCCS)- The New Jersey Core Curriculum Content Standards were originally adopted in 1996 in an effort to define what students should know and be able to do at the end of their K-12 public school education. The NJCCCS are revised on a 5-year cycle. The Standards seek to articulate the important knowledge and skills all students should master (New Jersey Department of Education, 2008a).

New Jersey Mathematics Standards. The New Jersey Mathematics Standards were originally adopted in 1996 and have undergone periodic revision. Following their adoption, the New Jersey Mathematics Curriculum Framework (Rosenstein, Caldwell, & Crown, 1996) was developed by the New Jersey Mathematics Coalition. Published in December 1996 and adopted by the New Jersey State Board of Education in February 1997, the Mathematics Curriculum Framework provides guidance for the teaching of K-12 mathematics and "builds on the Curriculum and Evaluation Standards for School


Reform-based mathematics program- A mathematics instructional program developed through the financial support of the National Science Foundation (NSF). These funded mathematics curriculum projects were developed to promote widespread implementation of mathematics curricula reflective of the NCTM Standards and were designed to increase students’ mathematics conceptual understanding and problem solving abilities.

Socioeconomic status- A student’s socioeconomic status is based on a combination of family income level, parents’ educational attainment, and parents’ occupation (Demarest, Reisner, Anderson, Humphrey, Farquhare, & Stein, 1993).

Standards-based mathematics program- A mathematics instructional program developed through the support of the NSF reflective of the NCTM Standards and designed to increase students’ mathematics conceptual understanding and problem solving abilities; a reform-based mathematics program.

Traditional program- A mathematics instructional program characterized by the use of grade level textbooks organized into chapters developed around instruction in traditional mathematics algorithms. Typically the teacher provides a demonstration of
some aspect of a mathematics concept or algorithm, students complete practice examples that demonstrate the concept or algorithm, and similar problems are assigned for further practice at home (Heid, 1997; Schoen, Fey, Hirsch, & Coxford, 1999).

Summary

The 1983 publication of *A Nation at Risk* (NCEE) raised alarms about the state of education in the US and was followed by additional reports designed to give new directions for public schools (NSBCPEMST, 1983; SCANS, 1991). Perhaps because *The Sandia Report* (SNL, 1993) provided a far less alarming view of our nation’s K-12 education system, it was not published until 10 years later. However, in the years surrounding *A Nation at Risk* (1983), national (NAEP, 1978-1988) and international reports (SIMS, 1987; TIMSS, 1995; PISA, 2000) demonstrated less than satisfactory mathematics performance for many groups of U.S. students. Publication of the *Standards* documents (NCTM, 1980, 1989, 1991, 1995, 2006) led to funding from the NSF for the creation of school mathematics programs designed to reflect a new vision for K-12 mathematics. The resulting dichotomy of traditional versus reform-based programs has led to emotional debates and challenges for school administrators as they seek to select and implement school mathematics programs that will yield improved student performance.

Organization of the Study

This study is organized into five chapters. Chapter I includes Introduction to the Research Problem, Statement of the Problem, Purpose of the Study, The Research
Question and Subsidiary Questions, Hypotheses, Significance of the Study, Limitations and Delimitations of the Study, Definition of Key Terms, and the Organization of the Study.

Chapter II provides a Review of Literature related to the study, including the history of Standards-based mathematics reform. Chapter III presents the Design-Methods and Procedures, Sampling, Instrumentation, Data, Data Collection, Data Analysis, and the Specific Treatment of each Sub-Problem. Chapter IV presents the study’s finding. Chapter V presents the Summary, Conclusions, and Implications and Recommendations for Further Study. References and Appendix are included for complete documentation.
Chapter II

REVIEW OF THE LITERATURE

Introduction

The review of relevant research begins with a history of mathematics education and the push for reform of mathematics instruction in the United States based on national and international assessment data. Key to this discussion are theories about how children learn mathematics and the evolving influence of several theories throughout the last century. These varying viewpoints are considered in the context of the on-going national debate, or math wars, over traditional versus Standards-based instructional methods and materials. Consideration of how students learn mathematics leads to a review of research on traditional and reform-mathematics programs as the basis for the current study.

Concerns about U.S. mathematics achievement are long-standing and ideas about school mathematics instruction much debated, with disagreements dating to the early 1900s. In some instances, international events have served as the basis for these debates—Sputnik, TIMSS, PISA—while at other times the impetus has been motivated by national reports—A Nation at Risk, NAEP (1983), What Work Requires of Schools (SCANS, 1991), NCES (2005a). Shifting views of what is important for children to learn, which children should learn it, and how it should be learned are discussed.

Literature Search Method

The review of the relevant research, theory, and literature related to this study encompasses articles and research from the early 1900s to 2009, with a focus on the last 50 years. Journal searches were completed, beginning at Seton Hall (SHU) and later at
Teachers College, Columbia University, using the reference sections of prior dissertations via Dissertation Abstracts Database from UMI ProQuest Digital Dissertations and the online data bases Academic Search Premier, ProQuest Multiple Databases, JSTOR, LexisNexis Academic, EBSCO, ERIC, Education Journals, and the Directory of Open Access Journals (DOAJ). Topic descriptors included: NCTM Standards, reform-based mathematics, mathematics learning, math wars, mathematics achievement, elementary mathematics education, mathematics curriculum reform, Everyday Mathematics, Trailblazers, Investigations in Number, Data, and Space, mathematics teaching, and NFS mathematics programs. In addition, the reference sections of journal articles and past research, as well as dissertation committee members, provided studies and information relevant to the research topic. In several instances, direct communication with researchers provided additional resources (J. Hiebert personal communication, November 26, 2008 and J. Johnson personal communication, November 24 & 25, 2008).

Criteria for Inclusion in the Literature Search

Literature which helped develop the historical context for the on-going debate concerning mathematics instruction was included in the literature review. Articles related to the math wars, whether from historical, cultural, or instructional perspectives, were included in order to develop an understanding of the debate's impact on elementary school instructional programs. Research surrounding how children learn mathematics was judged based on its historical importance to the field. Nearly all identified studies of student performance using K-5 Standards-based and traditional mathematics were
included to provide an understanding of the current research knowledge base surrounding various mathematics programs.

Review of Relevant Research, Theory and Literature

History of Mathematics Reform in the United States

Alarms were sounded about student achievement in mathematics as early as the mid 19th century. In 1845, Horace Mann, then Secretary to the State Board of Education in Massachusetts wrote,

Who of all the boys, aye, or girls either, shall cast the interest on a note, either as borrower or lender, when not one of them knows there is any difference between the value of a note of $200 payable in six months, and the value of two notes of $100 each,—one payable in three and the other in nine months! (Caldwell & Courtis, 1925, as cited in Senk & Thompson, 2003, p. 3)

The Committee of Ten, appointed by the National Education Association (NEA) in the 1890s, examined the mathematics programs in elementary and secondary schools and recommended modifications in both programs of study, including additional studies in algebra and geometry at both levels (Jones & Coxford, 1970). Disagreements about the manner in which students learn mathematics were evident in the early 1900s. Thorndike’s influence on American psychology and learning theory emphasized a sequential presentation of skills and concepts, while Judd, Sleight and others contended that student learning was a result of children’s experiences (Inglis, 1918; Senk & Thompson, 2003).
The Soviet Union’s launch of Sputnik in 1957 sparked renewed interest in American’s mathematics and science programs. In response, the National Science Foundation (NSF) provided funds for curriculum development projects aimed at improving student achievement at both the elementary and high school level. These materials provided greater emphasis on students’ understanding of the basic concepts and interrelationships of mathematics (Begle, 1973) and came to be known as the *new math*. Soon after implementation of these new curriculum materials, criticism of the methods and materials began to emerge. They were deemed difficult for parents and teachers to comprehend, with too much emphasis on theory and too little focus on basic skills (Kline, 1973).

These criticisms of the new math programs spurred a back-to-basics movement in elementary and secondary schools in the early 1970s. In response, many textbooks emphasized basic arithmetic computation and development of beginning algebra skills (Senk & Thompson, 2003, p. 9). Begun originally in 1964 with a grant from the Carnegie Corporation, the U.S. Congress established the National Assessment of Educational Progress (NAEP) in 1969, with the first nation-wide testing of students’ mathematics and reading skills taking place in 1972 (Vinovskis, 1998, p.8). This initial testing served as a point of comparison for future test administrations. Along with other states, New Jersey developed state-wide tests in the mid-1970s to measure student basic skills achievement. These Minimum Basic Skills (MBS) tests were administered to grades three, six, and nine from 1978-1982 and were the precursors of the current state-wide tests (New Jersey Department of Education, 2008b).
The results from the first NAEP tests of mathematics skills in 1972-73 and 1978 were a concern for both parents and educators (Kahl, 1979; Senk & Thompson, 2003; Wilson & Blank, 1999). Lower than expected results generated criticism of the emphasis on basic mathematics skills. A National Council of Supervisors of Mathematics (NCSM, 1977) position paper advocated for a broader definition of basic skills, and in 1980 the NCTM issued *An Agenda for Action* (NCTM, 1980) which mirrored the position of NCSM and described the future design of school mathematics (NCTM, 2006), launching the first steps, “in perhaps the greatest and longest-lasting ferment of educational research and development ever…” (Middleton et al. 2004, p. 76). Student performance on NAEP testing served as one data source in *Educating Americans for the 21st Century's* (NSBCPEM, 1983) call for changes in school mathematics programs.

In the years following *A Nation at Risk* (NCEE, 1983) and *Educating Americans for the 21st Century* (NSBSCPEM, 1983), a series of reports were released comparing U.S. performance to that of students in other nations. In 1987, data from the Second International Mathematics Study (SIMS) were reported, indicating below average performance by U.S. students on all SIMS tests (McKnight et al., 1987; Medrich & Griffith, 1992; Robitaille & Garden, 1989). In comparison to results from the First International Mathematics Study (FIMS) in the mid-1960s (Husen et al., 1967), U.S. performance of eighth graders had shown no improvement (McKnight et al., 1987; Medrich & Griffith, 1992), again calling into question the efficacy of the 1970s instructional focus on basic skills.

Following criticisms of U.S. curriculum and student achievement, the NCTM published *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989).
The *Standards* were, in part, a response to the NCEE's call for new, demanding curricula for all students (NCEE, 1983). The NCTM's agenda for reform continued with the publication of *Professional Standards for Teaching Mathematics* (NCTM, 1991) and *Assessment Standards for School Mathematics* (NCTM, 1995), presenting a long-term, comprehensive vision for K-12 mathematics instruction. As new curricula, instructional methods, and assessments were being put into place, critics began to challenge the "content goals, the pedagogical principles, and the assessment practices...at the heart of the reform agenda" (Schoen, Fey, Hirsch, & Coxford, 1999, p. 444). An analysis of the data from the Third International Mathematics and Science Study (TIMSS, 1995) characterized U.S. mathematics curricula as being "a mile wide and an inch deep" (Schmidt, Houang, & Cogan, 2002, p. 3). Concerns about student achievement and the most effective manner in which to improve mathematics teaching and learning continue more than 160 years after Horace Mann.

**The Math Wars**

Concerns about U.S student achievement in mathematics have been long standing and calls for reform recurring (Jones & Coxford, 1970; Kilpatrick & Stanic, 1995). Reports of student achievement on NAEP tests, the FIMS, and SIMS created a strong consensus for change in U.S. mathematics education (Schoen et al., 1999). Using input from other countries, classroom teachers, a growing body of research on teaching and learning, and the ideas of working mathematicians, the NCTM sought to create documents that would provide a roadmap for K-12 mathematics education in the United States.
Publication of the NCTM Standards documents did little to quiet the disagreements about how to teach mathematics. In response to the criticisms, NCTM published Curriculum Focal Points for Prekindergarten through Grade 8 Mathematics (NCTM, 2006). Designed to clarify key points of emphasis and identify the most important math topics at each grade level, Focal Points had little impact. In fact, their publication continued the controversy across the US.

The math wars have played out on local, state, and national levels with U.S. Secretary of Education Richard Riley calling, in 1998, for a “cease-fire” in the math wars because “people seem to be hunting for ways to disagree. ...This unhealthy habit of thinking in dogmatic ways does our children little good” (Colvin, 1999, ¶ 13). Despite this appeal, little has served to quiet the disagreements over mathematics instruction. Charges and counter charges (Becker & Jacob, 2000; Kilpatrick, 1997) have led to an “unprecedented level of national attention to mathematics education” (Reys, 2001, p. 6).

At the state level, these disagreements have frequently involved development of state mathematics standards or textbook selection. California, once considered a leader in mathematics reform, developed a framework for math educators in 1985 which foreshadowed the NCTM Standards (Colvin, 1999). In 1997, the state of California was developing its own mathematics standards for K-12 that reflected the NCTM Standards. According to Sherry Fraser (as cited in O’Brien, 2007), director of the Interactive Mathematics Program, in testimony to the National Mathematics Advisory Panel on November 6, 2006, the process was “hijacked by a state board member and given to four mathematicians to fix” (p. 665) so that standards would address content measured by multiple choice tests.
In November 2007, the Texas state board of education rejected the third grade version of *Everyday Mathematics*, the only text turned down of 164 considered. In spite of positive reviews by the U.S. government’s *What Works Clearinghouse* and use by an estimated 2.8 million students in 175,000 classrooms (University of Chicago Mathematics Project, 2008), Texas officials overruled a three-member textbook review committee’s recommendation due to concerns about the amount of attention to the basic skill of multiplication versus problem solving (Cavanagh, 2007). Almost simultaneously, the California state board of education approved the *Everyday Mathematics* program for use in grades K-6, a program it had rejected in 2001 (Cavanagh, 2007).

At the local district level, these disputes are most often evidenced in affluent, suburban school districts. Referring to educated parents, Rutgers University math professor Joseph Rosenstein (as cited in Hu, 2007) said, “They want their children’s education to resemble their education because they are successful. They say, ‘It worked for me, why won’t it work for them?’” (¶ 15). In January 1998, the Escondido, California school board in San Diego County moved to phase out the reform-based math program *Interactive Math* due to parent protests. Similarly, parents in the San Fernando Valley area of Los Angeles demanded that students be provided with traditional mathematics programs instead of reform-based programs. Ten years later, similar controversies have taken place in Ridgewood, NJ (Hu, 2007), Wayne, NJ (Alexander, 2008), Brooklyn, NY and Ridgefield, CT (Noveck, 2008) over the mathematics programs selected for elementary students.

Much of the disagreement is focused on emotional issues with little consideration of the significant research surrounding the NSF funded programs. Complaints have been
based on anecdotes rather than research, stories rather than real information when the issues are complex and involve staffing, leadership, and fidelity of implementation (Hiebert, 1999). Parents have raised the emotional level around mathematics instruction with accusations that school districts are using their children as guinea pigs when a Standards-based mathematics program has been selected. In reality, NSF programs have undergone extensive field testing prior to full nation-wide implementation. In comparison, traditional textbooks are subject to the market-driven nature of publishing and sales which does not allow for similar field testing due to the high costs of such endeavors (Reys, 2001). In reality, traditional programs have a less than stellar track record given the documented levels of student performance (Johnson & Rising, 1967; NAEP, 1983; Senk & Thompson, 2003; Wilson & Blank, 1999). At the National Mathematics Advisory Panel meeting, Fraser (as cited in O’Brien, 2007) stated the case for NSF programs:

Each of the projects included updates in content and in the context in which mathematics topics are presented. Each also affected the role of the teacher. Each has been through rigorous development that included design, piloting, redesign, field-testing, redesign, and publication. This amount of careful development is rare in textbook production... And in 2004 the National Academy of Sciences released a book, On Evaluating Curricular Effectiveness: Judging the Quality of K-12 Mathematics Programs, which looked at the evaluation studies for the thirteen NSF projects and six commercial textbooks. Based on the 147 research studies accepted, it is quite clear which curriculum programs have
promise to improve mathematics education in our country. They are the NSF-funded curriculum projects. (p. 665)

The intensity of the debate has resulted in an “emotionally charged atmosphere” with little evidence of “careful analysis of the complete curriculum and evaluation evidence” (Schoen, Fey, Hirsch, & Coxford, 1999, p. 65). The NCTM Standards were developed in response to poor student performance and were based on models of instruction from other countries where complex problem solving is more the norm. Advances in cognitive research have led to greater understanding of the need for students to connect new learning to past learning and to develop mathematical habits of the mind that “involve exploring, inventing, conjecturing, proving and problem solving” (Schoen, Fey, Hirsch, & Coxford, p. 446).

Some have viewed the disputes between parent groups and educators concerning the nature of mathematics instruction as based in issues of power rather than purely about mathematics methodology (Sarason, 1995). The struggle for power in public education has become more evident as parents have sought control over which schools their children attend. However, no clear role has been defined for parents in their attempts to decide what gets taught and how the instruction takes place (Peressini, 1998). The “back to basics” movement of 1970s was a reaction to parents’ dissatisfaction with the “new math,” but didn’t really include them in the change process. While the Standards-based reforms of the 1980s and 90s sought to inform parents of the goals for mathematics instruction, there was no real effort to involve parents in shaping the reforms.

The mathematics teaching profession recognizes and respects the rights of parents and society to hold it accountable for the mathematical competence of children.
However, in calling for particular programs of action, parents and society often mistakenly promote activities that are counterproductive to the realization of the goals they support. (NCTM, 1980, p. 27)

The NCTM expected parents to support the reform recommendations as they had been developed by the mathematics educators. Following a review of the NCTM Standards documents (NCTM, 1989, 1991), Peressini stated, “...parents receive minimal attention in the analysis and prescriptions that are advanced by the organization” (1988, p. 566). In fact, much of the need for educational reform, as prescribed by A Nation at Risk (NCEE, 1983), was driven not by parents but by the interests of industry and business and a need to remain competitive on a global level (SCANS, 1991). The failure to invite parents into the development of the mathematics reform agenda resulted in a lack of ownership on their part, and the Standards-based programs have not matched the parents’ school experiences, experiences from which many have built successful careers. As a result, parents’ only option has been to exercise power by influencing their local school board’s decision making (Peressini, 1988).

It is evident that parents, while having clear interest in what takes place in the classroom, had no role in the development of the Standards documents. And, while aspects of the math wars may be related to issues of power and control of public education, there is clearly more involved in the heated debates about mathematics instruction in the US. The recommendations inherent in the Standards have been challenged at local and state board meetings, through articles and editorials in newspapers, through professional journals, and via well-organized Internet websites such as NYCHOLD (http://www.nychold.com) and Mathematically Correct
Much of the debate centers around the manner in which children learn mathematics.

The disagreements over how math should be taught may also be viewed in an historical context with those advocating Standards-based mathematics programs as standard bearers for progressive education, similar in thinking to John Dewey and William Heard Kilpatrick, or even Jean Jacques Rousseau and the Romantic Movement (Klein, 2007). Beginning with An Agenda for Action (NCTM, 1980) and calls for "deceased emphasis" for "performing paper and pencil calculations with numbers of more than two digits" along with greater emphasis on "operation sense" and "collection and organization of data," the NCTM echoed the themes of progressive education by advocating for student-centered, discovery learning (NCTM, 1980 as cited in Klein, 2007, p. 22). The NCTM documents were promoted in support of two major thrusts: social justice and the needs of business and industry. The social justice advocates viewed traditional programs as elitist and aimed at favoring a limited segment of the population in their college applications. "Hence, lack of access to mathematics is a barrier—a barrier that leaves people socially and economically disenfranchised" (Schoenfeld, 2004, p. 255). Stronger mathematics preparation was seen as providing a more skilled workforce for business and industry. "Businesses no longer seek workers with strong backs, clever hands, and ‘shopkeeper’ arithmetic skills" (NCTM, 1989, p. 3). Along with greater opposition, came greater "polarization along political lines" (Klein, 2007, p. 27) with advocates of reform-based programs identifying critics as politically right wing because they were seen as arguing for more basic skills at the expense of conceptual understanding.
Disagreements about the manner in which students learn mathematics and its cultural importance may be traced to the late 19th century and the belief that the mind was similar to a muscle that could be improved through mental exercise (Cathcart, Pothier, Vance, & Bezuk, 2006). The humanists believed in the importance of mental discipline, as well as students' ability to reason (Schoenfeld, 2001). This view that students' mathematics learning could be developed via mental discipline was evident in much of the leading research of the time (Grouws & Cebulla, 2000; Jones & Coxford, 1970). Other views were also influential during the early twentieth century. The developmentalists looked for school activities to be aligned with the growing knowledge about child development. Social efficiency educators saw schools as designed to educate students according to their future role in society, as opposed to the social meliorists view of schools as sources of social equity or social justice (Jones & Coxford, 1970). With ideas popular since the time of Plato (Schoenfeld, 1992), many researchers thought the humanists' view, that discipline could strengthen reasoning ability, could be proven to generalize to other pursuits (Inglis, 1918).

Thorndike (as cited in Kilpatrick, 1992) demonstrated the shortcomings of this transfer training model. His work testing high school students supported his identical elements of transfer of learning theory (Thorndike, 1924) and did much to undermine the humanists views on the importance of mental discipline. Thorndike's work (as cited in Lagemann, 2002) in the area of animal behavior led to the development of his "law of effect" (p.58); animals learn by pleasurable or painful responses to stimuli. Thorndike
gave the name "connectionism" to his ideas about stimulus and response and emphasized the ability to strengthen them through practice (Kilpatrick, 1992). This work in stimulus-response theory moved the field of behavioral psychology toward a more behaviorist stance. The work of Thorndike and other behaviorists was dominant in American psychology for much of the early 1900s until the emergence of the neo-behaviorists Tolman and Hill and later the more radical Skinner (Kilpatrick, 1992, pp.9-10). Mathematics instructional materials and practices became dominated by drill work focused on discrete bits of content in an attempt to more frequently connect the correct answer to the stimulus. This approach was deeply impacted by the publication of two books by Thorndike, The Psychology of Arithmetic (1922) and The Psychology of Algebra (Thorndike et al., 1923), along with his series of arithmetic textbooks emphasizing drill and practice in the learning of mathematics (as cited in Kilpatrick, p. 10). This focus on students' attainment of procedural fluency in mathematics found in traditional programs may be linked to the theoretical foundations of behaviorism and its influence on curriculum and materials development.

While behaviorist ideas were predominant during the early years of the 20th century, contrasting views were evident even at the time of Thorndike's work. Judd (as cited in Kilpatrick, 1992) was particularly effective in providing research advocating an alternative view. His work attempted to demonstrate that transfer takes place through generalization, particularly in higher-order thinking. Judd, and later Dewey, sought to show that children's mathematical concepts develop from their experiences. It was a debate that continued for years and in some ways served as the precursor to the math wars.
Building on the ideas of Thorndike and Skinner (1938), Gagne (1965) helped move mathematics curriculum and instruction toward a focus on the analysis of a task or concept's structure prior to learning. By identifying the knowledge necessary to successfully complete a task, he demonstrated the importance of pre-requisite skill development. Gagne's (1965) "Nine Events of Instruction"—gain attention, inform learners of the objective, stimulate recall of prior learning, present the contents, provide learner guidance, elicit performance, provide feedback, assess performance, enhance retention and transfer—provided an influential framework for instructional design. This view of learning and mathematics instruction emphasized an analysis of mathematical structure and rigor evident in traditional classroom instruction and mathematics textbook programs, while reform-based instruction moved the focus toward the use of problem solving to help students develop meaning (Hiebert & Carpenter, 1992, p. 81).

The Standards documents (NCTM, 1989, 1991, 1995) reflected the growing knowledge gained from research into how students learn mathematics. Underpinning the documents was the increasing research base that students assimilate new information and experiences into their current conceptual understandings and construct their own new meanings related to the topic of instruction (Carpenter et al., 1989; Cobb & Steffe, 1983; Davis, 1992; Kamii & DeClark, 1985; Maher, Davis, & Alston, 1992a, Maher, Davis, & Alston, 1992b). This "constructivist" view of student learning meant that teachers would move from the role of transmitters of knowledge to facilitators of students' development of mathematical concepts (Carpenter, Fennema, & Franke, 1996; Cobb, Wood, Yackel, & McNeal, 1992; Ferrini-Mundy & Johnson, 1994; Maher, Davis, & Alston, 1992a; Maher, Davis, & Alston, 1992b).
Inherent in this shift in thinking about the teacher’s role in the classroom was an expectation that teachers had the necessary mathematical knowledge and could be taught new ways of interacting with students. Ma’s (1999) comparison of U.S. and Chinese teachers’ knowledge and understanding of mathematics, “…the knowledge gap between the U.S. and Chinese teachers parallels the learning gap between U.S. and Chinese students revealed by other scholars” (p. 144), reinforced the serious concerns about U.S. teachers identified by other researchers (Ball, 2003; Stevenson et al. 1990; Stevenson & Stigler, 1992). Ma’s research indicated that “…the key period during which Chinese teachers develop a teacher’s subject matter knowledge of school mathematics is when they teach it…” (Ma, 1999, p. 147).

Ball and Cohen (1996) identified five intersecting domains across which teachers enact curriculum with their students—what they think about students, teachers’ understanding of the material, teachers’ use of instructional materials, classroom group processes centered on the intellectual and social environment of the classroom, and teachers’ ideas about the larger community’s view of good teaching (p. 7). “Teachers guides could also support teachers’ learning of content…authors could discuss alterative representations of the ideas and connections among them” (p. 7). A number of these domains were explicit in the development of the NSF funded programs. As an example, the teacher materials for “How Many People? How Many Teams?” the grade 5 edition of Investigations in Number, Data, and Space discusses several division strategies in the teacher materials. These include: “Using groups of the divisor”; “Breaking the divided into parts”; “Making an equivalent problem”; and “Solving an easier related problem and then compensating” (TERC, 2008, p.125). Similarly the student and teacher materials for
4th grade *Everyday Mathematics* include examples and discussion of multiple strategies for performing multiplication. Included are partial-sums addition (Bell et al., 2004, p. 284), the partial-products algorithm (pp. 303-306), and lattice multiplication (pp. 315-320). In addition, both sets of teacher materials include discussion of possible student responses and ways in which teachers might handle the discussion of students' ideas.

**Research on Reform Mathematics vs. Traditional**

When the Mathematics and Science Expert Panel published *Exemplary and Promising Mathematics Programs* (Mathematics and Science Expert Panel, 1999), it included a list of 10 school mathematics programs deemed to be *Exemplary* or *Promising*. Programs were identified following a review by a national panel that included teachers, math program evaluators, mathematicians, and experts in the field of statistics. Publication of the list fanned the fires of the *math wars*. Nearly 200 research mathematicians signed a public letter to U. S. Secretary of Education, Richard Riley, requesting the list be withdrawn (Senk & Thompson, 2003, p. 17). Among the programs designated as *Promising* was *Everyday Mathematics* (Bell et al., 1988-1996).

In order to move beyond the emotional discourse surrounding school mathematics programs, it is important to examine the research related to student achievement in both traditional and reform-based mathematics programs. The available research should serve as a guide in the development and refinement of mathematics instructional programs, and it should lead to the next steps in answering questions about instructional programs (Hiebert, 1999; Middleton, et al 2004, p. 77.) While research serves as a guide, the relationship between research and the NCTM *Standards* (1989) is a multi-dimensional
issue, often without definitive answers. However, research can serve to document the current status, indicating those ideas that are most effective, as well as those least likely to be successful (Hiebert, 1999).

The research on traditional math programs and instruction has shown a recognizable approach in most mathematics classrooms with consistent routines among grade levels and an emphasis on practice of procedures for much of the last century (Fey, 1979; Hoetker & Ahlbrand, 1969; Stigler & Hiebert, 1997). The data from TIMSS (1995, 1999) demonstrated that the U.S. mathematics curriculum was repetitive and unfocused (Schmidt, McKnight, & Raizen, 1996). While few formal research studies had been conducted on the efficacy of traditional mathematics programs at the time the NCTM developed the Standards documents, “presuming that traditional approaches have proven to be successful is ignoring the largest database we have. The evidence indicates that the traditional curriculum and instructional methods in the United States are not serving our students well” (Hiebert, 1999, p.13).

As the emphasis on research studies to support the selection and use of school instructional materials has increased, more recent studies have examined the impact of traditional mathematics programs. Several studies considered the Scott Foresman-Addison Wesley Elementary Mathematics (Pearson Scott Foresman, 2004) program. Resendez and Manley (2005b) studied 388 second and 331 fourth graders in six schools (two urban, one rural, three suburban) with 35 teachers randomly assigned to the Scott Foresman program or another district program. Students were assessed using the TerraNova Basic Multiple Assessment for the appropriate grade level. The TerraNova CTBS was selected because it is a reliable and valid standardized test using multiple
choice, constructed response, and computation problems reflective of the NCTM Standards. Students in the intervention program demonstrated no statistically significant effects on the TerraNova Computation (Effect Size/ES = +0.05) score or the TerraNova Total Math score (ES = -0.04) as a result of their use of the Scott Foresman program. The overall effect size was reported as ES = +0.04.

A later study of the 2005 Scott Foresman-Addison Wesley Mathematics (Pearson Scott Foresman, 2005) program by Resendez and Azin (2006) examined the performance of 901 third and fifth graders in four elementary schools on the TerraNova Basic Multiple Assessment following the 2005-06 school year. Utilizing a randomized control trial with 20 teachers in the Scott Foresman and 19 in the control group, researchers found no significant differences between students in the Scott Foresman group and those in the control group (ES = -0.01 effect size). However, some low SES students in the Scott Foresman group demonstrated significantly accelerated growth over the control group.

Studies have examined the impact of Houghton Mifflin Mathematics, (Houghton Mifflin School Division, 2002) a traditional K-6 elementary mathematics program. Johnson, Yanyo and Hall (2002) studied 297 schools in 16 California districts matched for prior mathematics achievement, SES, ethnic diversity, and wealth indicators. Districts using Houghton Mifflin Mathematics were reported to have statistically significant greater gains in National Percentile Rank on the Stanford 9 test at all grade levels than the control districts using other mathematics programs. Overall effect sizes were reported as ES = +0.14 for Houghton Mifflin Mathematics. Similar results were reported for female, LEP and low SES students at all grade levels (Johnson, et al., 2002).
Johnson and Hall (2003) reported on 160 schools in eight California districts following their first year using the *Houghton Mifflin Mathematics* program and compared student performance on the California Stanford 9 test to 137 comparison schools in eight different districts. The schools were matched for district size, prior performance on the California Stanford 9 test, and student demographic characteristics. Johnson and Hall (2003) found significant, positive effects for *Houghton Mifflin Mathematics* for the intervention schools.

A study by EDSTAR, Inc. (2004) paired 519 schools from 32 districts in seven different states, including New Jersey, with 308 intervention schools and 211 comparison schools. Comparison schools used programs other than *Houghton Mifflin Mathematics* and included reform, traditional, and combination programs. Districts were matched for prior mathematics achievement, student demographic characteristics, district size, and average school size. Results indicated, “...the Houghton Mifflin districts had significantly greater percentages of students scoring at or above grade level in the 2002-2003 school year for many subgroups of students” (EDSTAR, 2004, p. 11). Effect sizes of 1.61 and 2.865 are reported for *Houghton Mifflin Mathematics* in comparison to districts using reform-based mathematics programs (p. 38) and district using “traditional math” programs (p. 39).

A recently released study (Agodini et al., 2009) compared the effects of four mathematics curricula on math achievement of first grade students in 39 disadvantaged schools. Using an experimental design, schools in each participating district were randomly assigned to one of four curricula: *Investigations in Number, Data, and Space* (Russell, S. J., Economopoulos, K., Murray, M., Mokros, J., & Goodrow, A., 2006);
Math Expressions (Fuson, 2006); Saxon Math (Larson, 2004); Scott Foresman-Addison Wesley Mathematics (Charles et al., 2005). The 39 schools were in three geographically dispersed regions of the country, including two urban, one suburban, and one rural district. This study of 1,309 first grade students and 131 first grade teachers did not use a control group but did measure student mathematics achievement at the baseline (fall) and upon completion of the programs (spring). Using hierarchical linear modeling techniques, the study found positive effects (ES = 0.30) for Math Expression and Saxon Math in comparison to Investigations and positive effects (ES = 0.24) for Math Expression and Saxon Math in comparison to Scott Foresman-Addison Wesley Mathematics. Investigations in Number, Data, and Space is considered to be a Standards-based curriculum.

In a matched post hoc study Resendez and Azin (2005) studied 340 Georgia public schools matched on SES and ethnicity. Use of the Saxon Math program produced no statistically significant different performance on the Georgia Criterion Referenced Competency Test in comparison to students using other textbooks, with an overall effect size of +0.06.

A significant body of research exists about reform teaching methods and reform-based mathematics programs. Following the NSF’s efforts in the 1990s to change the manner in which math and science instruction was conducted, the Rand Corporation was awarded a grant by the NSF to determine the relationships among reform-based instructional programs and practices and student achievement. The initial study sought to identify the elements of reform-based instruction that positively influenced student achievement and found “a generally weak but positive relationship between the
frequency with which a teacher used the reform practices and student achievement” (Klein et al., 2000, p. xiv). A follow-up study designed to extend the prior research was also conducted by Rand looking at instruction that engaged students as active participants in their learning. While the second study was longitudinal, included a more diverse population, and used multiple measures of student achievement, the results were similar. The use of reform-oriented instructional methods had non-significant or weak positive relationships to student achievement (Le et al., 2006).

The traditional approach to mathematics instruction is to teach a procedure and then assign practice problems as a follow up. Elements of reform mathematics programs are based on the theory that active involvement in problem solving will allow students to develop important mathematics skills (Carpenter et al., 1989; Cobb et al., 1991; Hiebert & Wearne, 1993; Hiebert & Wearne, 1996; Kamii & DeClark, 1985; Mack, 1990; Wearne & Hiebert, 1989), as evidence demonstrates that once students memorize procedures, it becomes difficult for them to understand the conceptual underpinnings at a later date (Brownell & Chazal, 1935; Wearne & Hiebert, 1988).

Reform-based mathematics programs emphasize the priorities and goals of the Standards, attempting to match the research on teaching and learning. The Curriculum and Evaluation Standards set five goals for students: “(1) that they learn to value mathematics, (2) that they become confident in their ability to do mathematics, (3) that they become mathematical problem solvers, (4) that they learn to communicate mathematically, and (5) that they learn to reason mathematically” (NCTM, 1989, p. 5). As an example, Hiebert and Wearne (1993) found that when instruction encouraged students to develop their own procedures and make sense of those presented by
classmates, it appeared to facilitate higher levels of understanding and connections between concepts and skills (p. 420). Similarly, Hiebert (1999) found a connection between students being given the opportunity to construct mathematical understandings and their resulting conceptual understanding and procedural skill. "When programs are implemented with fidelity for reasonable lengths of time, students have learned more and learned more deeply than in traditional programs" (p. 14).

Studies of reform-based mathematics programs have most frequently focused on those developed with NSF funding. While the results of these studies have generally demonstrated positive effects for reform-based mathematics programs, the studies have not been without problems due to the nature of educational research. The use of experimental designs with random assignment of students to treatment or control groups presents a number of ethical issues. In addition, conducting experiments in schools provides for significant challenges in attempting to control for the many variables present (Cline & Mandinach, 1999). More common has been the use of case studies and quasi-experimental studies.

Much of the research on reform-based mathematics programs for middle school students has shown positive relationships to student achievement, but there are studies questioning the effectiveness of particular programs. Alsup and Springer (2003) compared the Houghton-Mifflin mathematics program for eighth grade with the reform-based Cord Applied Math (Cord Communications, 2004) program and found no significant differences. However, this study involved only one teacher, who was also one of the researchers, and there were no pre-test results to equate the student groups.
Cain (2002) studied Louisiana schools using the *Connected Mathematics Project* (Fey et al., 1996) middle school curriculum materials and found that the *Connected Mathematics* schools significantly outperformed other schools on the Iowa Test of Basic Skills (ITBS) and the state assessment program. The researcher served as the lead teacher for the project, demonstrating a lack of rigorous research methodology. A quasi-experimental study conducted by Riordan and Noyce (2001) found that Massachusetts middle school students using *Connected Mathematics* outperformed students using traditional mathematics programs on the state-wide tests (Fey et al., 1996). Generally, students with more time using the program had greater advantage on the test. Effect sizes were not reported.

Studies of *Standards*-based elementary mathematics materials have focused on the three NSF-funded projects that resulted in *Investigations in Number, Data and Space* (TERC), *Math Trailblazers* (University of Illinois at Chicago), and *Everyday Mathematics* (Bell et al., 1988-1996). Carter et al. (2003) reported on the results of third grade Illinois students using *Math Trailblazers* for 2 years prior to taking the Illinois Goals Assessment Program (IGAP) and ITBS assessments. In schools with either inner-city, low-income families or middle class, suburban families, children performed equal to or better than students in the schools prior to implementation of *Math Trailblazers*. Carter et al. (2003) also cited several case studies indicating improvement in student performance following use of *Math Trailblazers*. The research design of these studies did not provide for comparative groups. Similarly, other studies have reported positive results for *Math Trailblazers* but have suffered from methodological flaws (Sconiers,
Isaacs, Higgins, McBride, & Kelso, 2003) such as a failure to establish the comparability of the groups under study.

Mokros' (2003) reported results of several studies involving the use of *Investigations*. Students learned mastery of basic facts and mastery of operations as well as students using traditional programs, and *Investigations* students performed better on word problems, complex calculations in word problems, and problems requiring an explanation of how an operation worked. They also demonstrated a better understanding of place value (Flowers, 1998; Goodrow, 1998). Other studies have reported positive results for *Investigations* but have suffered from methodological flaws (Sconiers, Isaacs, Higgins, McBride, & Kelso, 2003) such as a failure to establish the comparability of the groups under study.

The *Everyday Mathematics* (Bell et al., 1988-1996) program is based on eight underlying principles and their relationship to the research on mathematics education (Carroll & Isaacs, 2003): (a) Children begin school with considerable mathematics knowledge; (b) the elementary school mathematics curriculum should include topics in geometry, algebra, data, and statistics beginning in kindergarten; (c) the use of manipulative materials supports children's thinking and problem-solving skills; (d) students' flexible number sense should be developed along with paper and pencil calculation skills; (e) children "actively construct their knowledge" and should be supported via a balanced curriculum (Carroll & Isaacs, 2003, p. 81); (f) mathematics questions and observations should be considered throughout the curriculum, not just during mathematics lessons; (g) assessment of student progress should be an ongoing
process; and (h) the curriculum should be manageable for teachers given their many curricular responsibilities.

A significant number of studies involving *Everyday Mathematics* have been conducted to judge its effectiveness, including those by the program developers, Northwestern University under NFS funding, and individual schools, districts, and university researchers. Much of the original research on *Everyday Mathematics* was conducted as assessments during classroom field tests of the initial versions of the materials during the three-year development cycle (ARC Center, 2000a; ARC Center, 2000b; ARC Center, 2000c; Carroll, 1993; Carroll, 1995; Carroll, 1996b; Carroll & Porter, 1994; Everyday Learning Corporation, 1996a; Everyday Learning Corporation, 1996b; Everyday Learning Corporation, 1996c; Everyday Learning Corporation, 1996d; Everyday Learning Corporation, 1996e; Hedges, Stodolsky, & Mathison, 1987; Hedges, Stodolsky, & Mathison, 1988;). While the research indicated positive student performance results using *Everyday Mathematics*, some of the studies did not establish comparability of the groups or the materials evaluated during the field test differed from the final published version.

Carroll (1997) reported student results on the 1993 and 1994 Illinois Goal Assessment Program (IGAP) which was reflective of the NCTM *Standards* documents (NCTM, 1989; NCTM, 1991, NCTM, 1995) and had test items reflecting the six mathematical content strands. In the 26 schools using *Everyday Mathematics*, 25 had mean scores significantly above the Illinois state mean and none scored below the state mean. More than half the students using *Everyday Mathematics* since kindergarten met or exceeded the state math goals, which was more than two times the percentage of the
state-wide number. While this quasi-experimental study involved more than 1800 students, the study suffers from the fact that it did not demonstrate the comparability of the comparison groups. Other studies by Carroll, one of the developers of the First Edition of *Everyday Mathematics*, demonstrate the positive impact of the program but have methodological flaws (Carroll, 1996a).

More recent studies continue to add to the research supporting the efficacy of *Everyday Mathematics*. A quasi-experimental study of fourth grade student performance on the Massachusetts state-wide testing examined the performance of students using *Everyday Mathematics* or traditional programs (Riordan & Noyce, 2001). In this study of 3,781 *Everyday Mathematics* students in 67 schools and 5,102 students using 15 different mathematics textbooks in 78 schools, students using the Standards-based program outperformed the traditional students in all types of questions. Overall effect size for 2-3 years of *Everyday Math* was $ES = +0.15$ while more than 4 years was $ES = +0.34$, accounting for an overall effect size of $+0.25$. The "results attest to the effect of these curriculum programs as actually implemented under ordinary prevailing conditions in unselected schools, without regard to whether the programs were implemented optimally" (Riordan & Noyce, 2001, p. 383). In a study measuring longitudinal effects, Carroll (1998a, 1998b) compared the geometric knowledge of fifth and sixth graders of using *Everyday Mathematics* to students using traditional programs and found that *Everyday Mathematics* students "substantially outperformed their counterparts, and nearly all the differences were significant" (Carroll, 1998a p. 188).

Fuson, Carroll, and Drueck (2000) studied second grade students using *Everyday Mathematics* and compared their performance on selected assessments to that of Asian
students and United States students using traditional programs over a 5-year period. The heterogeneous groups of second graders scored higher than middle to upper middle class U.S. students in traditional programs on items measuring number sense, while scoring as well on other items. The *Everyday Mathematics* students were equivalent to middle-class Japanese students. The ARC Center study (Sconiers, Isaacs, Higgins, McBride, & Kelso, 2003) of more than 100,000 students in three states found the average math performance of students using reform programs significantly higher than the average scores of students in matched schools with effect sizes ranging from 0.02 to 0.14 (see Table 7) with an overall effect size of 0.10.

Two studies examined the performance of low-achieving students and learning disabled students on using *Everyday Mathematics*. On the ITBS and the Informal Mathematics Assessment test of problem-solving skills, *Everyday Mathematics* was found to be effective when used with average- and high-ability low achieving students (Baxter, Woodward, & Olson, 2001). Woodward and Baxter (1997), in a study of 104 learning disabled, third grade students using *Everyday Mathematics* and 101 learning disabled students in the comparison group, found that *Everyday Mathematics* students' success on the ITBS challenged the published criticism of Standards-based materials use with special education populations. Sood and Jitendra (2007) compared number sense instruction in three traditional programs and *Everyday Mathematics*. Results indicated that traditional programs had more direct opportunities for number relationship tasks. *Everyday Mathematics* provided greater emphasis on real-world connections and the more complex task of developing relational understanding.
Waite (2000) examined the performance of students using *Everyday Mathematics* in 6 schools with those from 12 schools using a traditional program. With schools matched by ethnicity, SES, and ITBS scores, "almost all comparisons showed that the experimental group taught with the *Everyday Mathematics* curriculum had higher scores on the 1999 Texas Assessment of Academic Skills mathematics test" than students taught with the district curriculum (Waite, 2000, abstract). The overall effect size for *Everyday Mathematics* was reported as +0.26.

**Achievement of Diverse Groups of Students**

A number of researchers examined the academic achievement of various groups defined by gender, socioeconomic status, linguistic diversity/proficiency, special education, and race or ethnicity. Many of these studies found distinct differences in student performance. Lockheed, Thorpe, Brooks-Gunn, Casserly, and McAloon (1985), in a comprehensive review of the research found greater differences based on race and ethnicity than on gender in middle school mathematics achievement. Secada’s (1992) review of research found the achievement gap between minority and White students to be closing for African-American students and limited only to mastery of basic skills test items. Secada identified SES as a consistent determinant of mathematics achievement; "Regardless of school SES, low-SES students achieved less than middle-class students; high SES-students scored best of all (Secada, 1992, p. 633). In examining English language proficiency and mathematics achievement, he found a more complex relationship with "...much variance to be explained" (Secada, 1992, p. 638).
An examination of the NAEP results between 1973 and 1992 indicates improvement for White, African-American, and Hispanic students, with significant disparities in the level of improvement on basic skills. African-American students demonstrated the largest gains for students at ages 9, 13, and 17, with Hispanic students also registering gains equal to or better than White students. However, while Hispanic and African-American made greater gains, White students outscored both groups at all three ages by 20 or more points (Tate, 1997). The 1992 NAEP tests saw the inclusion of more problem-solving skills as reflected in the NCTM Standards documents with questions requiring written extended-responses. While White students demonstrated stronger performance on these grade 4, 8, and 12 questions, all groups demonstrated low levels of proficiency (Tate, p. 657). Results of the National Educational Longitudinal Study of 1988 (NELS: 88), as reported by Green, Dugoni, Ingels, and Camburn (1995), found White and Asian students demonstrating advanced proficiency at two to three times the rate of African-American and Hispanic students. Nearly half of African-American and Hispanic students performed at the lowest levels.

When student performance was examined from 1973 to 2004, many of the same disparities continued to exist. Student performance on the NAEP for 13- and 17-year-olds improved for all ethnic groups, and the scores of African-American and Hispanic 13-year-olds was higher in 2004 than on any prior assessment (Fox, Connolly, & Snyder, 2005, p. 39). While White students consistently outscored Hispanic and African-American students, the performance gap narrowed between 1973 and 2004 (Fox et al., 2005, p. 39).
Some of the disparities in mathematics achievement at the high school level may be attributed to students’ exposure to course content. Hoffer, Rasinski, and Moore (1995) examined the relationship between the number of mathematics courses completed and achievement gains between grades 8 and 12. Results indicated that course taking tended to diminish differences in achievement gains, but Asian and White students continued to outscore Hispanic, African-American and Native American students on Grade-12 achievement tests.

Much of the research on mathematics achievement and its relationship to socio-economic status classifies students along an economic continuum based on parents’ education or the characteristics of the students’ community of residence due to the connection to family income. Rasinski, Ingels, Rock, and Pollack (1993) found improved performance over a 10-year period among all SES groups on NELS:88 and NAEP mathematics assessment data, however, higher SES groups consistently out performed lower SES groups. Hoffer, Rasinski, and Moore (1995) found that differences in Grades 9-12 mathematics achievement to be more a factor of mathematics course taking than differences in SES. Green et al. (1995) found math achievement levels varying by SES, with significantly greater percentages of high-SES students performing at the two highest proficiency levels. In addition, differences in performance based on ethnicity were evident even within SES groups, with significant differences between Whites’ and African-Americans’ test results (Green et al., 1995).

Leder’s (1992) review of the research on gender and mathematics learning found “few consistent differences in performance in mathematics...reported at the early primary school level” (p.607). Substantial evidence exists, however, that males often outperform
females during the secondary-school years. Determining factors in whether such differences are found include content and format of the assessments, age level of the students, and the type of assessment: classroom grades versus standardized tests (Leder, 1992, p. 607). While Hyde, Fennema, and Lamon (1990) and Leder (1986) have identified many of the issues associated with making generalizations about performance by gender, Benbow (1988) stated, “The ratios of high scoring boys to girls has remained relatively constant over the 15 years. Thus, sex differences in SAT-M scores among young adolescents are not temporary trends” (p. 172). In a review of the 30-year performance trends on NAEP testing, Fox, Connolly, and Snyder (2005) found that male 13- and 17-year-olds consistently outperformed females, although often by a narrow margin.

With the requirements of NCLB legislation, new sources of data involving large numbers of students have become available to researchers. Recently, Hyde, Lindberg, Linn, Ellis, and Williams (2008) conducted an analysis of data from 10 states, including New Jersey, with adequate statistical information to allow generalization to all 50 states. (Mean 2008 NAEP mathematics results for 8th-graders was 280.17 for all 50 states and 280.22 for the 10-state sample.) Their examination of the 10-state data set indicated no significant differences in performance between males and females. “In contrast to earlier findings, these very current data provide no evidence of a gender difference favoring males emerging in the high school years…” (Hyde et al., 2008, p. 494). In order to test the hypothesis of greater variability among male intellectual abilities with more males among top-scoring students, they analyzed the variance among test scores for males and
females. While slightly greater variance was found among male students, "...the causes remain unexplained" and inconsistent (Hyde et al., 2008, p.495).

Summary and Synthesis

Concerns about the mathematical capabilities of U.S. students have been longstanding and date to at least the early 1990s. The level of these concerns has ebbed and flowed over the years, often based on events apart from most classrooms. The 1957 launch of Sputnik by the Soviet Union sparked interest in K-12 mathematics and science education resulting in a wave of mathematics reform known as the *new math*. This more conceptual approach to school mathematics received wide-spread criticism from parents and teachers and a resulting back-to-basics movement resulted. Release of NAEP results in the early 1970s led the NCSM to advocate a broader definition of basic skills and NCTM to issue *An Agenda for Action* (NCTM, 1980). Results of cross cultural comparisons of international achievement (Husen et al., 1967; McKnight et al., 1987; Medrich & Griffith, 1992; Robitaille & Garden, 1989) ranked U.S. students below average and caused wide-spread concerns about our international standing. The publication of *A Nation at Risk* (NCEE, 1983) served to galvanize calls for education reform in spite of information to the contrary in *The Sandia Report*. NCTM was the first professional organization to respond with the publication of a series of *Standards* documents outlining the substance and direction for K-12 mathematics education.

Charting a new course for mathematics education, the *Standards* documents led to continuing disagreements about the manner in which K-12 mathematics education should be conducted. These disagreements have often taken place in affluent, suburban
communities where educated, successful parents have distinct ideas about how their children's schooling should be carried out. While these disagreements may be traced to matters of power, process, change, or learning theory, they have resulted in significant conflicts, even earning the term *math wars* to describe them. Only relatively recently has research begun to emerge that provides guidance for educators regarding the direction for mathematics education. With the advent of NCLB, requirements for increased testing, and focus on various subgroups, research on the efficacy of various mathematics programs has begun to allow more informed decision making.

However, in spite of the increasing research, decision making about best mathematics instructional programs and practices is influenced by local, state, and national politics and continues to be a challenge. Parents' views of schooling and learning are deeply embedded in their own experiences. The proper role for instructional materials varies among students, parents, teachers, administrators, and researchers. The push for increased achievement for all groups of students remains a daily concern for educational leaders. Despite improving mathematics performance for all subgroups on NAEP mathematics testing, the achievement gap among subgroups remains significant (Slavin & Lake, 2008), and educators continue to seek guidance from research about the best programs to address students' educational needs.

The research on elementary mathematics programs, while growing, still provides little firm direction for decision makers. Many of the studies suffer from methodology issues inherent in educational research. In a synthesis of effective elementary mathematics programs that included changes in curriculum, supplements to the curriculum including computer assisted instruction, and changes in classroom practices,
Slavin and Lake (2008) concluded, “More research is needed on all of these programs, but the evidence to date suggests a surprising conclusion that despite all the heated debates about the content of mathematics, there is limited high-quality evidence supporting differential effects of different math curricula” (p. 445).
Chapter III

RESEARCH DESIGN AND METHODOLOGY

Introduction

This research explored the influence of a reform-based elementary mathematics program and traditional mathematics programs on General Education student achievement on the NJASK3, NJASK4 and NJASK5. It examined whether the use of the reform-based mathematics program, *Everyday Mathematics*, or a traditional program influences mathematics achievement for General Education students in high-socioeconomic (i.e. DFG-I and DFG-J) districts. This investigation compared the mean Mathematics scores for several sample populations on third, fourth, and fifth grade New Jersey state tests.

For this study, mathematics programs were categorized as reform-based or traditional mathematics programs. Traditional programs included those published by Harcourt Brace, Scott Foresman, Silver-Burdett-Ginn, Addison Wesley, Houghton Mifflin and McGraw Hill. Third, fourth, and fifth grade General Education students in suburban northern New Jersey public school districts with similar SES, ethnicity, special education classification rates, and linguistic diversity comprised the sample populations. Mathematics performance referred to General Education student achievement data on the Mathematics portion of the NJASK3, NJASK4 and NJASK5 collected from publicly available sources at the New Jersey Department of Education. Districts using programs and materials other than *Everyday Mathematics* or a traditional elementary mathematics program were excluded from the research.
Research Design

While a number of researchers and organizations (Coalition For Evidence-Based Policy, 2002; Coalition For Evidence-Based Policy, 2003; Raudenbush, 2002) have advocated for greater use of research evidence in the selection of school-based programs, the passage of NCLB (U.S. Department of Education, 2002) and its frequent use of the term “scientifically-based research” moved the topic to the forefront of discussions about school reading and mathematics programs. *Reading First* required the use of programs based on “scientifically-based research” in order for states and districts to qualify for Federal funds (U.S. Department of Education, 2008). Similarly, President Bush asked the National Mathematics Advisory Panel to base its recommendations on the best scientific research (U.S. Department of Education, 2008, p.15).

The similarities and differences between causal-comparative research and correlational research have been delineated by several authors. Mertler and Charles (2005) and Gay et al. (2006) suggest that causal-comparative research provides a stronger link between cause and effect than that of correlational research. Mertler and Charles assert that causal-comparative research “strongly suggests cause and effect” (p. 315) while correlational research examines “the possible existence of causation” (p. 295). Gay et al. take a similar but slightly more nuanced stance with respect to causal-comparative and correlational research. “Causal-comparative studies attempt to identify cause-effect relationship; correlational studies do not...causal-comparative studies involve comparison, whereas correlational studies involve relationship” (p. 218).
Johnson (2001) identifies the differing views of other authors (Fraenkel & Wallen, 2000; Gall, M. D., Borg, W. R., & Gall, J. P., 1996) with regard to distinctions between causal-comparative and correlational research. "However, these authors do not make the claim that causal-comparative research provides superior evidence for establishing cause and effect than does correlational research" (Johnson, p. 4). Johnson contends that, 

"...causal-comparative research is neither better nor worse in establishing evidence of causality than correlational research," when defined as simple studies "with two variables and no controls, and advanced cases as studies where controls are included" (p. 5).

Causal-comparative and correlational research have several similarities. Both are nonexperimental in nature, as they lack manipulation of an independent variable and random assignment of subjects, leading to the possibility that any observed relationship between independent and dependent variables may be spurious (Davis, 1985, as cited in Johnson, 2001). Neither correlational nor causal-comparative research produce definitive research outcomes, but both may indicate relationships among variables, as well as productive directions for future experimental studies.

Differences between correlational and causal-comparative research include the types of variables used; "...according to popular textbooks (e.g. Charles, 1998; Fraenkel & Wallen, 2000; Gall et al., 1996; Gay & Airasian, 2000), causal-comparative studies include at least one categorical variable and correlational studies include quantitative variables" (Johnson, 2001, p. 5). However, in both instances, the independent variable cannot be manipulated or is not manipulated due to ethical or logistical concerns.

In order to move beyond the discussion of the relative advantages and disadvantages of correlational versus causal-comparative research as presented in the
various research methods textbooks (Fraenkel & Wallen, 2000; Gall, Borg, & Gall, 1996; Gay et al., 2006; Mertler & Charles, 2005), Johnson (2001) provided a new classification schema for nonexperimental research using time (retrospective, cross-sectional, longitudinal) and research objective (descriptive, predictive, explanatory) to describe nine types of nonexperimental research, as follows:

- Type 1 - Retrospective, descriptive study
- Type 2 - Cross-sectional, descriptive study
- Type 3 - Longitudinal, descriptive study
- Type 4 - Retrospective, predictive study
- Type 5 - Cross-sectional, predictive study
- Type 6 - Longitudinal, predictive study
- Type 7 - Retrospective, explanatory study
- Type 8 - Cross-sectional, explanatory study
- Type 9 - Longitudinal, explanatory (Johnson, 2001, p. 10)

This schema removes the scaling of variables as categorical or quantitative in determining the evidence of causality currently present in correlational and causal-comparative research. It classifies nonexperimental quantitative research according the primary research objective: descriptive, predictive, or explanatory (Johnson, 2001, p. 8). The type of data collected is considered along a time dimension—cross-sectional (single point in time), longitudinal (more than one point in time), and retrospective (looking backward in time)—during which the data is gathered. This three-by-three matrix creates a schema resulting in nine classes of nonexperimental quantitative research.
The current research is a nonexperimental, quantitative, cross-sectional, explanatory study (Johnson 2001, p. 10). The study examined sample data from various populations at a single point in time. The use of correlational data analysis advances the intent of this study, determining what relationships, if any, exist between the implementation of a reform-based elementary mathematics program and student achievement on NJASK tests. Use of the correlation coefficient squared ($r^2$) enables school leaders to better understand the extent to which a reform-based mathematics program may be related to student achievement on the state-wide NJASK tests. The nature of the data collected required use of a special case of the Pearson $r$, point-biserial correlation coefficient. When the independent variable is dichotomous (reform-based mathematics program/traditional mathematics program) and the dependent variable (NJASK) is measured on a ratio scale, the point-biserial correlation coefficient is used (Hinkle, Wiersma, & Jurs, 2003).

It is not the intent of this research to imply cause-effect relationships. Rather, the intent of this research is to provide educational decision makers with important information for consideration in how school resources (money, time, professional development initiatives) will be allocated. Additionally, the study may provide indications of a relationship warranting study via future experimental research.

Participants

The researcher collected data from New Jersey DFG-I and DFG-J schools regarding the selection and implementation of an elementary school mathematics program for grades kindergarten through 5. Schools implementing a consistent
mathematics program across grade levels K-5 for at least 3 years prior to the 2007 and 2008 state assessments were included in the study. Schools implementing an elementary mathematics program for fewer than 3 years and schools implementing multiple or mixed programs prior to the testing were excluded from the study. The researcher collected publically available NJASK3, NJASK4, and NJASK5 data for 2007 and 2008 from the New Jersey Department of Education website. As student level data was not used for this study, the researcher has requested and received approval to collect data via an exemption from the Institutional Review Board (IRB).

Setting for the Study

The study examined NJASK state test results for third, fourth and fifth graders in New Jersey DFG I and J districts using the reform-based mathematics program Everyday Mathematics or a traditional mathematics program. Districts using other reform programs such as Math Trailblazers (Institute for Mathematics and Science Education, 1999), or Investigations (TERC, 1998), or programs that cannot be classified as a traditional mathematics program were not included in the study. DFG I and J districts are among the most affluent districts in New Jersey based on the criteria established by the State. Districts are ranked from A to J, with J districts having the highest SES. Criteria include percent of adults with no high school diploma, percent of adults with some college education, occupational status, unemployment rate, percent of individuals in poverty, and median family income. Current DFG rankings are based on the 2000 Decennial Census and remain largely consistent over time.
Treatment

*Everyday Mathematics* is a kindergarten through sixth grade mathematics instructional curriculum developed by the University of Chicago School Mathematics Project (USCSMP) under a grant from the National Science Foundation (NSF), including teacher, student, and resource materials. As such, it is reflective of the NCTM *Standards* and emphasizes the priorities expressed in the *Standards* documents which called for a “deceased emphasis” on “performing paper and pencil calculations” along with greater emphasis on “operation sense” and “collection and organization of data,” (NCTM, 1980 as cited in Klein, 2007, p. 22). The *Standards* documents (NCTM, 1989, 1991, 1995) reflected the theory that active involvement in problem solving, incorporation of manipulative materials, opportunities for students to develop their own procedures and actively construct their own knowledge would allow students to develop important mathematics skills and deeper conceptual understanding. Students in the reform-based mathematics group (independent variable) used the *Everyday Mathematics* program for at least 3 years.

Research Methods

*Data Collection*

The researcher collected data from New Jersey DFG-I and DFG-J schools regarding the selection and implementation of an elementary school mathematics program for grades kindergarten through 5. This information was collected from e-mail contact with individuals responsible for district and/or school mathematics curriculum implementation (superintendents, assistant superintendents, curriculum coordinators, school principals), district websites, and phone conversations with school and district
personnel. Schools implementing a consistent mathematics program across grade levels K-5 for at least 3 years prior to the 2007 and 2008 state assessments were included in the study. Schools implementing an elementary mathematics program for fewer than three years and schools implementing multiple or mixed programs prior to the testing were excluded from the study.

The researcher collected historical, publically-released NJASK3, NJASK4, NJASK5 data for 2007 and 2008 from the New Jersey Department of Education website. In instances where schools included in the study had student performance data suppressed due to small student population size, those cohorts of students were excluded for a particular year.

Variables

The independent variable in each correlation is the use of a Standards-based mathematics program, Everyday Mathematics, or a traditional mathematics program for at least 3 years prior to the NJASK testing at that grade level. The dependent variable is the school mean scale score on NJASK3, NJASK4, and NJASK5 mathematics test for 2007 or 2008 for DFG I & J schools. The Pearson $r$ values were analyzed to test the significance ($p \leq .05$) of each value to determine how the relationships identified differed from what would be expected in the general population. Statistical analyses were completed using SPSS 17.0 software programmed to calculate the Pearson $r$ value (point-biserial correlation coefficient) using a two-tailed test.
**Sampling**

General Education students in third, fourth or fifth grade during the 2006-07 and 2007-08 school years from each elementary school selected for this research were studied as sample populations. Results from the NJASK3, NJASK4 and NJASK5 state mathematics tests were used as the measure of mathematics achievement.

The schools included in this research study have been designated schools in DFG-I or DFG-J districts, representing the upper-most range of the New Jersey District Factor Groups. This designation indicates that the school districts are from among the wealthiest communities in the state. The schools are from several counties and geographic regions in the state and have comparable percentages of ethnicity, economically disadvantaged student populations, Limited English Proficient student populations, and special education populations.

The data were collected from the New Jersey State Department of Education’s database of scores for NJASK testing. The researcher solicited and received approval to collect data via an exemption from the Institutional Review Board, as all documents exist in the public domain and only school level data were used. In several instances, the NJASK results of a particular grade level cohort at an individual school were suppressed due to small sample size and the potential identification of individual students. In such instances, results for that cohort of students were not included in the study.

**Instrumentation**

The NJASK3, NJASK4 and NJASK5 tests are designed to measure student achievement in language arts literacy, mathematics, and science (grade 4) on the New
Jersey Core Curriculum Content Standards in each area. The NJASK tests are administered in grades 3-8 and serve to meet the requirements outlined under NCLB. The New Jersey Department of Education provides results disaggregated by General Education, gender, race, ethnicity, LEP, special education classification, and economic status. The disaggregation of data began with the 2001 administration of the Elementary School Proficiency Assessment (ESPA). Data for General Education students were included in the proposed study.

New Jersey developed state-wide tests in the mid-1970s to measure student basic skills achievement. These Minimum Basic Skills (MBS) tests were the precursors of the current state-wide tests and were administered to grades 3, 6, and 9 from 1978-1982. The MBS tests eventually led to the ESPA in 1997, testing fourth graders in language arts, mathematics, and science. In 2003, the NJASK tests replaced the ESPA, following field testing in 2002. The NJASK was expanded to grades 3-8 as required by NCLB legislation (New Jersey Department of Education, 2008b).

Prior to the spring 2008 testing, the various grade level NJASK tests were developed by several different test companies including National Computer Systems (NCS) and Pearson Educational Measurement. In order to gain greater consistency in the testing program, New Jersey issued a contract in 2003 to Educational Testing Service (ETS) working with Pearson Educational Measurement to produce the tests for all grade levels 3-8. NJASK scaled scores range from 100 to 300 and are classified as Partially Proficient (below 200), Proficient (200-249), or Advanced Proficient (250 and above).

The NJASK tests are designed to measure student progress toward attainment of the New Jersey Core Curriculum Content Standards (NJCCS). The NJCCCS were
originally adopted in 1996 in an effort to define what students should know and be able to
do at the end of their K-12 public school education. The Standards seek to articulate the
important knowledge and skills all students should master (New Jersey Department of
Education, 2008a). Standards were originally developed in language arts literacy,
mathematics, science, visual and performing arts, work place readiness, and world
language. Over the past 12 years, the various standards documents have undergone
several revisions and are currently on a 5-year revision cycle. The New Jersey
mathematics standards are reflective of the NCTM Standards documents in their
emphasis and approach to mathematics. The NJASK tests include seven mathematical
strands; Numeration and Number Theory, Whole Number Operations, Fractions and
The test specifications developed by the New Jersey Department of Education indicate
that the questions are largely made up of problem-solving tasks (Tienken & Wilson,
2001). As a result, the NJASK mathematics tests are well aligned with the NCTM
Standards.

Procedures

The list of DFG-I and DFG-J districts was downloaded from the New Jersey
Department of Education web site (New Jersey Department of Education, 2004) and the
districts’ web sites were reviewed for relevant information regarding each district’s
mathematics curriculum and program of instruction. In cases where a district’s web site
was not up-to-date and specific about the elementary mathematics program used in each
school, contact with district or school administrative personnel was attempted via e-mail
or phone call. These contacts were made using the same set of questions (see Appendix A). Follow-up communication was made, as necessary, to determine the program of instruction during the 3 years prior to 2007 and 2008 NJASK testing. Schools using a traditional program or the *Everyday Mathematics* program were included in the study, while schools using mixed programs or other reform-based programs were excluded from the study. The *State Summary* data for 2007 and 2008 NJASK tests (New Jersey Department of Education, 2008d) were downloaded from the New Jersey Department of Education web site in *Excel* format.

The data were redacted, leaving only those DFG-I and DFG-J schools included in the study. Individual correlation coefficients were generated for NJASK3, NJASK 4, and NJASK5 for 2007 and 2008 by DFG using SPSS 17.0 software. Pearson $r$ values were tested for statistical significance using two-tailed tests. Pearson $r$ was calculated using the point-biserial correlation coefficient, and Cohen’s $d$ was calculated, using an online calculator, to determine effect size (http://web.uccs.edu/lbecker/Psy590/escalc3.htm).

**Methods of Data Analysis**

Descriptive statistical analyses utilizing a series of $t$-tests were conducted on the collected data. The $t$-tests compared the sample means of students using the reform-based mathematics program, *Everyday Mathematics*, and those using traditional mathematics programs. Cohen’s $d$ (Cohen, 1998) was used as a calculation of effect size. Cohen’s $d$ is the difference between the means, $M_1 - M_2$, divided by standard deviation, $s$, of either group. The pooled standard deviation is calculated as the root mean square of the two standard deviations (Cohen, 1988, p. 44), $d = M_1 - M_2 / s_{pooled}$. Cohen defined
effect sizes for $d$ as small, $d = 0.2$, medium, $d = 0.5$, and large, $d = 0.8$ and indicated
"there is a certain risk inherent in offering conventional operational definitions for those
terms for use in power analysis in as diverse a field of inquiry as behavioral science"
(Cohen, p.25). Effect sizes are considered to be the estimated differences between the
population means (Witte & Witte, 2004, p. 376). A test of correlation was used to
determine r values using the point biserial correlation coefficient calculated via SPSS.

Discussion of Controls

"Nonexperimental quantitative research is an important area of research for
educators because there are so many important but nonmanipulable independent variables
needing further study in the field of education" (Johnson, 2001. p. 3). Because many
variables do not easily fit the requirements of experimental research, the responsibility
remains for researchers to eliminate possible rival explanations when conducting
nonexperimental quantitative studies. "It is essential that we understand what is always
important when attempting to make causal attributions is the elimination of plausible
rival explanations" (Johnson, 2001, p. 3). While it is not possible to control for all
variables in nonexperimental quantitative research, it is important to attempt to control
for extraneous variables (Gay et al., 2006, p. 222).

"One way to control extraneous variables is to compare groups that are
homogeneous with respect to the extraneous variable" (Gay et al., 2006, p. 222).
Controlling for such variables will reduce the possibility that any observed relationship
between independent and dependent variables may be spurious (Davis, 1985, as cited in
Johnson, 2001). The current research controls for a number of extraneous variables
through the use of DFG-I and J General Education NJASK data. Controls are necessary to account for possible rival explanations of correlation values found to be significant \( (p \leq .05) \).

**Socioeconomic status (SES).** The DFG system was developed for ranking New Jersey school districts by their SES. Introduced by the New Jersey Department of Education in 1975 based on 1970 Census data, DFG is periodically updated taking into account new Census data. The 1984 revision slightly changed the theoretical model for determining SES. The most recent revision took place in 2004 using the 2000 Census. DFG ranks school districts from A to J, with J districts having the highest SES (New Jersey Department of Education, 2004).

The DFG system was based on the research evidence of the relationship between SES and educational outcomes (Green et al., 1995; Rasinski et al., 1993; Secada, 1992) and was developed for reporting test scores. The intent of the DFG system was to reduce the variation in reported scores due to non-school factors. SES is not measured directly by the DFG system, but it is a function of other measures including income, occupation, and education. DFG is a statistical index derived from various socioeconomic indicators. These include: percent of population with no high school diploma, percent of population with some college, occupation, population density, income, unemployment, and poverty. Districts are "grouped so that each group would consist of districts having factor scores within an interval of one tenth of the distance between the highest and lowest scores" (New Jersey Department of Education, 2004, ¶ 5).

**Limited English Proficient (LEP).** In New Jersey, Limited English Proficient students are those whose performance on an approved test of listening, speaking, reading,
and writing of English identifies them as needing additional, specialized English
instruction from an appropriately certificated teacher. New Jersey state regulations
mandate the use of multiple criteria for identifying and exiting students from a language
assistance program.

The policies, regulations, and practices surrounding the collection of data regarding
LEP students vary considerably from state to state, making generalizations about their
status on a national basis difficult (Kindler, 2002). A national survey for educational
status indicators of LEP students for 2000-2001 indicated approximately 9.1% of
secondary level (grades 7-12) students were retained in grade. With 41 states reporting
data on both participation and student success on English reading comprehension tests,
only 18.7% of LEP students assessed scored above the state norm on the state-
administered tests (Kindler, 2002, p.13).

New Jersey LEP students demonstrate significantly lower performance on NJASK
testing. In 2007, only 51.3% of third grade students identified as Current LEP achieved
Proficiency or Advanced Proficiency on the Language Arts Literacy subtest, with 65.1%
reaching proficiency levels in the Mathematics subtest. In contrast, third grade General
Education students’ proficiency levels were 89.1% in Language Arts Literacy and 91.0%
Mathematics. The General Education mean scale score for third grade mathematics was
222.0, while the Current LEP mean scale score was 196.0 for mathematics (New Jersey
Department of Education, 2008c).

Results for 2007 fourth grade NJASK testing were similar. Eighty-seven and four
tenths percent of General Education students achieved proficiency levels on the
Language Arts Literacy subtest, while only 44.6% of Current LEP students reached
proficiency levels. In Mathematics, 89.7% of fourth grade General Education students score at Proficient or Advanced Proficient levels, with only 54.8% of Current LEP students achieving either proficiency level. The mean scale score for fourth grade General Education students was 239.3 for mathematics, while the mean score for Current LEP students was 204.7 (New Jersey Department of Education, 2008c). The lower achievement of LEP students will be controlled for through the use of General Education achievement data, as General Education scores exclude Current LEP students.

Step-by-Step Procedures

The following steps were followed in conducting the study:

*Step One: Identification of schools for inclusion in the study.* The list of DFG-I and J districts was downloaded from the New Jersey Department of Education web site (New Jersey Department of Education, 2004). The web sites of DFG-I and J districts were reviewed for relevant data regarding the district’s math curriculum and program of instruction. In some instances the information from the district web site was up-to-date, clear, and specific about the elementary mathematics program used in each elementary school. In most cases a direct contact with district or school administrative personnel was required. These contacts were made via e-mail to district and/or school level administrators, using the same set of questions (see Appendix A).

When responses to the e-mail questions were unclear, follow-up email or phone contact was made to determine the elementary mathematics program in place for the 3 years prior to NJASK testing in 2007 and 2008. Schools that did not use *Everyday Mathematics* or a traditional program were excluded from the data sample. These
exclusions included other reform-based mathematics programs (Investigations, Math Trailblazers) and schools using mixed programs that included both traditional and reform-based instructional materials.

**Step Two: Collection of the 2007 and 2008 NJASK test data.** Within 8-10 months following the administration of NJASK tests, the New Jersey Department of Education publishes New Jersey Statewide Assessment Reports on the Department web site (New Jersey Department of Education, 2008c, 2008d). In addition to an Executive Summary including highlights of grade level test results, the Department publishes as a State Summary of data in Excel Spreadsheet format. This State Summary includes results from every public school in New Jersey. In some instances, elements of the data are suppressed due to small sample size and the possibility that a person with first-hand knowledge of an individual school could determine the performance of an individual student or students (New Jersey Department of Education, 2005). The New Jersey Department of Education web site was accessed to download the State Summary files for DFG-I and J districts. In the few instances where the NJASK results of a particular grade level cohort at an individual school were suppressed due to small sample size and the potential identification of individual students, results for that cohort of students were not included in the study.

**Step Three: Correlation analysis.** Individual correlation coefficients were generated for NJASK3, NJASK4, and NJASK5 2007 and 2008 tests using SPSS 17.0 software. Pearson $r$ values were tested for significance using SPSS software and two-tailed tests. The nature of the data collected required use of a special case of the Pearson $r$, point-biserial correlation coefficient. When the independent variable is dichotomous (reform-
based mathematics program/traditional mathematics program) and the dependent variable (NJASK) is measured on a ratio scale, the point-biserial correlation coefficient is used (Hinkle, Wiersma, & Jurs, 2003).

*Step Four: Generation of Cohen's d values.* Cohen's $d$ values were calculated to determine the effect size of *Everyday Math* or traditional programs.

*Step Five: Discussion.* The data generated in the study was reviewed and conclusions drawn regarding the influence between implementation of a reform-based elementary mathematics program and performance on the NJASK mathematics assessments. Inferences were drawn and implications discussed.

*Step Six: Results compared.* Results from the proposed study will be compared to previous studies as reported in the Literature Review in Chapter II as a basis for conclusions.

**Summary of Chapter III and Description of Chapter IV**

Chapter III described the proposed research design and the methods for data collection and analysis. The chapter gave a detailed step-by-step process for this study. Chapter IV will provide a statistical analysis of the data collected in this study. The chapter will also summarize findings.
Chapter IV
DATA ANALYSIS

Introduction

The purpose of this study was to investigate the possible influence of a reform-based elementary school mathematics program in comparison to traditional mathematics programs within New Jersey DFG-I and J districts on student achievement as measured by NJASK tests at grades 3, 4, and 5. This chapter provides a presentation of the data obtained during the study, along with analysis of results.

The study used the results of the state-mandated NJASK standardized tests administered at grades 3, 4, and 5 as a tool for this nonexperimental, quantitative, cross-sectional, explanatory study (Johnson, 2001, p. 10). Student achievement outcomes were analyzed and compared for grades 3, 4, and 5 for the 2006-07 and 2007-08 school years. This study determined through analysis of student achievement outcomes whether the implementation of a reform-based elementary mathematics program influences student performance on NJASK3, NJASK4, and NJASK5 mathematics tests. The statistical analysis software SPSS 17.0 was utilized for data analysis. One research question and six subsidiary questions are analyzed and discussed.

Results

The purpose of this chapter is to present the results of the descriptive statistics analyses utilizing a series of t-tests. The t-tests were used to compare the sample means of students using the reform-based mathematics program, *Everyday Mathematics*, and those using traditional mathematics programs. Several additional statistical measures
were employed when t-tests yielded statistically significant results. Cohen’s $d$ (Cohen, 1988) was used as one calculation of effect size. Cohen’s $d$ is the difference between the means, $M_1 - M_2$, divided by standard deviation, $s$, of either group. The pooled standard deviation is calculated as the root mean square of the two standard deviations (Cohen, 1988, p. 44), $d = M_1 - M_2 / s_{\text{pooled}}$. Cohen defined effect sizes for $d$ as small, $d = 0.2$, medium, $d = 0.5$, and large, $d = 0.8$ and indicated “there is a certain risk inherent in offering conventional operational definitions for those terms for use in power analysis in as diverse a field of inquiry as behavioral science” (p.25). Effect sizes are considered to be the estimated differences between the population means (Witte & Witte, 2004, p. 376).

A test of correlation was used to determine $r$ values using the point biserial correlation coefficient calculated via SPSS using the following formula

$$
 r_{pb} = \frac{M_1 - M_0}{s_n} \sqrt{\frac{n_1 n_0}{n^2}},
$$

The calculated $r$ values will be measuring the relationship between independent and dependent variables will be interpreted according to the following Rule of Thumb (Hinkle et al., 2003, p. 109):
Table 1

Rule of Thumb for Interpreting Correlation

<table>
<thead>
<tr>
<th>Size of Correlation</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>+/- .90 to 1.00</td>
<td>Very high correlation</td>
</tr>
<tr>
<td>+/- .70 to .90</td>
<td>High correlation</td>
</tr>
<tr>
<td>+/- .50 to .70</td>
<td>Moderate correlation</td>
</tr>
<tr>
<td>+/- .30 to .50</td>
<td>Low correlation</td>
</tr>
<tr>
<td>+/- .00 to .30</td>
<td>Little if any correlation</td>
</tr>
</tbody>
</table>

The chapter will focus on investigating and answering the primary research question and associated subsidiary questions.

Research Question

Research Question: What is the influence of implementing a reform-based mathematics program on the mathematics achievement, as measured by NJASK tests, of General Education students in grades 3 through 5 who attend school in New Jersey school districts classified as DFG-I and DFG-J?

This research question will be addressed and supported by 6 subsidiary questions. The 6 subsidiary questions are supported by 12 hypotheses, which are investigated through the analysis of t-tests and calculation of Cohen's $d$. In those instances where statistically significant differences occur between the means, further analysis will be provided via correlational tests including calculation of Pearson $r$ via SPSS 17.0 software and Cohen's $d$ correlation coefficient using the calculator at http://web.uccs.edu/lbecker/Psy590/escalc3.htm.
Subsidiary Question 1

What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-I General Education students on NJASK3 mathematics tests?

This question is supported by Hypotheses 1 and 2, which are investigated and analyzed using t-tests displayed in Tables 2-7.

Hypothesis 1. There is no significant difference in 2007 NJASK3 mathematics scores between DFG-I General Education students using the Everyday Mathematics program and those using a traditional mathematics program.

Table 2 indicates the 2007 third grade DFG-I mean mathematics scale score was higher for Everyday Mathematics (EM) schools (246.118) than for schools using traditional programs (245.286). While there was a small difference in the mean scores, results of the t-test for two independent samples (see Table 3) indicate the means do not differ significantly at the $p < .05$ level ($p = .371$). Calculation of Cohen's $d$ (0.1349) indicates a small effect size for Everyday Mathematics.

Table 2

<table>
<thead>
<tr>
<th>2007 NJASK3 DFG-I Mean Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>GE Scale Math</td>
</tr>
<tr>
<td>EM</td>
</tr>
<tr>
<td>GE Scale Math</td>
</tr>
<tr>
<td>EM</td>
</tr>
</tbody>
</table>
Table 3

*T-Test Comparing DFG-I Everyday Math and Traditional Program Schools on 2007 NJASK3*

<table>
<thead>
<tr>
<th>Levene's Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Sig.</td>
<td>t</td>
</tr>
<tr>
<td>---</td>
<td>------</td>
<td>---</td>
</tr>
<tr>
<td>GE Scale Math variances assumed</td>
<td>Equal</td>
<td>2.235</td>
</tr>
<tr>
<td>GE Scale Math variances not assumed</td>
<td>Equal</td>
<td>-872</td>
</tr>
</tbody>
</table>

Table 4

*2007 NJASK3 DFG-I Correlations*

<table>
<thead>
<tr>
<th>Program</th>
<th>GE Scale Math</th>
<th>Program</th>
<th>GE Scale Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>1</td>
<td>.068</td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.371</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>177</td>
<td>177</td>
<td></td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>.068</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.371</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>177</td>
<td>177</td>
<td></td>
</tr>
</tbody>
</table>
Subsidiary Question 1

What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-I General Education students on NJASK3 mathematics tests?

Hypothesis 2. There is no significant difference in 2008 NJASK3 mathematics scores between DFG-I General Education students using the Everyday Mathematics program and those using a traditional mathematics program.

Table 5 indicates the 2008 third grade DFG-I mean mathematics scale score was higher for traditional program schools (245.880) than for schools using Everyday Mathematics (244.525). While there was a small difference in the mean scores, results of the t-test for two independent samples (see Table 6) indicate the means do not differ significantly at the $p < .05$ level ($p = .163$). Calculation of Cohen’s $d$ (0.2094) indicates a small effect size for traditional programs.

Table 5

<table>
<thead>
<tr>
<th>Program</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE Scale Math</td>
<td></td>
<td>Traditional</td>
<td>245.880</td>
<td>7.1493</td>
</tr>
<tr>
<td>EM</td>
<td>103</td>
<td>244.525</td>
<td>5.7112</td>
<td>.5627</td>
</tr>
</tbody>
</table>
Table 6

*T-Test Comparing DFG-I Everyday Math and Traditional Program Schools on 2008 NJASK3*

<table>
<thead>
<tr>
<th></th>
<th>Levene's Test for</th>
<th>t-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Equality of Variances</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GE Scale variances</td>
<td>3.391 .067 1.400 175</td>
<td>1.400 .163 1.3545 .9678 -.5556 3.2645</td>
<td></td>
</tr>
<tr>
<td>Math assumed</td>
<td>Equal variances not assumed</td>
<td>1.350 134.976 .179 1.3545 1.0037 -.6305 3.3395</td>
<td></td>
</tr>
</tbody>
</table>

Table 7

*2008 NJASK3 DFG-I Correlations*

<table>
<thead>
<tr>
<th></th>
<th>Program</th>
<th>GE Scale Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program</td>
<td>Pearson Correlation</td>
<td>1</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>177</td>
<td>177</td>
</tr>
<tr>
<td>GE Scale Math</td>
<td>Pearson Correlation</td>
<td>-.105 1</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.163</td>
</tr>
<tr>
<td>N</td>
<td>177</td>
<td>177</td>
</tr>
</tbody>
</table>
Subsidiary Question 2

What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-J General Education students on NJASK3 mathematics tests?

This question is supported by Hypotheses 3 and 4, which are investigated and analyzed using t-tests displayed in Tables 8-13.

Hypothesis 3. There is no significant difference in 2007 NJASK3 mathematics scores between DFG-J General Education students using the Everyday Mathematics program and those using a traditional mathematics program.

Table 8 indicates the 2007 third grade DFG-J mean mathematics scale score was higher for Everyday Mathematics schools (249.268) than for schools using a traditional program (247.200). While there was a difference in the mean scores, results of the t-test for two independent samples (see Table 9) indicate the means do not differ significantly at the $p < .05$ level ($p = .194$). Calculation of Cohen’s $d$ (0.4574) indicates a small to medium effect size for Everyday Mathematics.

Table 8

<table>
<thead>
<tr>
<th>Program</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE Scale Math</td>
<td>14</td>
<td>247.200</td>
<td>5.0666</td>
<td>1.3541</td>
</tr>
<tr>
<td>Traditional</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EM</td>
<td>19</td>
<td>249.268</td>
<td>3.9007</td>
<td>.8949</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 9

*T-Test Comparing DFG-J Everyday Math and Traditional Program Schools on 2007 NJASK3*

<table>
<thead>
<tr>
<th>Levene's Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>Sig.</td>
</tr>
<tr>
<td>GE Scale Equal variances assumed</td>
<td>.066</td>
<td>.758</td>
</tr>
<tr>
<td>GE Scale Equal variances not assumed</td>
<td>-1.274</td>
<td>23.566</td>
</tr>
</tbody>
</table>

Table 10

*2007 NJASK3 DFG-J Correlations*

<table>
<thead>
<tr>
<th>Program</th>
<th>GE Scale Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>GE Scale Math</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>33</td>
</tr>
<tr>
<td>GE Scale Math Pearson Correlation</td>
<td>0.232</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.194</td>
</tr>
<tr>
<td>N</td>
<td>33</td>
</tr>
</tbody>
</table>
**Subsidiary Question 2**

What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-J General Education students on NJASK3 mathematics tests?

*Hypothesis 4.* There is no significant difference in 2008 NJASK3 mathematics scores between DFG-J General Education students using the Everyday Mathematics program and those using a traditional mathematics program.

Table 11 indicates the 2008 third grade DFG-J mean mathematics scale score was higher for traditional program schools (250.850) than for schools using *Everyday Mathematics* (248.058). While there was a small difference in the mean scores, results of the t-test for two independent samples (see Table 12) indicate the means do not differ significantly at the $p < .05$ level ($p = 132.$). Calculation of Cohen’s $d$ (0.5461) indicates a medium effect size for traditional programs.

Table 11

<table>
<thead>
<tr>
<th>2008 NJASK3 DFG-J Mean Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>GE Scale Math Traditional</td>
</tr>
<tr>
<td>EM</td>
</tr>
</tbody>
</table>
Table 12

T-Test Comparing DFG-J Everyday Math and Traditional Program Schools on 2008 NJASK3

<table>
<thead>
<tr>
<th></th>
<th>Levene’s Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>Sig.</td>
<td>t</td>
</tr>
<tr>
<td>GE Equal</td>
<td>.080</td>
<td>.779</td>
<td>1.546</td>
</tr>
<tr>
<td>Scale variances assumed</td>
<td>Equal variances not assumed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.555</td>
<td>28.751</td>
<td>.131</td>
</tr>
</tbody>
</table>

Table 13

2008 NJASK3 DFG-J Correlations

<table>
<thead>
<tr>
<th></th>
<th>Program</th>
<th>GE Scale Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program</td>
<td>Pearson Correlation</td>
<td>1</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.132</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>GE Scale Math</td>
<td>Pearson Correlation</td>
<td>-.268</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.132</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>33</td>
<td>33</td>
</tr>
</tbody>
</table>
Subsidiary Question 3

What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-I General Education students on NJASK4 mathematics tests?

This question is supported by Hypotheses 5 and 6, which are investigated and analyzed using $t$-tests displayed in Tables 14-19.

Hypothesis 5. There is no significant difference in 2007 NJASK4 mathematics scores between DFG-I General Education students using the Everyday Mathematics program and those using a traditional mathematics program.

The findings are noted in Tables 14 and 15. The mean score for DFG-I Everyday Mathematics schools (252.964) for 2007 NJASK4 mathematics was higher than for schools using a traditional program (248.494). Table 15 displays results of the $t$-test for two independent samples indicating observed differences between Everyday Mathematics schools and traditional schools at the $p < .000$ level of significance. Calculation of Pearson $r$ ($+.322$) indicates a low positive correlation between NJASK scores and implementation of Everyday Mathematics. Calculation of Cohen’s $d$ (0.6804) indicates a medium effect size for Everyday Mathematics.

Table 14

<table>
<thead>
<tr>
<th>Program</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE Scale Math</td>
<td>71</td>
<td>248.494</td>
<td>6.8793</td>
<td>.8164</td>
</tr>
<tr>
<td>Traditional</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EM</td>
<td>101</td>
<td>252.964</td>
<td>6.2430</td>
<td>.6212</td>
</tr>
</tbody>
</table>

2007 NJASK4 DFG-I Mean Scores
Table 15

*Table 15*

**T-Test Comparing DFG-I Everyday Math and Traditional Program Schools on 2007 NJASK4**

<table>
<thead>
<tr>
<th></th>
<th>Levene's Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>Sig.</td>
<td>t</td>
</tr>
<tr>
<td>GE Scale Math</td>
<td>Equal</td>
<td>.855</td>
<td>.356</td>
</tr>
<tr>
<td></td>
<td>variances</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>assumed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Equal</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>variances not assumed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 16

*Table 16*

**2007 NJASK4 DFG-I Correlations**

<table>
<thead>
<tr>
<th>Program</th>
<th>GE Scale Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program</td>
<td>Pearson Correlation</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
</tr>
<tr>
<td></td>
<td>Sum of Squares and Cross-products</td>
</tr>
<tr>
<td></td>
<td>Covariance</td>
</tr>
<tr>
<td></td>
<td>N</td>
</tr>
<tr>
<td>GE Scale Math</td>
<td>Pearson Correlation</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
</tr>
<tr>
<td></td>
<td>Sum of Squares and Cross-products</td>
</tr>
<tr>
<td></td>
<td>Covariance</td>
</tr>
<tr>
<td></td>
<td>N</td>
</tr>
</tbody>
</table>

**Correlation**

| **Correlation is significant at the 0.01 level (2-tailed).** |
Subsidiary Question 3

What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-I General Education students on NJASK4 mathematics tests?

Hypothesis 6. There is no significant difference in 2008 NJASK4 mathematics scores between DFG-I General Education students using the Everyday Mathematics program and those using a traditional mathematics program.

The findings are noted in Tables 17 and 18. The mean score for DFG-I Everyday Mathematics schools (253.234) for 2008 NJASK4 mathematics was higher than for schools using a traditional program (249.889). Table 18 displays results of the t-test for two independent samples indicating observed differences between Everyday Mathematics schools and traditional schools at the \( p < .001 \) level of significance. Calculation of Pearson \( r \) (+.257) indicates little correlation between NJASK scores and implementation of Everyday Mathematics. Calculation of Cohen’s \( d \) (0.537) indicates a medium effect size for Everyday Mathematics.

Table 17

<table>
<thead>
<tr>
<th>Program</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE Scale Math</td>
<td>70</td>
<td>249.889</td>
<td>6.2375</td>
<td>.7455</td>
</tr>
<tr>
<td>Traditional</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EM</td>
<td>100</td>
<td>253.234</td>
<td>6.2171</td>
<td>.6217</td>
</tr>
<tr>
<td>Traditional</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 18

<table>
<thead>
<tr>
<th></th>
<th>Levene's Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>Sig.</td>
<td>t</td>
</tr>
<tr>
<td>GE Equal Scale variances assumed</td>
<td>.238</td>
<td>.626</td>
<td>-3.448</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>-3.446</td>
<td>148.338</td>
<td>.001</td>
</tr>
</tbody>
</table>

Table 19

2008 NJASK4 DFG-I Correlations

<table>
<thead>
<tr>
<th></th>
<th>Program</th>
<th>GE Scale Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program</td>
<td>Pearson Correlation</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>170</td>
</tr>
<tr>
<td>GE Scale Math</td>
<td>Pearson Correlation</td>
<td>.257**</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>170</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).
Subsidiary Question 4

What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-J General Education students on NJASK4 mathematics tests?

This question is supported by Hypotheses 7 and 8, which are investigated and analyzed using t-tests displayed in Tables 20-25.

Hypothesis 7. There is no significant difference in 2007 NJASK4 mathematics scores between DFG-J General Education students using the Everyday Mathematics program and those using a traditional mathematics program.

The findings are noted in Tables 20 and 21. The mean score for DFG-J Everyday Mathematics schools (257.405) for 2007 NJASK4 mathematics was higher than for schools using a traditional program (252.025). Table 21 displays results of the t-test for two independent samples indicating observed differences between Everyday Mathematics schools and traditional schools at the \( p < .003 \) level of significance. Calculation of Pearson \( r \) (+.513) indicates a moderate positive correlation between NJASK scores and implementation of Everyday Mathematics. Calculation of Cohen’s \( d \) (1.112) indicates a large effect size for Everyday Mathematics.

Table 20

<table>
<thead>
<tr>
<th>Program</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE Scale Math</td>
<td>12</td>
<td>252.025</td>
<td>5.9419</td>
<td>1.7153</td>
</tr>
<tr>
<td>Traditional</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EM</td>
<td>19</td>
<td>257.405</td>
<td>3.3922</td>
<td>.7782</td>
</tr>
</tbody>
</table>

2007 NJASK4 DFG-J Mean Scores
Table 21

**T-Test Comparing DFG-J Everyday Math and Traditional Program Schools on 2007 NJASK4**

<table>
<thead>
<tr>
<th></th>
<th>Levene’s Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>Sig.</td>
<td>t</td>
</tr>
<tr>
<td>GE Equal Scale variances Math assumed</td>
<td>1.711</td>
<td>.201</td>
<td>-3.220</td>
</tr>
<tr>
<td>GE Equal variances not assumed</td>
<td>-2.856</td>
<td>15.591</td>
<td>.012</td>
</tr>
</tbody>
</table>

Table 22

**2007 NJASK4 DFG-J Correlations**

<table>
<thead>
<tr>
<th></th>
<th>Program Correlation</th>
<th>GE Scale Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program</td>
<td>Pearson Correlation</td>
<td>.513&lt;sup&gt;*&lt;/sup&gt;</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.003</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>GE Scale Math</td>
<td>Pearson Correlation</td>
<td>.513&lt;sup&gt;**&lt;/sup&gt;</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.003</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>31</td>
<td>31</td>
</tr>
</tbody>
</table>

<sup>*</sup>. Correlation is significant at the 0.01 level (2-tailed).
Subsidiary Question 4

What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-J General Education students on NJASK4 mathematics tests?

Hypothesis 8. There is no significant difference in 2008 NJASK4 mathematics scores between DFG-J General Education students using the Everyday Mathematics program and those using a traditional mathematics program.

The findings are noted in Tables 23 and 24. The mean score for DFG-J Everyday Mathematics schools (258.042) for 2008 NJASK4 mathematics was higher than for schools using a traditional program (250.192). Table 24 displays results of the t-test for two independent samples indicating observed differences between Everyday Mathematics schools and traditional schools at the $p < .000$ level of significance. Calculation of Pearson $r$ (.735) indicates a high positive correlation between NJASK scores and implementation of Everyday Mathematics. Calculation of Cohen’s $d$ (2.125) indicates a large effect size for Everyday Mathematics.

Table 23

<table>
<thead>
<tr>
<th>Program</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE Scale Math</td>
<td>12</td>
<td>250.192</td>
<td>3.8609</td>
<td>1.1145</td>
</tr>
<tr>
<td>EM</td>
<td>19</td>
<td>258.042</td>
<td>3.5172</td>
<td>.8069</td>
</tr>
</tbody>
</table>
Table 24

*T-Test Comparing DFG-J Everyday Math and Traditional Program Schools on 2008 NJASK4*

<table>
<thead>
<tr>
<th>GE Equal Scale variances assumed</th>
<th>Levene's Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>Sig.</td>
<td>t</td>
</tr>
<tr>
<td>GE Equal Scale variances assumed</td>
<td>.163</td>
<td>.689</td>
<td>-5.831</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>.163</td>
<td>.689</td>
<td>-5.705</td>
</tr>
</tbody>
</table>

Table 25

*2008 NJASK4 DFG-J Correlations*

<table>
<thead>
<tr>
<th>Program</th>
<th>GE Scale Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>.735*</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
</tr>
<tr>
<td>N</td>
<td>31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GE Scale Math</th>
<th>Pearson Correlation</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>31</td>
<td></td>
</tr>
</tbody>
</table>

*Correlation is significant at the 0.01 level (2-tailed).*
Subsidiary Question 5

What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-I General Education students on NJASK5 mathematics tests?

This question is supported by Hypotheses 9 and 10, which are investigated and analyzed using t-tests displayed in Tables 26-31.

Hypothesis 9. There is no significant difference in 2007 NJASK5 mathematics scores between DFG-I General Education students using the Everyday Mathematics program and those using a traditional mathematics program.

Table 26 indicates the 2007 fifth grade DFG-I mean mathematics scale score was higher for Everyday Mathematics schools (246.527) than for schools using traditional programs (245.752). While there was a difference in the mean scores, results of the t-test for two independent samples (see Table 27) indicate the means do not differ significantly at the $p < .05$ level ($p = .570$). Calculation of Cohen’s $d$ (0.0929) indicates a small effect size for Everyday Mathematics.

Table 26

<table>
<thead>
<tr>
<th>Program</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE Scale Math</td>
<td>66</td>
<td>245.752</td>
<td>8.5303</td>
<td>1.0500</td>
</tr>
<tr>
<td>Traditional</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EM</td>
<td>85</td>
<td>246.527</td>
<td>8.1333</td>
<td>.8822</td>
</tr>
</tbody>
</table>

2007 NJASK5 DFG-I Mean Scores
Table 27

*T-Test Comparing DFG-I Everyday Math and Traditional Program Schools on 2007 NJASK5*

<table>
<thead>
<tr>
<th>Levene's Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>Sig.</td>
</tr>
<tr>
<td>GE Equal Scale variances assumed</td>
<td>1.003</td>
<td>.318</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 28

*2007 NJASK5 DFG-I Correlations*

<table>
<thead>
<tr>
<th>Program</th>
<th>GE Scale Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>151</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GE Scale Math</th>
<th>Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td></td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>151</td>
</tr>
</tbody>
</table>
Subsidiary Question 5

What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-I General Education students on NJASK5 mathematics tests?

Hypothesis 10. There is no significant difference in 2008 NJASK5 mathematics scores between DFG-I General Education students using the Everyday Mathematics program and those using a traditional mathematics program.

Table 29 indicates the 2008 fifth grade DFG-I mean mathematics scale score was higher for Everyday Mathematics schools (248.071) than for schools using traditional programs (245.079). There was a difference in the mean scores with results of the t-test for two independent samples (see Table 30) indicating the means do differ significantly at the $p = .045$ level. This finding indicates that observed mean differences exist between DFG-I schools using Everyday Mathematics and traditional programs on the 2008 NJASK5. Calculation of Pearson r (+.164) indicates little if any positive correlation between NJASK scores and implementation of Everyday Mathematics. Calculation of Cohen's $d$ (0.334) indicates a small effect size for Everyday Mathematics.

Table 29

<table>
<thead>
<tr>
<th>2008 NJASK5 DFG-I Mean Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>GE Scale Math</td>
</tr>
<tr>
<td>Traditional</td>
</tr>
<tr>
<td>EM</td>
</tr>
</tbody>
</table>
Table 30

*T-Test Comparing DFG-I Everyday Math and Traditional Program Schools on 2008 NJASK5*

<table>
<thead>
<tr>
<th></th>
<th>Levene's Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>Sig.</td>
<td>t</td>
</tr>
<tr>
<td>GE Scale variances</td>
<td>.000</td>
<td>.997</td>
<td>-2.022</td>
</tr>
<tr>
<td>Math assumed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal variances not</td>
<td>-2.041</td>
<td>144.048</td>
<td>.043</td>
</tr>
<tr>
<td>assumed</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 31

*2008 NJASK5 DFG-I Correlations*

<table>
<thead>
<tr>
<th></th>
<th>Program</th>
<th>GE Scale Math</th>
</tr>
</thead>
</table>
| Program                 | Pearson Correlation | 1 | .164*
| Sig. (2-tailed)         | .045    |               |
| N                       | 150     | 150           |
| GE Scale Math           | Pearson Correlation | .164* | 1
| Sig. (2-tailed)         | .045    |               |
| N                       | 150     | 150           |

*Correlation is significant at the 0.05 level (2-tailed).*
Subsidiary Question 6

What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-J General Education students on NJASK5 mathematics tests?

This question is supported by Hypotheses 11 and 12, which are investigated and analyzed using t-tests displayed in Tables 32-37.

Hypothesis 11. There is no significant difference in 2007 NJASK5 mathematics scores between DFG-J General Education students using the Everyday Mathematics program and those using a traditional mathematics program.

Table 32 indicates the 2007 third grade DFG-J mean mathematics scale score was higher for Everyday Mathematics schools (252.800) than for schools using traditional programs (247.792). While there was an observed difference in the mean scores, results of the t-test for two independent samples (see Table 33) indicate the means do not differ significantly at the $p < .05$ level ($p = .133$). Calculation of Cohen's $d$ (0.552) indicates a medium effect size for Everyday Mathematics.

Table 32

<table>
<thead>
<tr>
<th>Program</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE Scale Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traditional</td>
<td>12</td>
<td>247.792</td>
<td>10.1295</td>
<td>2.9241</td>
</tr>
<tr>
<td>EM</td>
<td>19</td>
<td>252.800</td>
<td>7.8522</td>
<td>1.8014</td>
</tr>
</tbody>
</table>
Table 33

*T-Test Comparing DFG-J Everyday Math and Traditional Program Schools on 2007 NJASK5*

<table>
<thead>
<tr>
<th>Levene's Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Sig.</td>
<td>t</td>
</tr>
<tr>
<td>---</td>
<td>------</td>
<td>---</td>
</tr>
<tr>
<td>GE Scale Math variances assumed</td>
<td>.724</td>
<td>.402</td>
</tr>
<tr>
<td>GE Scale Math variances not assumed</td>
<td>-1.458</td>
<td>19.240</td>
</tr>
</tbody>
</table>

Table 34

*2007 NJASK5 DFG-J Correlations*

<table>
<thead>
<tr>
<th>Program</th>
<th>GE Scale Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>1</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.133</td>
</tr>
<tr>
<td>N</td>
<td>31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Program</th>
<th>GE Scale Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>.276</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.133</td>
</tr>
<tr>
<td>N</td>
<td>31</td>
</tr>
</tbody>
</table>
Subsidiary Question 6

What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-J General Education students on NJASK5 mathematics tests?

Hypothesis 12. There is no significant difference in 2008 NJASK5 mathematics scores between DFG-J General Education students using the Everyday Mathematics program and those using a traditional mathematics program.

The findings are noted in Tables 35 and 36. The mean score for DFG-J Everyday Mathematics schools (253.395) for 2008 NJASK5 mathematics was higher than for schools using a traditional program (246.108). Table 36 displays results of the $t$-test for two independent samples indicating observed differences between Everyday Mathematics schools and traditional schools at the $p < .012$ level of significance. Calculation of Pearson $r$ (+.444) indicates a low positive correlation between NJASK scores and implementation of Everyday Mathematics. Calculation of Cohen's $d$ (0.9121) indicates a large effect size for Everyday Mathematics.

Table 35

<table>
<thead>
<tr>
<th>2008 NJASK5 DFG-J Mean Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program</td>
</tr>
<tr>
<td>----------------------------</td>
</tr>
<tr>
<td>GE Scale Math Traditional</td>
</tr>
<tr>
<td>EM</td>
</tr>
</tbody>
</table>
Table 36

*T-Test Comparing DFG-J Everyday Math and Traditional Program Schools on 2008 NJASK5*

<table>
<thead>
<tr>
<th></th>
<th>Levene's Test for Equality of Variances</th>
<th>t-test for Equality of Means</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>Sig.</td>
<td>t</td>
</tr>
<tr>
<td>GE Equal variances assumed</td>
<td>3.701</td>
<td>.064</td>
<td>-2.666</td>
</tr>
</tbody>
</table>

Table 37

*2008 NJASK5 DFG-J Correlations*

<table>
<thead>
<tr>
<th></th>
<th>Program</th>
<th>GE Scale Math</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program Pearson Correlation</td>
<td>1</td>
<td>.444*</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>31</td>
<td>.012</td>
</tr>
<tr>
<td>N</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>GE Scale Math Pearson Correlation</td>
<td>.444*</td>
<td>1</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.012</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>31</td>
<td>31</td>
</tr>
</tbody>
</table>

*: Correlation is significant at the 0.05 level (2-tailed).
Summary of Findings

This chapter will conclude with a brief discussion of the 6 Subsidiary Questions and the associated hypotheses, as well as the overarching research question. A complete evaluation of each hypothesis, along with future recommendations, will be included in Chapter V. An analysis of the null hypothesis data presented in Chapter IV highlights the following results of the study.

Subsidiary Question 1

What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-I General Education students on NJASK3 mathematics tests?

Hypothesis 1. There is no significant difference in 2007 NJASK3 mathematics scores between DFG-I General Education students using the Everyday Mathematics program and those using a traditional mathematics program. Hypothesis 1 was retained. The findings within the administered t-test concluded that there were differences in mean NJASK scores between Everyday Mathematics schools (246.1 18) and traditional schools (245.286). These differences were not statistically significant at the p < .05 level. Calculation of Cohen's d (0.1349) indicates a small effect size for Everyday Mathematics.

Hypothesis 2. There is no significant difference in 2008 NJASK3 mathematics scores between DFG-I General Education students using the Everyday Mathematics program and those using a traditional mathematics program. Hypothesis 2 was retained. The findings within the administered t-test concluded that there were differences in mean
NJASK scores between *Everyday Mathematics* schools (244.525) and traditional schools (245.880). These differences were not statistically significant at the $p < .05$ level.

Calculation of Cohen’s $d$ (0.2094) indicates a small effect size for traditional programs.

**Subsidiary Question 2**

What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-J General Education students on NJASK3 mathematics tests?

**Hypothesis 3.** There is no significant difference in 2007 NJASK3 mathematics scores between DFG-J General Education students using the *Everyday Mathematics* program and those using a traditional mathematics program. Hypothesis 3 was retained.

The findings within the administered $t$-test concluded that there were differences in mean NJASK scores between *Everyday Mathematics* schools (249.268) and traditional schools (247.200). These differences were not statistically significant at the $p < .05$ level.

Calculation of Cohen’s $d$ (0.4574) indicates a small to medium effect size for *Everyday Mathematics*.

**Hypothesis 4.** There is no significant difference in 2008 NJASK3 mathematics scores between DFG-J General Education students using the *Everyday Mathematics* program and those using a traditional mathematics program. Hypothesis 4 was retained.

The findings within the administered $t$-test concluded that there were differences in mean NJASK scores between *Everyday Mathematics* schools (248.058) and traditional schools (250.850). These differences were not statistically significant at the $p < .05$ level.
Calculation of Cohen’s $d$ (0.5461) indicates a medium effect size for traditional programs.

**Subsidiary Question 3**

What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-I General Education students on NJASK4 mathematics tests?

*Hypothesis 5.* There is no significant difference in 2007 NJASK4 mathematics scores between DFG-I General Education students using the *Everyday Mathematics* program and those using a traditional mathematics program. Hypothesis 5 is rejected.

The findings within the administered $t$-test concluded that there were differences in mean NJASK scores between *Everyday Mathematics* schools (252.964) and traditional schools (248.494).

The critical $t$-test value for a two-tailed test with 170 degrees of freedom ($df$) at the $p < .05$ level of significance is +/- 1.960. The SPSS printout indicates a $t$ ratio of -4.432 demonstrating a statistically significant difference in the means at the .000 level. Further statistical tests of correlation result in a Cohen’s $d$ value of 0.680. Based on Cohen’s (1988) parameters, this could be considered a medium effect size on 2007 NJASK4 DFG-I for the use of *Everyday Mathematics*. Calculation of Pearson $r = 0.322$, indicated a low positive correlation.

*Hypothesis 6.* There is no significant difference in 2008 NJASK4 mathematics scores between DFG-I General Education students using the *Everyday Mathematics* program and those using a traditional mathematics program. Hypothesis 6 is rejected.
The findings within the administered t-test concluded that there were differences in mean NJASK scores between *Everyday Mathematics* schools (253.234) and traditional schools (249.889).

The critical t-test value for a two-tailed test with 168 degrees of freedom ($df$) at the $p < .05$ level of significance is +/- 1.960. The SPSS printout indicates a $t$ ratio of -3.448 demonstrating a statistically significant difference in the means at the .001 level. Further statistical tests of correlation result in a Cohen’s $d$ value of 0.537. Based on Cohen’s (1988) parameters, this could be considered a medium effect size on 2008 NJASK4 DFG-I for the use of *Everyday Mathematics*. Calculation of Pearson $r = 0.257$, indicated little positive correlation.

**Subsidiary Question 4**

What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-J General Education students on NJASK4 mathematics tests?

*Hypothesis 7.* There is no significant difference in 2007 NJASK4 mathematics scores between DFG-J General Education students using the Everyday Mathematics program and those using a traditional mathematics program. Hypothesis 7 is rejected. The findings within the administered $t$-test concluded that there were differences in mean NJASK scores between *Everyday Mathematics* schools (257.405) and traditional schools (252.025). Calculation of Cohen’s $d$ (1.112) indicates a large effect size for *Everyday Mathematics.*
The critical $t$-test value for a two-tailed test with 29 degrees of freedom ($df$) at the $p < .05$ level of significance is $+/− 2.045$. The SPSS printout indicates a $t$ ratio of $-3.220$ demonstrating a statistically significant difference in the means at the .003 level. Further statistical tests of correlation result in a Cohen’s $d$ value of $1.112$. Based on Cohen’s (1988) parameters, this could be considered a large effect size on 2007 NJASK4 DFG-J for the use of Everyday Mathematics. Calculation of Pearson $r = 0.513$, indicated a moderate positive correlation.

**Hypothesis 8.** There is no significant difference in 2008 NJASK4 mathematics scores between DFG-J General Education students using the Everyday Mathematics program and those using a traditional mathematics program. Hypothesis 8 is rejected. The findings within the administered $t$-test concluded that there were differences in mean NJASK scores between Everyday Mathematics schools (258.042) and traditional schools (250.192).

The critical $t$-test value for a two-tailed test with 29 degrees of freedom ($df$) at the $p < .05$ level of significance is $+/− 2.045$. The SPSS printout indicates a $t$ ratio of $-5.831$ demonstrating a statistically significant difference in the means at the .000 level. Further statistical tests of correlation result in a Cohen’s $d$ value of $2.125$. Based on Cohen’s (1988) parameters, this could be considered a large effect size on 2008 NJASK4 DFG-J for the use of Everyday Mathematics. Calculation of Pearson $r = 0.735$, indicated a high positive correlation.
Subsidiary Question 5

What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-I General Education students on NJASK5 mathematics tests?

Hypothesis 9. There is no significant difference in 2007 NJASK5 mathematics scores between DFG-I General Education students using the Everyday Mathematics program and those using a traditional mathematics program. Hypothesis 9 was retained. The findings within the administered t-test concluded that there were differences in mean NJASK scores between Everyday Mathematics schools (246.527) and traditional schools (245.752). These differences were not statistically significant at the \( p < .05 \) level.

Calculation of Cohen's \( d \) (0.0929) indicates a small effect size for Everyday Mathematics.

Hypothesis 10. There is no significant difference in 2008 NJASK5 mathematics scores between DFG-I General Education students using the Everyday Mathematics program and those using a traditional mathematics program. Hypothesis 10 is rejected. The findings within the administered t-test concluded that there were differences in mean NJASK scores between Everyday Mathematics schools (248.071) and traditional schools (245.079).

The critical t-test value for a two-tailed test with 148 degrees of freedom (df) at the \( p < .05 \) level of significance is +/- 1.960. The SPSS printout indicates a t ratio of -2.022 demonstrating a statistically significant difference in the means at the .045 level. Further statistical tests of correlation result in a Cohen’s \( d \) value of 0.334. Based on Cohen’s (1988) parameters, this could be considered a small effect size on 2008 NJASK5 DFG-I
for the use of *Everyday Mathematics*. Calculation of Pearson $r = 0.164$, indicated little if
any positive correlation.

*Subsidiary Question 6*

What is the difference in student achievement, if any, between implementing a
reform-based mathematics program and the performance of DFG-J General Education
students on NJASK5 mathematics tests?

*Hypothesis 11*. There is no significant difference in 2007 NJASK5 mathematics
scores between DFG-J General Education students using the *Everyday Mathematics*
program and those using a traditional program. Hypothesis 11 was retained. The
findings within the administered $t$-test concluded that there were differences in mean
NJASK scores between *Everyday Mathematics* schools (252.800) and traditional schools
(247.792). These differences were not statistically significant at the $p < .05$ level.
Calculation of Cohen’s $d$ (0.552) indicates a medium effect size for *Everyday
Mathematics*.

*Hypothesis 12*. There is no significant difference in 2008 NJASK5 mathematics
scores between DFG-J General Education students using the *Everyday Mathematics*
program and those using a traditional mathematics program. Hypothesis 12 is rejected.
The findings within the administered $t$-test concluded that there were differences in mean
NJASK scores between *Everyday Mathematics* schools (253.395) and traditional schools
(246.108).

The critical $t$-test value for a two-tailed test with 29 degrees of freedom ($df$) at the $p$
< .05 level of significance is +/- 2.045. The SPSS printout indicates a $t$ ratio of -2.666
demonstrating a statistically significant difference in the means at the .012 level. Further statistical tests of correlation result in a Cohen's $d$ value of 0.912. Based on Cohen's (1988) parameters, this could be considered a large effect size on 2008 NJASK5 DFG-J for the use of *Everyday Mathematics*. Calculation of Pearson $r = 0.444$, indicated a low positive correlation.

Summary of Chapter IV and Description of Chapter V

Chapter IV provided an analysis of statistical data collected in this study. The chapter also summarized the findings.

Chapter V includes the summary of findings from the data analysis in Chapter V. The chapter includes discussion of the findings and conclusions related to the findings of the study. Chapter V concludes with recommendations for practice and policies, as well as recommendations for future research regarding the use of reform-based and traditional elementary mathematics programs.
Chapter V

SUMMARY OF FINDINGS, CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS

Introduction and Research Questions

The primary purpose of this study was to provide data about the performance of elementary students using a reform-based or traditional textbook program on New Jersey state tests. The study provides data to assist school and district leaders with information about the existence of a positive statistical correlation to student achievement, as measured by NJASK tests, given the implementation of a reform-based elementary mathematics program. A total of 177 schools classified as DFG-I with third grade students (2007-12,642 students; 2008-12,463 students), 172 schools classified as DFG-I with fourth grade students (2007-12,721 students; 2008-12,535 students), and 151 schools classified as DFG-I with fifth grade students (2007-12,979 students; 2008-12,717 students) were included in the study. Thirty-three schools classified as DFG-J with third grade students (2007-2,607; 2008-2,651 students), 31 schools classified as DFG-J with fourth grade students (2007-2,715 students; 2008-2,638) students, and 31 schools classified as DFG-J with fifth grade students (2007-2,633 students; 2008-2,737 students) were included in the study (see Appendix B).

The study was a nonexperimental, quantitative, cross-sectional, explanatory study designed to determine what difference, if any, existed between implementing a reform-based mathematics program, Everyday Mathematics, or a traditional mathematics program and the performance of DFG-I & J students on NJASK mathematics tests. Possible confounding variables were addressed through the use of DFG data. District
Factor Grouping results in similar districts being grouped together for comparative purposes. Controlling for SES indirectly results in districts similar in percentage of LEP, economically disadvantaged, and ethnic diversity being grouped together. Additionally, the use of General Education data eliminates special education and LEP students who may receive test accommodations.

This study used public domain data retrieved from the New Jersey Department of Education website. The information consisted of school level NJASK mathematics mean scores, DFG designation, and responses from districts regarding their elementary mathematics program. Schools included in the study were identified as using the reform-based mathematics program, *Everyday Mathematics*, or a traditional mathematics program. Schools using mixed programs or using a program for less than 3 years were excluded from the study. The study used the variables of State test data and elementary mathematics program to measure the correlation between mathematics program and student achievement.

A series of *t*-tests were used to determine if differences existed between elementary mathematics program and student achievement. The statistical analysis software SPSS 17.0 was programmed to evaluate the level of significance using a two-tailed test. In instances where \( p < .05 \), indicated existence of a significant relationship, calculation of Pearson \( r \), via the squared point biserial correlation coefficient, was utilized on the prescribed data to determine the strength of the relationship between elementary mathematics program and student achievement. Calculation of Cohen's \( d \) was used to determine the effect on student achievement of a traditional or reform-based mathematics program. Findings and
recommendations from this study should help school leaders make decisions about program selection and implementation.

The study had a single research question with six subsidiary questions. The research question examined the influence of implementing a reform-based mathematics program on the mathematics achievement, as measured by NJASK tests, of General Education students in grades 3 through 5 who attend school in New Jersey school districts classified as DFG-I and DFG-J? The six subsidiary questions and associated hypotheses follow.

Summary of Findings

This study examined the research question via the testing of two hypotheses for each subsidiary question. The data generated were designed to test the hypothesis that there is no patterned influence of the use of the reform-based elementary mathematics program, Everyday Mathematics, on student achievement as measured by the NJASK3, NJASK4, or NJASK5.

Subsidiary Question 1: What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-I General Education students on NJASK3 mathematics tests?

The first t-test computed the difference between DFG-I General Education third grade mathematics performance on 2007 NJASK and implementation of a traditional mathematics program or Everyday Mathematics. While the mean mathematics score for Everyday Mathematics (246.118) schools was higher than traditional schools (245.286), the means do not differ significantly at the $p < .05$ level ($p = .371$). Calculation of
Cohen’s $d$ (0.1349) indicated a small difference in the mean scores existed between implementation of *Everyday Mathematics* or a traditional program and student achievement on NJASK mathematics.

The second $t$-test computed the difference between DFG-I General Education third grade mathematics performance on 2008 NJASK and implementation of a traditional mathematics program or *Everyday Mathematics*. While the mean mathematics score for *Everyday Mathematics* (244.525) schools was lower than traditional schools (245.880), the means do not differ significantly at the $p < .05$ level ($p = .163$). Calculation of Cohen’s $d$ (0.2094) indicated a small difference in mean scores existed between implementation of a traditional mathematics program or *Everyday Mathematics* and student achievement on NJASK mathematics.

Subsidiary Question 2: What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-J General Education students on NJASK3 mathematics tests?

The third $t$-test computed the difference between DFG-J General Education third grade mathematics performance on 2007 NJASK and implementation of a traditional mathematics program or *Everyday Mathematics*. While the mean mathematics score for *Everyday Mathematics* (249.268) schools was higher than traditional schools (247.200), the means do not differ significantly at the $p < .05$ level ($p = .194$). Calculation of Cohen’s $d$ (0.4574) indicated a small to medium difference in mean scores existed between implementation of *Everyday Mathematics* or a traditional program and student achievement on NJASK mathematics.
The fourth *t*-test computed the difference between DFG-J General Education third grade mathematics performance on 2008 NJASK and implementation of a traditional mathematics program or *Everyday Mathematics*. While the mean mathematics score for *Everyday Mathematics* (248.058) schools was lower than traditional schools (250.850), the means do not differ significantly at the *p* < .05 level (*p* = .132). Calculation of Cohen’s *d* (0.5461) indicated a medium difference in the mean scores existed between implementation of *Everyday Mathematics* or a traditional program and student achievement on NJASK mathematics.

Subsidiary Question 3: What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-I General Education students on NJASK4 mathematics tests?

The fifth *t*-test computed the difference between DFG-I General Education fourth grade mathematics performance on 2007 NJASK and implementation of a traditional mathematics program or *Everyday Mathematics*. The mean mathematics score for *Everyday Mathematics* (252.964) schools was higher than traditional schools (248.494) and statistically significant significantly at the *p* < .05 level (*p* = .000). Calculation of Cohen’s *d* (0.6804) indicated a medium difference in the mean scores existed between implementation of *Everyday Mathematics* or a traditional program and student achievement on NJASK mathematics. Calculation of Pearson *r* = .322 indicated a low positive correlation.

The sixth *t*-test computed the difference between DFG-I General Education fourth grade mathematics performance on 2008 NJASK and implementation of the *Everyday Mathematics* program. The mean mathematics score for *Everyday Mathematics*
(253.234) schools was higher than traditional schools (249.889) and statistically significant significantly at the $p < .05$ level ($p = .001$). Calculation of Cohen’s $d$ (0.537) indicated a medium difference existed between implementation of *Everyday Mathematics* or a traditional program and student achievement on NJASK mathematics. Calculation of Pearson $r = 0.257$, indicated little positive correlation.

Subsidiary Question 4: What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-J General Education students on NJASK4 mathematics tests?

The seventh $t$-test computed the difference between DFG-J General Education fourth grade mathematics performance on 2007 NJASK and implementation of a traditional mathematics program or *Everyday Mathematics*. The mean mathematics score for *Everyday Mathematics* (257.405) schools was higher than traditional schools (252.025) and statistically significant significantly at the $p < .05$ level ($p = .003$). Calculation of Cohen’s $d$ (1.112) indicated a large difference in the mean scores existed between implementation of *Everyday Mathematics* or a traditional program and student achievement on NJASK mathematics. Calculation of Pearson $r = 0.513$, indicated a moderate positive correlation. The size of this correlation allows the researcher to infer a positive relationship between the reform-based mathematics program and student achievement on the NJASK4 test. The size of the correlation (.41-.60) is “Large enough to be of practical as well as theoretical use” (Fraenkel & Wallen, 1996, p. 233).

The eighth $t$-test computed the difference between DFG-J General Education fourth grade mathematics performance on 2008 NJASK and implementation of a traditional mathematics program or *Everyday Mathematics*. The mean mathematics score for
Everyday Mathematics (258.042) schools was higher than traditional schools (250.192) and statistically significant significantly at the \( p < .05 \) level \( (p = .000) \). Calculation of Cohen's \( d \) (2.125) indicated a large difference existed between implementation of Everyday Mathematics or a traditional program and student achievement on NJASK mathematics. Calculation of Pearson \( r = 0.735 \), indicated a high positive correlation. The size of the correlation (.61-.80) is "Very important, but rarely obtained in educational research" (Fraenkel & Wallen, 1996, p. 233).

Subsidiary Question 5: What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-I General Education students on NJASK5 mathematics tests?

The ninth \( t \)-test computed the difference between DFG-I General Education fifth grade mathematics performance on 2007 NJASK and implementation of a traditional mathematics program or Everyday Mathematics. While the mean mathematics score for Everyday Mathematics (246.527) schools was higher than traditional schools (245.752), the means do not differ significantly at the \( p < .05 \) level \( (p = .570) \). Calculation of Cohen's \( d \) (0.0929) indicated a small difference in the mean scores existed between implementation of Everyday Mathematics or a traditional program and student achievement on NJASK mathematics.

The tenth \( t \)-test computed the difference between DFG-I General Education fifth grade mathematics performance on 2008 NJASK and implementation of a traditional mathematics program or Everyday Mathematics. The mean mathematics score for Everyday Mathematics (248.071) schools was higher than traditional schools (245.079) and statistically significant significantly at the \( p < .05 \) level \( (p = .045) \). Calculation of
Cohen’s $d$ (0.334) indicated a small difference in the mean scores existed between implementation of *Everyday Mathematics* or a traditional program and student achievement on NJASK mathematics. Calculation of Pearson $r = 0.164$, indicated little positive correlation.

Subsidiary Question 6: What is the difference in student achievement, if any, between implementing a reform-based mathematics program and the performance of DFG-J General Education students on NJASK5 mathematics tests?

The eleventh $t$-test computed the difference between DFG-J General Education fifth grade mathematics performance on 2007 NJASK and implementation of a traditional mathematics program or *Everyday Mathematics*. While the mean mathematics score for *Everyday Mathematics* (252.800) schools was higher than traditional schools (247.792), the means do not differ significantly at the $p < .05$ level ($p = .133$). Calculation of Cohen’s $d$ (0.552) indicated a medium difference in the mean scores existed between implementation of *Everyday Mathematics* or a traditional program and student achievement on NJASK mathematics.

The twelfth $t$-test computed the difference between DFG-J General Education fifth grade mathematics performance on 2008 NJASK and implementation of a traditional mathematics program or *Everyday Mathematics*. The mean mathematics score for *Everyday Mathematics* (253.395) schools was higher than traditional schools (246.108) and statistically significant significantly at the $p < .05$ level ($p = .012$). Calculation of Cohen’s $d$ (0.9121) indicated a large difference in the mean scores existed between implementation of *Everyday Mathematics* or a traditional program and student achievement on NJASK mathematics. Calculation of Pearson $r = 0.444$, indicated a low
positive correlation, but the size of the correlation (.41-.60) is "Large enough to be of practical as well as theoretical use" (Fraenkel & Wallen, 1996, p. 233).

Discussion and Conclusions

These data and statistical analyses indicated mixed results for the existence of a relationship between implementation of a reform-based elementary mathematics program and student achievement on NJASK. Results indicated statistically significant (p < .05) differences existed between fourth grade DFG-I and DFG-J General Education students’ performance on 2007 and 2008 NJASK and schools implementing Everyday Mathematics. A similar positive correlation between fifth grade DFG-I and DFG-J General Education students’ performance on 2008 NJASK existed for schools using the Everyday Mathematics program. Three of the Pearson r calculations (2007 Grade 4, DFG-J; 2008 Grade 4, DFG-J; 2008 Grade 5, DFG-J) were large enough to be useful to educators.

The data and statistical analyses indicated there were no statistically significant differences between either implementation of the Everyday Mathematics program or a traditional program and third grade DFG-I and DFG-J General Education students’ performance on 2007 and 2008 NJASK. Similarly, no statistically significant differences existed between fifth grade DFG-I and DFG-J General Education students’ performance on 2007 NJASK. While the data and statistical analyses indicated mixed results for the existence of a relationship between implementation of a reform-based elementary mathematics program and student achievement on NJASK, several points deserve further exploration.
Mean Scale Scores

1. Of the 12 mean scale score comparisons between reform-based and traditional schools on NJASK performance, the 6 indicating statistically significant differences favored schools using *Everyday Mathematics*. None of the 12 comparisons indicated statistically significant higher mean performance by traditional program schools.

2. Of the 12 comparisons between reform-based and traditional schools on NJASK performance at grades 3-5, 10 indicated higher mean scores by *Everyday Mathematics* schools.

3. Of the 8 comparisons between *Everyday Mathematics* schools and traditional schools on NJASK performance at grades 4 and 5, all indicated higher mean performance by *Everyday Mathematics* schools. Seventy-five percent of the comparisons at grades 4 and 5 indicated statistically significant differences favoring schools using *Everyday Mathematics*.

These results indicate generally higher mean scale score performance for schools implementing the *Everyday Mathematics* program than schools using a traditional program. These differences were more consistent at grades 4 and 5 with more frequent statistical significance. The results suggest support for the theories and research about how students learn and assimilate new information and experiences into their current conceptual understandings (Carpenter et al., 1989; Cobb & Steffe, 1983; Davis, 1992; Kamii & DeClark, 1985; Maher, Davis, & Alston, 1992a), which served as the basis for reform-based mathematics programs and the changing role for teachers (Carpenter, Fennema, & Franke, 1996; Cobb, Wood, Yackel, & McNeal, 1992; Ferrini-Mundy & Johnson, 1994; Hiebert & Wearne, 1993; Maher, Davis, & Alston, 1992a). The results of
the current research somewhat suggest that the longer students are in the reform-based program, the stronger their performance is relative to a traditional program. This is similar to the results noted by Sconiers et al. (2003, p. 7) in the ARC Center Study, where the performance of students in a reform-based program improved over time relative to students in traditional programs.

**Effect Sizes**

1. Ten of 12 grade 3-5 comparisons between traditional programs and *Everyday Mathematics* indicated effect sizes ranging from small (0.13) to large (2.13) for the *Everyday Mathematics* program.

2. Of the 2 comparisons that favored traditional programs, one had a small effect size (2007 DFG-I, Grade 3, ES = 0.20) and the other a medium effect size (2008 DFG-J, Grade 3 ES = 0.55).

3. In the 8 comparisons at grades 4 and 5, six of the effect sizes for *Everyday Mathematics* were medium (2007 DFG-I, Grade 4, ES = 0.68; 2008 DFG-I, Grade 4, ES = 0.54, 2007 DFG-J, Grade 5, ES = 0.55) or large (2007 DFG-J, Grade 4, ES = 1.11; 2008 DFG-J, Grade 4, ES = 2.13; 2008 DFG-J, Grade 5, ES = 0.91).

**Additional effect size comparisons**

While longitudinal cohort comparisons (Example- DFG-I 2007 Gr. 3 to 2008 Gr. 4) do not provide identical student populations, the low student mobility rates in DFG-I and DFG-J districts (typically under 10%) allow for some interesting comparisons.

1. While not statistically significant (p < .05), the mean scores for 2007 grade 3 DFG-I schools favored *Everyday Mathematics* (246.118) over traditional schools (245.286) with a small effect size of 0.13. A year later, the difference in 2008 grade 4
DFG-I mean scores was statistically significant with a medium effect size of 0.53 favoring *Everyday Mathematics* schools (253.234) over traditional schools (249.889).

2. The mean scores for 2007 grade 3 DFG-J schools favored *Everyday Mathematics* (249.268) over traditional schools (247.200) and were not statistically significant, but calculation of Cohen’s $d$ indicated a small to medium effect size of 0.457 for *Everyday Mathematics*. In 2008, the DFG-J grade 4 scores favored *Everyday Mathematics* (258.042) over traditional programs (250.192) with a large effect size of 2.13.

3. The mean scores for 2007 grade 4 DFG-I schools favored *Everyday Mathematics* (252.964) over traditional schools (248.494), were statistically significant, and calculation of Cohen’s $d$ indicated a medium effect size of 0.68 for *Everyday Mathematics*. In 2008, the DFG-I grade 5 scores were again statistically significant and favored *Everyday Mathematics* (248.071) over traditional programs (245.079) but had a small effect size of 0.33.

4. The mean scores for 2007 grade 4 DFG-J schools favored *Everyday Mathematics* (257.405) over traditional schools (252.052) and were statistically significant, with a large effect size of 1.112 for *Everyday Mathematics*. In 2008, the DFG-J grade 5 scores again favored *Everyday Mathematics* (253.395) over traditional programs (246.108) with a large effect size of 0.912.

While the differences in school level, mean scale scores between the traditional and reform-based program are inconclusive, the data indicate generally stronger performance by *Everyday Mathematics* schools. The effect sizes for both *Everyday Mathematics* and a traditional mathematics program are small to moderate at the third grade level. While not statistically significant, two third grade comparisons favored *Everyday Mathematics*
(2007 DFG-I and 2007 DFG-J) and two favored traditional programs (2008 DFG-I and 2008 DFG-J). At fourth grade, all comparisons indicated medium to large effect sizes suggesting the longer students are in a reform program the greater the effect size (Riordan & Noyce, 2001; Sconiers et al., 2003, p. 7). Slavin and Lake (2008) suggest alternative reasons for this increasing effect size, indicating it may be a result of “survivors” being more likely to be included in a study of this type. Schools that begin with a reform program sometimes abandon it, leaving the more capable schools as a part of the study (p.433).

When examining the comparisons between differences in the mean scores for third grade DFG-I and DFG-J to fourth grade DFG-I and DFG-J (number 1 and 2 above), the relationship between Everyday Mathematics and NJASK scores appears to strengthen with higher effect size at grade 4 than at grade 3. This may indicate that the influence of a reform-based program on student achievement is seen over a longer period of time and immediate effects should not be expected by parents, teachers, or school leaders.

Continuing the comparison from grade 4 to grade 5 (number 3 and 4 above), the effect sizes do not increase in a similar manner. While the effect sizes at fifth grade continued to favor Everyday Mathematics, they were smaller than fourth grade. This finding does not yield an easy explanation and is not supported by either the views of Sconiers et al. (2003) or Slavin and Lake (2008). However, these results need further exploration, as the number of items on the third, fourth, and fifth grade NJASK tests have varied from test to test and year to year. In 2007 and 2008 the third grade test had 33 possible raw score points comprised of 27 multiple choice items 24 points) and 3 open-ended questions (9 points). The fourth grade test has traditionally been a longer test, and
in 2007 and 2008 had 43 points divided among 32 multiple choice (28 points) and 5 open-ended items (15 points). The 2007 fifth grade test had 39 points divided among 30 multiple choice and 9 open-ended items, but the 2008 test contained 50 raw score points (excluding field test items) with 42 multiple choice, 8 short constructed response, and 5 extended-response questions. It is interesting to note that tests with statistically significant differences in mean scores between Everyday Mathematics and traditional programs (fourth grade 2007 and 2008; fifth grade 2008) were the tests with more total items. Beginning in 2009, the NJASK 3-5 tests will all have 50 possible raw score points divided among multiple choice, short constructed-response, and extended constructed-response items, allowing for further exploration of the potential influence of reform-based programs on student achievement.

**Standard Deviations**

Examination of the standard deviations for mean scale scores for traditional (T) and Everyday Mathematics (E.M.) schools reveals generally lower standard deviations for Everyday Mathematics (see Table 38). In only two instances (2008 Gr. 3 DFG-J and 2008 Gr. 5 DFG-I) were the standard deviations for Everyday Mathematics greater than traditional programs. This would suggest that students using the Everyday Mathematics program more frequently have less deviation from the mean score, indicating a narrower band of student achievement. This finding indicates that Everyday Mathematics may have the potential to reduce achievement gaps among students. Combined with generally higher mean scores, the lower standard deviations may indicate that the Everyday Mathematics program results in higher mean NJASK mathematics performance with less deviation from the mean. One broad objective of the NCTM Standards
documents (NCTM, 1989, 1991, 1995, 2006) was to increase the levels of achievement for all students and to provide greater access and equity (Jones & Coxford, 1970; Schoenfeld, 2004). The standard deviation results of the current study suggest that *Everyday Mathematics* may successfully narrow the gap between high scoring and low scoring students while raising the level of achievement for all children.

Table 38

*Standard Deviations by Test and Program*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2007 Gr. 3 DFG-I</td>
<td>5.660</td>
<td>6.627</td>
</tr>
<tr>
<td>2008 Gr. 3 DFG-I</td>
<td>5.711</td>
<td>7.149</td>
</tr>
<tr>
<td>2007 Gr. 3 DFG-J</td>
<td>3.900</td>
<td>5.066</td>
</tr>
<tr>
<td>2008 Gr. 3 DFG-J</td>
<td>5.206</td>
<td>5.015</td>
</tr>
<tr>
<td>2007 Gr. 4 DFG-I</td>
<td>6.243</td>
<td>6.879</td>
</tr>
<tr>
<td>2008 Gr. 4 DFG-I</td>
<td>6.217</td>
<td>6.237</td>
</tr>
<tr>
<td>2007 Gr. 4 DFG-J</td>
<td>3.392</td>
<td>5.941</td>
</tr>
<tr>
<td>2008 Gr. 4 DFG-J</td>
<td>3.517</td>
<td>3.860</td>
</tr>
<tr>
<td>2007 Gr. 5 DFG-I</td>
<td>8.133</td>
<td>8.530</td>
</tr>
<tr>
<td>2008 Gr. 5 DFG-I</td>
<td>9.297</td>
<td>8.602</td>
</tr>
<tr>
<td>2007 Gr. 5 DFG-J</td>
<td>7.852</td>
<td>10.129</td>
</tr>
<tr>
<td>2008 Gr. 5 DFG-J</td>
<td>5.199</td>
<td>10.029</td>
</tr>
</tbody>
</table>

The data and statistical analyses from this study indicated mixed results for the existence of a relationship between implementation of a reform-based elementary mathematics program and student achievement, and the results may be placed in the
context of prior research. Authors have suggested that the levels of U.S. student performance using traditional programs have been well documented as less than satisfactory on national (NAEP, 1983) and international tests (Hiebert, 1999; Johnson & Risling, 1967; Schoen, et al., 1999; Senk & Thompson, 2003; Wilson & Blank, 1999). Several previous studies found no statistically significant differences in student achievement for the traditional programs published by Scott Foresman-Addison Wesley and Harcourt Achieve (Resendez & Manley, 2005; Resendez & Azin, 2005; Resendez & Azin, 2006). While a number of studies found positive results for specific traditional programs (Agodini, et al., 2009; EDSTAR, 2004; Johnson & Hall, 2003; Johnson, J., Yanyo, L., & Hall, 2002), the current research found no statistically significant differences in mean scores favoring traditional programs.

Student performance on NJASK tests was stronger for the reform-based schools on the longer versions of the tests at grade 4 (2007 and 2008) and 5 (2008), indicating possible support for theories that active involvement in problem solving, incorporation of manipulative materials, and opportunities for children to develop their own procedures and actively construct their own knowledge would allow students to develop important mathematics skills and deeper conceptual understanding (Carpenter et al., 1989; Cobb et al., 1991; Hiebert, 1999; Hiebert & Wearne, 1996; Kamii & DeClark, 1985; Mack, 1990; Wearne & Hiebert, 1989). The longer tests may have allowed students in the reform-based schools to better demonstrate their mathematics skills and conceptual understanding.

Student results on NJASK tests indicated generally higher mean scores for *Everyday Mathematics* schools. These results were similar to results found in prior
research. Carroll (1997) reported results for 26 schools using *Everyday Mathematics* on the Illinois Goal Assessment Program, in which 25 schools had mean scores significantly above the Illinois state mean and none below the state mean. A quasi-experimental study of fourth grade student performance on Massachusetts state-wide testing compared *Everyday Mathematics* schools with traditional program schools (Riordan & Noyce, 2001). Schools using *Everyday Mathematics* outperformed traditional schools in all types of questions and effect sizes were larger for schools using *Everyday Mathematics* for more than 4 years (ES = 0.34) than for 2-3 years (ES = 0.15). Other studies of *Everyday Mathematics* have found positive results while making different comparisons (Baxter et al., 2001; Carroll, 1998a; Fuson et al., 2000; Sconiers et al., 2003; Waite, 2000; Woodward & Baxter, 1997).

**Recommendations for Policy and Practice**

While this study suggests sometimes strong relationships between *Everyday Mathematics* and student achievement on New Jersey state tests, without similar results for traditional programs, neither the results of this study nor previous studies provide clear guidance for educational decision makers. “More research is needed on all of these programs but the evidence to date suggests a surprising conclusion that despite all the heated debates about the content of mathematics, there is limited high-quality evidence supporting differential effects of different math curricula” (Slavin & Lake, p. 445).

Given the frequent debate about the selection and implementation of elementary mathematics programs, several suggestions are offered to educational leaders.

The relationships between classroom instructional programs and student learning are complex with multiple variables involved in determining student achievement and
success. Research should guide the decision making of schools and districts in the selection, purchase, and implementation of educational programs. However, parents often want to be a part of the decision making around their child’s education (Sarason, 1995), not only in what schools they attend but in the type of instruction they receive. In order to avoid struggles over what gets taught and how the instruction takes place, a clear role needs to be defined for parents (Peressini, 1998) around the selection of elementary mathematics programs.

Similarly, clearly defined roles for teachers’ involvement in the decision making process will need to be developed, as changes in teaching practices often require changes in teacher beliefs about mathematics and learning (Battista, 1994; Remillard, 2000). Leaders will need to be fully cognizant of the nature of schools (Fullan, 1995) and the requirements for bringing about change in teacher practices (Fullan, 1996; Hinde, 2003).

Considerable debate continues to take place about the selection and implementation of mathematics curricula and programs, with few definitive answers. As such, school leaders should look to other avenues for improving students’ mathematics performance. Computer assisted instruction (CAI) is one area needing additional study. Slavin and Lake (2008) identified 38 studies of CAI, with 15 involving randomized or randomized quasi-experimental designs for elementary mathematics materials (p. 445). In most instances the CAI intervention involved no more than three 30-minute sessions per week, and the median effect size was +0.19 (p.459).

Perhaps more promising are studies involving professional development and improvement of teachers’ instructional process strategies (cooperative learning, mastery learning, math content knowledge, direct instruction). Slavin and Lake (2008) identified
36 studies evaluating instructional process strategies, with 19 using randomized or randomized quasi-experimental designs (p. 475). The median effect size for 9 studies involving cooperative learning at the elementary level was +0.29. Studies involving peer tutoring and peer-assisted learning were also found to have positive effects. Given these results, it is incumbent upon school leaders to explore up-to-date research and multiple methods for improving students' mathematics learning.

Suggestions for Future Research

The implementation of the reform-based elementary mathematics program, *Everyday Mathematics*, as examined in this study indicated a generally positive relationship to student achievement as measured by NJASK tests. These results provide guidance for additional recommendations in clarifying the role of reform-based mathematics programs in student learning. Some areas for future research are noted below:

1. Identify several comparable cohorts of students to participate in a research study using an experimental design with random assignment of reform and traditional programs beginning in kindergarten or first grade. Measure student mathematics achievement at the baseline and at the completion of first, second, and third grade years (Agodini et al. 2009).

2. Identify several comparable cohorts of students in reform-based schools and traditional schools that can be matched at the baseline for student mathematics achievement. Match at the baseline using an assessment different than NJASK scores (e.g. Terra Nova). Compare achievement in subsequent years to identify significant differences in performance.
3. Identify several comparable cohorts of students in reform-based schools and traditional schools that can be matched at the baseline for student mathematics achievement. Given the lack of significant difference found in third grade NJASK scores between reform-based and traditional schools in this study, use third grade NJASK scores as part of the baseline matching process. Compare achievement in subsequent years to identify possible relationships between student performance and mathematics program.

4. Identify several comparable cohorts of students in specific reform-based, traditional, or mixed-program schools that can be matched at the baseline using for student mathematics achievement using third grade NJASK or a different standardized measure (e.g. Terra Nova). Compare achievement in subsequent years to identify significant differences in performance.

5. Identify several reform-based and traditional districts willing to release student-level data. Select students in reform-based and traditional programs who can be matched for across a number of factors (mathematics achievement, reading achievement, ethnicity, SES). Measure student achievement on a longitudinal basis for differential effects.

6. Replicate the current study using data from different District Factor Groups within the state to determine whether similar or different relationships are found. Identify whether differences exist among traditional and reform-based schools in different DFG groups which may indicate stronger or weaker relationships among various student populations.

7. Conduct an experimental study measuring the influence of professional development in mathematics (mathematics teaching methods, mathematics content knowledge) on student mathematics achievement in either a traditional or reform-based
program. Using two groups equal at the baseline, measure the relationship of professional development to student achievement.

8. Compare student achievement of reform-based and traditional schools on problem solving measures (e.g. raw data from NJASK on open-ended problems, NJASK items identified as Problem Solving, other standardized measures) to determine possible relationships among type of math program and student performance on problem solving tasks.

9. Identify schools consistently achieving the highest levels of performance on NJASK testing within various District Factor Groups. Develop a qualitative study that will identify common school-level factors impacting student performance (professional development, parent involvement, teacher experience level, teacher educational attainment, student grouping for instruction).

Summary

Chapter I of this research study provided background information on the growing consensus of the need for improvement in U.S. students' mathematics achievement. Chapter I included the statement of the problem, purpose of the study, research question, research hypothesis, significance of the study, limitations and delimitations of the study, and definition of key terms. Chapter II provided a review of the relevant literature, including the literature search method and criteria for inclusion in the literature review. The review was divided into a history of mathematics instructional reform in the U.S., information about mathematics learning, research studies on reform-based and traditional mathematics programs, and student achievement. Chapter III reviewed the research
design, participants, setting for the study, treatment, methods, data collection, variable, sampling, instrumentation, procedures, methods of data analysis, controls and step-by-step procedures. In Chapter IV, the researcher presented research findings for the 6 subsidiary questions and 12 associated hypotheses, including acceptance or rejection of each hypothesis. Chapter V presented the findings with discussion and conclusions. Recommendations for policy and practice, along with recommendations for future study, were also presented.


http://www.aasa.org/publications/saarticledetail.cfm?ItemNumber=3320


Houghton Mifflin Company.


New Jersey Department of Education. (2004). *New Jersey Department of Education district factor groups (DFG) for school districts.* Retrieved June 14, 2008, from The State of New Jersey Web Site:  
http://www.state.nj.us/education/finance/sf/dfg/shtml.


http://education.state.nj.us/cccs/?_intro.


http://everydaymath.uchicago.edu


Appendix A

Questions to DFG-I and DFG-J Schools and Districts
Appendix A presents the E-mail and phone questions to DFG-I and DFG J schools and districts.

Dear __________:

Your assistance with the following questions is appreciated.

What math program do you use in the following grade(s) in your school(s)?

K-
1-
2-
3-
4-
5-

How long have you been using the program?

If less than 3 years, what program did you use previously and when did you begin using it?

Who is the best person in your school to contact regarding any further questions about the mathematics program?

Thank you.

Bill Ward
Superintendent of Schools
Old Tappan, NJ
Appendix B

Number of Schools and Students by DFG and Year
Appendix B presents the number of schools and students in each DFG for each year of NJASK Testing.

<table>
<thead>
<tr>
<th>Year</th>
<th>DFG</th>
<th>Grade</th>
<th>Prog.</th>
<th># of schools</th>
<th># of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>I</td>
<td>3</td>
<td>Trad.</td>
<td>73</td>
<td>4,922</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E.M.</td>
<td>104</td>
<td>7,720</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>177</td>
<td>12,642</td>
</tr>
<tr>
<td>2008</td>
<td>I</td>
<td>3</td>
<td>Trad.</td>
<td>74</td>
<td>4,943</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E.M.</td>
<td>103</td>
<td>7,520</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>177</td>
<td>12,463</td>
</tr>
<tr>
<td>2007</td>
<td>J</td>
<td>3</td>
<td>Trad.</td>
<td>14</td>
<td>980</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E.M.</td>
<td>19</td>
<td>1,627</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>33</td>
<td>2,607</td>
</tr>
<tr>
<td>2008</td>
<td>J</td>
<td>3</td>
<td>Trad.</td>
<td>14</td>
<td>1,046</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E.M.</td>
<td>19</td>
<td>1,605</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>33</td>
<td>2,651</td>
</tr>
<tr>
<td>2007</td>
<td>I</td>
<td>4</td>
<td>Trad.</td>
<td>71</td>
<td>4,962</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E.M.</td>
<td>101</td>
<td>7,759</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>172</td>
<td>12,721</td>
</tr>
<tr>
<td>2008</td>
<td>I</td>
<td>4</td>
<td>Trad.</td>
<td>70</td>
<td>5,013</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E.M.</td>
<td>100</td>
<td>7,522</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>170</td>
<td>12,535</td>
</tr>
<tr>
<td>2007</td>
<td>J</td>
<td>4</td>
<td>Trad.</td>
<td>12</td>
<td>1,042</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E.M.</td>
<td>19</td>
<td>1,673</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>31</td>
<td>2,715</td>
</tr>
<tr>
<td>2008</td>
<td>J</td>
<td>4</td>
<td>Trad.</td>
<td>12</td>
<td>990</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E.M.</td>
<td>19</td>
<td>1,648</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>31</td>
<td>2,638</td>
</tr>
<tr>
<td>2007</td>
<td>I</td>
<td>5</td>
<td>Trad.</td>
<td>66</td>
<td>4,888</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E.M.</td>
<td>85</td>
<td>8,091</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>151</td>
<td>12,979</td>
</tr>
<tr>
<td>Year</td>
<td>DFG</td>
<td>Grade</td>
<td>Prog.</td>
<td># of schools</td>
<td># of students</td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
<td>-------</td>
<td>-------</td>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td>2008</td>
<td>I</td>
<td>5</td>
<td>Trad.</td>
<td>66</td>
<td>5,047</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E.M.</td>
<td>84</td>
<td>7,670</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>150</td>
<td>12,717</td>
</tr>
<tr>
<td>2007</td>
<td>J</td>
<td>5</td>
<td>Trad.</td>
<td>12</td>
<td>1,020</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E.M.</td>
<td>19</td>
<td>1,717</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>31</td>
<td>2,633</td>
</tr>
<tr>
<td>2008</td>
<td>J</td>
<td>5</td>
<td>Trad.</td>
<td>12</td>
<td>1,032</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E.M.</td>
<td>19</td>
<td>1,075</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>31</td>
<td>2,737</td>
</tr>
</tbody>
</table>