12-2000

Effects of hand-held calculators on precollege students in mathematics classes - a meta-analysis

Aimee Joy Ellington

To the Graduate Council:

I am submitting herewith a dissertation written by Aimee Joy Ellington entitled "Effects of hand-held calculators on precollege students in mathematics classes - a meta-analysis." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Education.

Donald J. Dessart, Major Professor

We have read this dissertation and recommend its acceptance:

E. Dale Doak, Stephanie O. Robinson, Carl G. Wagner

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Aimee Joy Ellington entitled "Effects of Hand-Held Calculators on Precollege Students in Mathematics Classes – A Meta-Analysis." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Education.

Donald J. Dessart, Major Professor

We have read this dissertation
and recommend its acceptance:

E. Dale Doak

Stephanie O. Robinson

Carl G. Wagner

Accepted for the Council:

Interim Vice Provost and
Dean of The Graduate School

EFFECTS OF HAND-HELD CALCULATORS ON PRECOLLEGE STUDENTS IN MATHEMATICS CLASSES – A META-ANALYSIS

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Aimee Joy Ellington

December 2000

ACKNOWLEDGEMENTS

Any undertaking of significance is seldom attributable to a single individual. I would like to acknowledge a few of the people who have contributed, in unique ways, to the completion of my graduate school experience. I am indebted to many friends, family members, and colleagues for their suggestions, guidance, support, and encouragement.

I would like to thank my advisor and committee chairperson, Dr. Donald J. Dessart. His advice regarding course and research opportunities was invaluable. His constructive comments during the research and writing phases of the dissertation kept me focused on my purpose. My thanks, also, to the other members of my committee: Dr. Stephanie Robinson, whose courses in mathematics education gave me a solid foundation on which to build my graduate research program; Dr. Carl Wagner, who challenged me mathematically and motivated me to communicate my ideas as clearly as possible; and Dr. Dale Doak, whose technology-based courses inspired me to undertake a research project involving technology in the mathematics classroom.

Many friends and colleagues have provided me with moral support from the time I took my first graduate course to the moment the last sentence of the dissertation was written. They helped me believe in myself at every stage of this momentous undertaking: Mayme Kay Banasiak, Kimberly Brookshire, Connie Dalton, Dr. Cynthia Gettys, Dr. Sandoval Melim, Dr. Susanna Reid, and Dr. Margie Ribble. Two other wonderful friends provided assistance with many technical aspects of the dissertation writing process: Kathy Manns, who read every word of the constantly evolving manuscript and provided wonderful editorial suggestions; Cara Dalton, who provided clarification regarding the many subtle nuances of the English language. A wonderful mentor, Sarah Jane Liers, encouraged me to explore my interests in mathematics. She motivated me to pursue a career in the education field. As a life-long learner, she continues to be an inspiration for me.

Finally, my family has been a tremendous source of unconditional love and understanding as I have pursued my educational dreams. My parents, Joyce and Douglas Ellington, were supportive at every stage of my educational career. They reminded me that education is an adventure and stressed the importance of taking time to enjoy the processes involved in achievement. My brother, Eric Ellington, provided emotional support and humor at pivotal times during the dissertation process. My great aunt and uncle, Lucille and George Skinner, encouraged me through all of my educational endeavors and were wonderful reminders that my grandparents would be proud of my accomplishments.

ABSTRACT

This study was conducted to determine the effect of hand-held calculators on students in precollege mathematics classes. Achievement and attitude were the educational constructs under consideration. The findings of fifty-three calculator-based research studies were integrated through meta-analysis. The methods of data analysis were based on Hembree's model, featuring fundamental meta-analytic procedures developed by Glass. Meta-analytic methods advocated by Hedges were also incorporated. Hembree conducted a similar meta-analysis in 1984. This study was an update to his work.

Data collection from all locatable studies from 1984 through June 2000 resulted in 307 effect magnitudes. Meta-analytical evaluation of mean effect sizes and their corresponding confidence intervals was conducted. Where appropriate, the validity of the mean effect sizes was assessed with fail-safe N values. With respect to skills acquisition, students using calculators either maintained or improved their operational and problem solving skills. Due to the minimal amount of available data, results regarding skills retention were not statistically significant. Analysis of the calculator's role in skills transfer was not possible due to insufficient data.

The following results were based on the inclusion of calculators in traditional mathematics instruction.

1. When calculators were used during testing, operational and problem solving skills of students in all grades and all ability levels realized significant improvement.

2. When calculators were not used during testing, paper-and-pencil skills of low ability students in all grades and average high school students improved. The operational skills of average students in grades K-8 and high ability students in all grades were neither helped nor hindered by calculator use. The problem solving skills of students in all grades improved after calculator involvement in mathematics instruction.

3. Students using calculators possessed better attitudes toward mathematics than their non-calculator counterparts.

The results of this study reveal students' operational and problem solving skills may improve and will not be hindered by calculator use in mathematics classes. Also, students may realize a significant improvement in their attitudes toward mathematics after using calculators. The benefits of calculator use should be most significant when students have access to calculators during testing as well as instruction.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I

Introduction

The latter half of the twentieth century featured a wealth of amazing technological innovations. Many have become significant elements of everyday life. In the last three decades, various electronic devices have been introduced to educational settings. Twenty-first century educators cannot imagine a classroom without technology. Resource documents published by the National Council of Teachers of Mathematics (NCTM) highlight the relationship between the mathematics classroom and technology. The Curriculum and Evaluation Standards for School Mathematics (NCTM, 1989) called for the availability of calculators and computers for students' use in problem solving situations. The successor document Principles and Standards for School Mathematics (NCTM, 2000) went a step further by describing technological devices as "essential tools for teaching, learning, and doing mathematics" (p. 24).

The hand-held calculator was the electronic device featured in this study. The calculator has experienced a metamorphosis during the last thirty years. In 1974, the NCTM encouraged the use of calculators in the classroom (NCTM, 1974). However, at nearly $50, the four-function device was still an expensive purchase. As a result, it could only be found in a select number of homes and very few classrooms. Within a few years a significant decline in price resulted in the presence of a calculator in virtually every home. The use of basic and scientific calculators was prevalent during the early 1980s. However, many mathematics teachers were still hesitant to incorporate them into their lessons on a regular basis. In particular, less than twenty percent of elementary teachers and less than thirty-six percent of high school teachers allowed students access to calculators during mathematics instruction (Suydam, 1982).

The 1990s witnessed another significant step in the evolution of the calculator. The graphing calculator, a hand-held device with the power of a small computer, was introduced to classrooms and homes. It had all the features of a scientific calculator plus many other capabilities. For example, a decade earlier, precise graphing techniques would have required a computer. The NCTM immediately recognized the impact the graphing calculator could have on secondary education (NCTM, 1989). In particular, the Standards document (NCTM, 1989) gave the graphing calculator credit for "the emergence of a new classroom dynamic in which teachers and students become natural partners in developing mathematical ideas and solving mathematical problems" (p. 128). During the later half of the decade, textbooks began to provide discussion of mathematical concepts, examples, and problem sets in which graphing technology was necessary.

During the last thirty years, calculators slowly made their mark on mathematics classrooms. They became influential in all aspects of the learning process from basic computations to the examination of abstract mathematical ideas. Calculators allowed students access to examples of equations that were impossible to generate by paper-and-pencil alone. As a result, more time was allotted for students to engage in conjecturing, conceptualizing, and modeling mathematical ideas (NCTM, 2000). Technology influenced the choice and placement of topics in the mathematics curriculum as well as methods of assessment (NCTM, 2000). At the end of the century, mathematics classrooms were vastly different from those that participated in the beginning of the calculator evolution. Especially at the high school level, graphing calculators were the rule, not the exception. With the encouragement of the NCTM, all types of calculators were being incorporated into students' explorations of mathematical concepts.

Throughout this metamorphosis, a plethora of studies have analyzed various aspects of the relationship between calculators and mathematics education. The effects on teachers, the attitudes of parents, and the methods of calculator use are just a few areas of focus. The literature contained over 120 studies which researched the most important issue in the relationship between calculators and classrooms – the effects of calculators on students. Technology had the potential for helping students understand mathematics on a deeper level. Knowing the actual effect of the calculator on students' understanding of mathematical concepts would allow educators and administrators to determine if the calculator's potential was being realized. Based on this premise, the current study was undertaken.

Problem Statement

A thorough, statistical examination of the effect of calculators on students in the K-12 mathematics classroom was the primary focus of this research. Hembree conducted a similar study in 1984. The current study was an update to Hembree's original research, his amendment with Dessart (Hembree & Dessart, 1992), and comparable research by Smith (1996). In particular, this document contains an examination and synthesis of results provided by a set of calculator-based research studies. All of the studies featured precollege mathematics students. Meta-analysis, a statistically sound process for integrating a collection of findings, was used to evaluate calculator effects on achievement and attitude.

Organization of the Study

This document is organized into six chapters. This chapter explains the rationale of the study. It contains the introduction, the study's purpose, the definition of relevant terms, and research questions. Chapter two is an examination of the traditional approaches involved in reviewing literature and contains a historical look at meta-analysis. The review of literature is found in chapter three. Reviews of calculator-based research and meta-analyses conducted in the field of mathematics education are the central theme of the chapter. The methods used to conduct the study are discussed in the fourth chapter and specific details of meta-analysis as they apply to this study are explained. These include the identification, collection, and coding of significant studies, as well as the calculation and evaluation of effect sizes. The methodology is based on the mathematical model of meta-analysis presented in Hembree's (1984) work. Chapter five presents the results of data analysis coupled with a discussion of its significance. The final chapter contains the researcher's conclusions and recommendations for future research.

Purpose of the Study

This study sought an answer to a complex question regarding the calculator's significance in the mathematics classroom. How does the calculator affect students' achievement and attitude in the study of precollege mathematics? This question was analyzed through a series of research questions listed in this chapter.

Definition of Terms

Classification of Calculators

Three types of calculators were discussed in the literature review and in the studies integrated by meta-analysis.

1. The basic calculator is the four-function or multi-function variety equipped with algebraic logic and an eight-digit display with floating decimal. It is the calculator most often found in elementary and junior high school classrooms.

2. The scientific calculator is capable of supplying the user with numerical evaluations of the basic functions, including trigonometric and

logarithmic functions, studied at the high school level. Most scientific calculators allow the user to perform operations with parenthetical grouping symbols.

3. The graphing calculator is the newest hand-held innovation to impact the mathematics classroom. With a 2½ by 1½ inch display screen, the user can investigate and compare mathematical concepts through graphic, symbolic, and numeric methods.

Educational Constructs

Since achievement and attitude were the two constructs under consideration in the studies gathered by the researcher, they were the focus of the meta-analysis. The organizational structure chosen to evaluate these constructs was the result of two important factors. First, since the work of Hembree (1984) was a precursor to this research and involved studies of the same constructs, Hembree's framework was a natural and obvious method of organization for the current study. Second, during initial analysis, the researcher recognized natural subdivisions in each study with regard to the constructs. The research questions established by the various authors followed these natural subdivisions. Through this analysis, it was determined that the natural breakdown of the constructs closely matched the Hembree framework. Therefore, the categories and subcategories of analysis in the current study were similar to those established and analyzed by Hembree (1984; Hembree & Dessart 1992).

The achievement construct referred to the acquisition, retention, and transfer of mathematical skills. Thus, mathematical skills were divided into two basic categories depending on how they were used within the studies. These categories also appeared in the writings of Hembree (1984; Hembree & Dessart 1992).

1. The operational category contained skills related to the solution of specific mathematical problems. Along with the general category of composite operational skills, the subcategories of computational skills and conceptual skills were analyzed separately.

2. The problem solving category was comprised of skills not implied by the mathematical problems at hand. Instead, these were skills students selected from their mathematical repertoire. The overall category of composite problem solving skills were evaluated as well as the subcategories of problem solving productivity skills and problem solving selectivity skills. The definitions of these subcategories were originally described by Hembree (1984). Productivity referred to the number of problems attempted by students. Selectivity considered the number of appropriate strategies used by students.

As in Hembree's (1984) meta-analysis, the attitude construct included the six attitudinal factors of the Mathematics Attitude Inventory developed by the Minnesota Research and Evaluation Center. The factors were attitude toward mathematics; anxiety toward mathematics; self-concept in mathematics; motivation to increase mathematical knowledge; attitude toward mathematics teachers; and perception of the value of mathematics in society. Several studies uncovered by the current researcher assessed students' attitude toward the use of calculators in mathematics. Therefore, this category was also included under the attitude construct.

Experimental Design of Integrated Studies

The studies followed a quasi-experimental treatment/control group design. "Quasi-experimental" referred to the use of intact classes as opposed to random sampling techniques to determine treatment and control groups. In all studies, two groups of students were taught by equivalent methods of mathematical instruction with one significant difference. The treatment group used calculators while the control group had no access to calculators. The effects of calculator use were measured by comparing the groups' responses to post-treatment evaluations. Standardized and teacher-designed achievement tests were the general means used to measure calculator effects on achievement.

1. Skills acquisition was measured immediately after treatment.
2. Skills retention was measured after a predetermined time lapse following treatment.
3. Skills transfer was measured by evaluating the ways students used the skills in other mathematical areas.

Standardized and teacher-designed survey instruments were the most common methods used to measure calculator effects on attitude.

Several studies involved a confounding variable – the creation and use of special curriculum materials for teaching mathematics with the calculator. These studies were conducted with the quasi-experimental design described above. However, the two groups differed in two significant respects – calculator use and method of instruction. Therefore, it was not appropriate to evaluate the data through meta-analysis. Descriptive statistics were generated and discussed.

A few studies reported gender-related differences in treatment and control groups. Female students who used calculators were compared to male students without access to calculators, and vice versa. These differences were not conducive to evaluation through meta-analysis. Therefore, a non-statistical analysis of the data was conducted.

The Calculator's Role in the Studies

The limits placed on students' use of calculators were also important in the design of this study. In particular, reading through various research reports revealed that some researchers allowed students to use calculators during testing while others did not. Hembree (1984) succinctly defined an extension effect to be the effect that resulted from the use of calculators during testing. When calculators were not used during testing, Hembree (1984) called the effect a maintenance effect. These terms adequately describe the two methods of calculator use significant to in the current study. Hence, calculator effects on achievement were evaluated in terms of maintenance and extension.

One aspect of the current study, which deviated from the framework established by Hembree, was the consideration of calculator effects for a specific type of technology. The graphing calculator is an innovation that was not available at the time of Hembree's (1984) original work. Nearly half of the studies involved in the current meta-analysis featured graphing calculators. Hence, the areas of skills acquisition and students attitude toward mathematics were analyzed in two ways.

1. All calculator types – A meta-analytical integration of all relevant studies involving basic, scientific, and graphing calculators.

2. Graphing calculator only – A meta-analytical integration of relevant graphing calculator studies.

The purpose of separate analyses was to determine if the effect of the graphing calculator was different than the effect of the calculator in general. The studies did not provide sufficient data to conduct two separate analyses regarding skills retention and transfer or other aspects of the attitude construct.

Research Questions

The following research questions were used to analyze calculator effects on achievement and attitude. They are similar to those established by Hembree (1984). In all cases, pre-college students in the mathematics classroom were under consideration. The achievement research questions (1 – 6) were analyzed in terms of maintenance and extension.

The <u>Achievement</u> Research Questions:

1. What are the effects of calculators on the acquisition of composite operational skills?
   a. What are the effects of calculators on the acquisition of computational skills?
   b. What are the effects of calculators on the acquisition of conceptual skills?
2. What are the effects of calculators on the acquisition of problem solving skills?
   a. What are the effects of calculators on the acquisition of problem solving productivity skills?
   b. What are the effects of calculators on the acquisition of problem solving selectivity skills?
3. What are the effects of calculators on the retention of operational skills?
4. What are the effects of calculators on the retention of problem solving skills?
5. What are the effects of calculators on the transfer of operational skills?
6. What are the effects of calculators on the transfer of problem solving skills?
7. What are the effects of calculators on students' estimation skills?

The Attitude Research Questions:

8. What are the effects of calculators on students' attitude toward mathematics?

9. What are the effects of calculators on students' attitude toward the use of the calculator in mathematics?

10. What are the effects of calculators on students' anxiety toward mathematics?

11. What are the effects of calculators on students' self-concept in mathematics?

12. What are the effects of calculators on students' motivation to learn mathematics?

13. What are the effects of calculators on students' attitude toward mathematics teachers?

14. What are the effects of calculators on student perception of the value of mathematics in society?

Research Questions Not Analyzed by Meta-Analysis:

15. Are the effects of calculators on achievement and attitude different for male and female students?

16. What are the effects of calculators on achievement and attitude when special curricula are involved?

Summary

The definitions provided in this chapter were essential elements in the researcher's evaluative process. Analysis of the sixteen research questions provided information necessary to answer the fundamental question of this study: How does the calculator affect students' achievement and attitude in the precollege mathematics classroom? Subsequent chapters provide the research-based foundation for this study, the methods of analysis used to evaluate the research questions, and the results of data analysis.

Chapter II

An Examination of Research Review Techniques

The literature review is an important component of educational research endeavors. Through the medium of scholarly writing, academicians are expected to review previously conducted research as a precursor to the presentation of their own ideas and findings. A chapter of each doctoral dissertation is devoted to a review of relevant literature. In the current study, chapter three is dedicated to this important task. Most published studies provide a historical explanation of the topic under consideration before launching into the researcher's current findings. Furthermore, literature reviews are significant as scholarly activities in their own right. Light and Pillemer (1984) emphasize this point through the following statement: "For science to be cumulative, an intermediate step between past and future research is necessary: synthesis of existing evidence" (p. 3). Therefore, the literature review should not be taken lightly. Since the current study is a "synthesis of existing evidence" regarding the use of calculators in the K-12 classroom, an examination of research review techniques is appropriate at this juncture.

Methods of Review

There are a variety of quantitative methods available for reviewing academic research. The approaches discussed here are both precursors and contemporaries of meta-analysis – the method of review at the heart of this study. This discussion concerns quantitative techniques of review most frequently applied to the field of education with a particular focus on mathematics education. Subsequent pages contain a discussion of four particular methods:

1. the narrative approach
2. the vote-counting method
3. the method of combining p-values
4. meta-analysis

Strengths and weaknesses of each approach are described. The discussion of meta-analysis is preceded by a consideration of the reasons leading to the development of this statistical method of review. All of the approaches have an interesting history and the first three are not obsolete. However, meta-analysis satisfies needs not met by other methods of review.

The Narrative Approach

Providing narrative descriptions of research findings is one of the oldest summary procedures and is still popular today. It requires the reviewer to supply brief descriptions of studies conducted on a specific topic. The results are displayed at face value in the same manner they were initially presented by the original researcher (Hunter, Schmidt, & Jackson, 1982). In some cases, the reviewer attempts to find a theory encompassing all of the research results in an overarching conclusion. There is no statistical analysis involved in this process.

With only a few studies on which to report, the narrative approach can be quite manageable. In fact, a theory may exist which integrates the findings and takes the conflicting elements into consideration. However, when utilizing the narrative approach on a large collection of studies, there is much more room for disagreement. The subjects being covered, as well as the design and measurement techniques used by the studies can be called into question (Hunter et al., 1982). This often makes the process more difficult for the reviewer and results in findings that are cumbersome and tedious for the reader to interpret.

When a review includes a large collection of reports, processing all of the information provided is a difficult task. As a result, the researcher generally selects one of three possible alternatives (Hunter et al., 1982):

1. The reviewer may summarize the studies without integrating the results into some overarching theory.

2. The reviewer may provide descriptions of all of the studies involved but base his theoretical conclusions on only a select few.

3. The researcher may attempt to generate a comprehensive theory including all of the studies. This typically results in a theory that does not accurately represent each study's conclusions.

The Vote-counting Method

The vote-counting method was initially developed to help with the information-processing aspects of reviewing (Hunter et al., 1982). Vote-counting involves the reviewer placing a study into one of three categories based on the statistical significance of its outcomes. The selection of the appropriate category depends on the relationship between the dependent and independent variables. The reviewer must determine if the relationship is significantly positive, significantly negative, or statistically insignificant. The reviewer's conclusions are based on the category containing the most studies. Vote-counting is quite simple and, as with the narrative approach, may be sufficient for reviewing a small number of studies. However, if the review is more complex or involves a large number of studies, the process exhibits three crucial flaws.

First, the method may result in an overall conclusion that the studies have no significant outcomes while some important positive and negative effects are being ignored (Light & Pillimer, 1984). Consider the scenario in which the number of studies with significant positive effects is relatively close to the number of studies with significant negative effects. However, the number of studies with no significant effect exceeds them both. In this case, the vote-counting method will result in a conclusion of no effect. As a result, the large number of significant effects has been ignored in the process.

Second, the vote-counting method is not effective with studies characterized by small sample sizes and small effect sizes. Many educational research endeavors with these qualities produce interesting findings (Hedges & Olkin, 1985). However, the vote-counting method will not adequately explain the significance of these studies. Hedges and Olkin (1985) proved that when a true effect exists and the mean statistical power of the studies is less than ½, vote-counting is not an appropriate method of review. Under these conditions, as the number of reviewed studies increases, the probability that vote-counting will yield accurate conclusions decreases (Hedges & Olkin, 1985). Therefore, other reviewing techniques would be more appropriate for this type of situation.

Third, the category containing the most studies does not always adequately describe the magnitude of an effect (Light & Pillimer, 1984). Consider the scenario in which studies are separated into three categories. One category may contain more studies than either of the other two categories, but the vote-counting method can not explain whether the results are overwhelmingly in favor of a particular treatment or barely significant. This flaw is similar to another difficulty regarding sample sizes (Light & Pillimer, 1984). The statistical significance of research results is greatly influenced by sample size. However, "reviewers using the voting method treat all studies alike and completely ignore the fact that studies with different sample sizes have a completely different meaning for 'significant'" (Hunter et al., 1982, p. 132). In other words, the vote-counting method is incapable of reporting these subtle, but important, details.

Combining P-values

This method requires the researcher to combine significance levels across all studies in order to produce a p-value representing the entire group (Hunter et al., 1982). If the p-value is small enough, the reviewer can report the existence of a significant effect. Rosenthal (1978) was an early advocate of this approach and his research involved nine methods by which a pooled p-value could be generated. His work in this area is self-described as a precursor to meta-analysis (Rosenthal, 1991). This method is more powerful than the two previously mentioned approaches because the results of individual studies are combined in a statistical manner. In particular, the method of combining p-values can yield significant results the vote-counting procedure would be unable to distinguish (Light & Pillimer, 1984). When a reviewer incorporates several studies with small, possibly insignificant, p-values, it is possible for the pooled p-value to describe a statistically significant effect for the combination of studies (Rosenthal, 1991).

However, this method also has its faults. For example, while many combinations of studies may result in a significant pooled p-value, the magnitude of the effect may not be represented by the value (Hunter et al., 1982). Two criticisms of this approach were described in the writings of Light and Pillimer (1984). First, a pooled p-value is not able to explain the distribution of study results. Therefore, even if a significant pooled p-value is generated, it may represent one largely significant study outweighing several other statistically insignificant studies. The second criticism refers to the type of studies deemed fit for publication. Most published studies contain statistically conclusive results. Therefore, studies unable to generate significant findings are most likely underrepresented in the expanse of research available in print. Therefore, a decisive pooled p-value may be the result of publication bias (Light & Pillimer, 1984).

The Need for a New Approach

While all of these techniques are useful methods of review, meta-analysis has been slowly making its mark on the world of research. The need for another form of statistical analysis was realized over sixty years ago. Rosenthal's (1991) early work evolved into a form of analysis featuring p-values and effect sizes. Through the work of Rosenthal and other researchers in the 1970s, the process of meta-analysis began to take shape. What follows is a discussion of significant features of meta-analysis.

First, a research-based investigation of the need for a statistical method of integrating research is considered. A comparison of meta-analytic procedures and traditional forms of review was conducted by Cooper and Rosenthal (1980). The study produced some interesting results. Forty-one graduate students and faculty members were asked to conduct a review of literature. They reviewed seven studies containing statistically significant findings on the relationship between gender differences and task persistence. Prior to their requests for literature reviews, Cooper and Rosenthal (1980) knew the results of the seven studies revealed females were more task persistent than their male counterparts.

The participants were randomly assigned to conduct either a meta-analytic review or a review by more traditional methods. Cooper and Rosenthal (1980) found 73% of the participants using traditional methods of review were unable to determine a significant relationship between gender and task persistence. Only 32% of the meta-analytic reviewers reached a similar conclusion. This is a clear example of why educational research needs a statistically rigorous method of review. It is also an example of the ability of meta-analysis to satisfy that need.

Meta-Analysis

"Meta-analysis is the quantitative cumulation and analysis of descriptive statistics across studies" (Hunter et al., 1982, p. 137). The fundamental steps in the process are gathering studies relevant to the topic, extracting quantifiable information from the studies, and organizing the information into an overall conclusion (Hunter et al., 1982). While this method does not require access to the original data, meta-analysis is a statistically sound procedure for research integration (Glass, McGaw, & Smith, 1981). The basic concepts involved in this process were first implemented by Thorndike and Ghiselli over sixty years ago (Hunter et al., 1982). Their work used average correlations to integrate the results from a group of studies. The early work of Rosenthal (1991) also involved some basic meta-analytic procedures. However, the credit for combining the essential processes involved in the methodology and first coining the term "meta-analysis" belongs to Glass (Hunter et al., 1982).

While Glass is credited with the pioneer efforts in this field, three other academicians were involved in the early stages of research falling under the umbrella of meta-analysis. As mentioned previously, Rosenthal's (1991) initial work contained elements of meta-analysis. He continued to develop his procedures and eventually produced a form of meta-analysis similar to the method advocated by Glass. Schmidt and Hunter were two other researchers involved in the development of meta-analysis (Hunter et al., 1982). Their work also generated the fundamental meta-analytic principles independently of Glass. Today, the work of Schmidt and Hunter is largely considered an extension of Glassian meta-analysis. In particular, their additional procedures handle problems like sampling error and unreliability (Hunter et al., 1982).

In the last twenty years, Hedges has become an invaluable source of updates to the methodology and theory encompassed by the term "meta-analysis". Statistical Methods for Meta-Analysis by Hedges and Olkin (1985) provides statistical justification and expansion of the ideas originally formulated by Glass, Hunter, and Schmidt. The Practical Guide to Modern Methods of Meta-Analysis (Hedges, Shymansky, & Woodworth, 1989) is an excellent source of examples of ways the fundamental meta-analytic procedures can be used. In 1984, Hembree produced a model for meta-analysis and an example of the precedures necessary to synthesize research in education. His model is a precursor to the Hedges guide published in 1989. Glass' basic characteristics of meta-analysis are outlined below, followed by an explanation of the Schmidt-Hunter extensions of Glass' work. Discussion of the "file drawer" problem and a description of the Hembree model for meta-analysis will conclude this chapter.

The Glass Method

Meta-analysis is not primary analysis – the analysis of original data in a research study (Glass et al., 1981). It is also not secondary analysis – the reanalysis of an original research question with different statistical techniques or the consideration of new questions with the old data. Instead, meta-analysis is the process of analyzing data provided by quantitative explanations of research studies with descriptive statistical techniques (Glass et al., 1981). There are several basic steps to meta-analysis:

1. gathering the data;
2. organizing the properties of the studies involved;
3. organizing the findings generated by the studies;
4. assessing the results.

Each of these is described below.

<u>Gathering the Data</u>

The researcher involved in conducting a meta-analysis gathers a set of studies on a topic in which he has a particular interest. These studies may or may not address the same research questions. Actually, in most cases, the studies will not contain results on all the same questions. This has been a criticism from some who feel meta-analysis "mixes apples and oranges" (Glass et al., 1981, p. 22). However, if all of the studies were similar in every respect, they would simply produce the same results except for the usual statistical error

(Glass, 1978). There would be no reason to integrate the findings. Therefore, the use of studies involving different research questions is a crucial component of meta-analysis.

Obtaining pertinent studies requires a thorough search of the available literature on the research topic. The researcher must make every effort to uncover every possible research study relevant to the subject at hand. Even though the meta-analysis itself only features primary sources, the search should also include secondary literature sources (Glass et al., 1981). Primary sources include journal articles, doctoral dissertations, and master's theses as well as papers and reports created for scholarly meetings. Searching the bibliographies of primary sources is a good method of locating other possible studies. Secondary sources are those that organize and review the data generated by primary sources. These can be found in a variety of abstract archives, and many are published in journals. For reviewers interested in conducting meta-analyses in subjects within the field of mathematics education, some valuable secondary sources include: Educational Resources Information Center (ERIC), Review of Educational Research, Dissertation Abstracts International, and the July issues of the Journal for Research in Mathematics Education (JRME) for the years 1970 through 1994.

A thorough, exhaustive search for primary sources is an essential step in the integration of research (Glass et al., 1981). "Locating studies is the stage at which the most serious form of bias enters a meta-analysis" (Glass et al., 1981,

p. 57). It is important that the studies integrated by the review represent most of the data available on the question at hand including unpublished research. If the data is not representative of everything available, the results are questionable at best. Therefore, the reviewer should provide a complete description of the location procedures used in the gathering of studies (Glass et al., 1981). This will allow the reader to make a knowledgeable and informed assessment of the meta-analysis and its relevance to his needs.

Organizing the Properties of a Study

Once the primary studies are located, the immense task of organizing the data begins. This includes quantifying both the findings and characteristics of the studies (Glass et al., 1981). The researcher must determine quantitative labels for study properties. This is crucial to evaluating the relationship between the characteristics of a study and its results. However, establishing the appropriate definitions and finding ways to quantify them require much thought and consideration (Glass et al., 1981). In particular, the codes chosen by the reviewer are a direct result of the amount and type of information reported in the primary study. In some cases, the reviewer might need to request missing information or explanation of unclear information from the author of the primary resource.

Coding the characteristics of the studies in measurable terms involves a variety of statistical classifications spanning the scale of the nominal through ratio categories. Many of the characteristics may be represented with the typical measurement scales. Examples of items relatively easy to code are number of students, length of treatment, student grade level, and years of teaching experience. Other characteristics require the use of "indicator variables" (Glass, 1978, p. 365) for accurate coding. The socio-economic status of students and the type of treatment used are two examples of characteristics which are non-ordinal, but nonetheless may be significant during data analysis.

The characteristics important to meta-analysis fall into two categories: substantive and methodological (Glass et al., 1981). The properties fundamental to the specific problem or treatment under consideration are called substantive. Methodological characteristics are those directly resulting from the method of the primary study: experimental, correlational, survey, and the like. Regardless of which form the properties assume, the purpose in coding them is to allow the researcher easy access to the characteristics during data analysis. This is important since one aspect of meta-analysis is to determine the statistical similarities of research results for different study characteristics (Glass et al., 1981).

Organizing the Findings of a Study

If every research study expressed its results with similar statistical calculations, then the process of quantifying the findings would be quite simple. However, this is obviously not the case in most situations. Just as research studies ask different questions, they also use different statistical methods to arrive at their conclusions. Therefore, a variety of statistical methods are available for integrating the findings of research studies (Glass et al., 1981; Hedges & Olkin, 1985). In general, the meta-analysis involves the results of experimental and correlational studies. Effect sizes are generated for experimental studies and correlation coefficients are the values gleaned from correlational studies. The basic components for calculating and evaluating effect sizes and the fundamental process used for analyzing correlational studies are discussed below. However, the extensive details and statistical formulas necessary for more difficult cases of meta-analytical integration are not described here. Instead, the reader is invited to peruse Meta-Analysis in Social Research by Glass, McGaw, and Smith (1981) or Statistical Methods for Meta-Analysis by Hedges and Olkin (1985).

Calculations for Experimental Studies

The results of experimental studies in which a treatment and control group or two treatment groups are compared are best described with standardized mean differences, or effect sizes (ES), between pairs of treatment conditions (Glass et al., 1981). In particular, "the most informative and straightforward measure of experimental effect size is the mean difference between experimental and control groups divided by within-group standard deviation" (Glass et al., 1981, p. 102):

$$ES = \frac{\overline{X}_E - \overline{X}_C}{s_x} \tag{1}$$

The meaning of effect size (ES) is fairly easy to understand. In many cases, an ES can be interpreted through careful consideration of the study's properties (Glass et al., 1981). For example, consider the scenario in which a treatment group is compared to a control group based on posttest scores and an ES of 1.00 is calculated. This means the average test score of students receiving treatment is one standard deviation above the average test score of students who did not participate in treatment. In general, this value is fairly easy to understand. However, sometimes it may be necessary to consider the known effects of a treatment condition in order to comprehend the magnitude of an effect size (Glass et al., 1981). Consider a review of research in which one of two different treatments is compared to traditional instruction in a series of experimental studies. If the average effect size of treatment A is twice as large as the average effect size of treatment B, the benefits of treatment A outweigh the benefits of treatment B.

In equation (1) there are three possibilities for s_x. Glass (Glass et al., 1981) prefers the use of the control group standard deviation since this "at least has the advantage of assigning equal effect sizes to equal means" (p. 107). Hedges and Olkin (1985) provide two other possible values for s_x. They state the standard deviation of the experimental group can be used. In fact, Glass makes this suggestion as well. However, Hedges and Olkin's (1985) strongest support is for a pooled standard deviation combining the standard deviations from both the treatment and control groups:

$$s_x = \sqrt{\frac{(n_E - 1)s_E^2 + (n_C - 1)s_C^2}{n_E + n_C - 2}} \qquad (2)$$

where n_E and n_C are the respective experimental and control group sample sizes and s_E and s_C are the respective experimental and control group standard deviations. Glass (Glass et al., 1981) does not support this recommendation. When several treatment groups are being compared to a control group, Glass contends it would be possible for different effect sizes to result from identical mean differences. Hedges and Olkin (1985) assert a pooled standard deviation is the best possibility for s_x. They claim Glass' argument is not relevant since, in most cases, it is safe to assume population variances are equal.

Hedges and Olkin (1985) go one step further and provide a correction factor for the calculated ES. Hedges (1981) proved that Glass' original ES estimate has a bias based on the number of degrees of freedom for s_x. So each ES should be multiplied by

$$J(m) = 1 - \frac{3}{4m - 1} \qquad (3)$$

where m = $n_E + n_C - 2$. This formula is an excellent approximation of the actual bias correction values provided by Hedges and Olkin (1985, p. 80). The bias is small and even inconsequential for large degrees of freedom but can significantly inflate ES for small degrees of freedom. For example, if a study contains a combined sample of 21 students, then the correction factor would be J(19) = 0.96, so the unbiased ES estimate is 4% smaller than the original estimate (Hedges et al., 1989). Hedges and Olkin (1985) advocate every ES should be corrected for bias before further calculations are conducted.

A novice utilizing meta-analysis will quickly find most studies provide statistical values other than the basic means and standard deviations needed for equations (1) and (2). Hedges and Olkin (1985; Hedges et al., 1989) are excellent references for the appropriate formulas to convert statistical data like t-values or information from ANOVA tables into effect sizes.

Calculations for Correlational Studies

For correlational studies, the researcher integrates the correlation coefficients that describe the relationship between two variables. Hedges and Olkin (1985) assert the correlation coefficient is a good candidate for explaining the magnitude of an effect. Their explanation is that correlation coefficients are "invariant under substitution of different but linearly equatable measures of the same construct", (Hedges & Olkin, 1985, p. 223). Working with correlation coefficients is not as complicated as the process of generating effect sizes from experimental studies.

Glass (Glass et al., 1981) states integration can be conducted on any of the correlation scales: r_{xy}, r_{xy}^2, or Fisher's Z transformation of r_{xy}. However, the results of the analysis should be stated in terms of the traditional Pearson's product-moment, r_{xy}, scale. Glass provides specific guidelines for conversion when findings are presented in forms other than r_{xy}. Once the coefficients are all represented on the same metric, comparison follows a similar pattern as with effect sizes from experimental studies. The sizes of correlation coefficients are evaluated in terms of the study's properties that represent the relationship described by r_{xy}.

Before proceeding with the assessment of results, this is an appropriate juncture to emphasize the need for meta-analysis to include dissertations and other unpublished research. Creating effect sizes for all studies, published and unpublished, and using quantitative measures to analyze the data allows the researcher to statistically determine whether or not published studies are more rigorously designed than unpublished studies (Glass et al., 1981). This is also an opportunity for the researcher to examine how the strength of a research design influences the size of the effect. Therefore, the use of unpublished research in a meta-analysis is essential.

Assessing the Results

The process of meta-analysis seeks to statistically integrate a set of studies related to a particular topic. For this to be possible, the results of studies must be stated in terms of a common measurement scale. The dependent variables in statistical analysis are the research results while the independent variables are the aforementioned substantive and methodological study characteristics (Glass et al., 1981). The methods used to transform study information into analyzable data have been discussed. Next, the methods for reporting the findings will be examined. In general, all traditional methods for analyzing statistical data can be useful in a meta-analysis. All forms of descriptive statistics can be used to illustrate effect sizes from experimental studies and correlation coefficients. However, the particular methods advocated by Glass (Glass et al., 1981) and Hedges and Olkin (1985) are discussed here.

Graphical Analysis

For visual representation, Glass (Glass et al., 1981) recommends a method originally described by Tukey (1977) in Exploratory Data Analysis. The "schematic box-and-whisker plot" provides a picture of the effect size distribution for different constructs (Glass et al., 1981). An example appears in Figure 1. The main box represents the second and third quartiles of the distribution. The median, defined in the traditional way, lies within the box. The inner fences are positioned on each side of the box exactly 1.5 times the length of the box. They are marked "f". Data values beyond the inner fences are called outliers. Outer

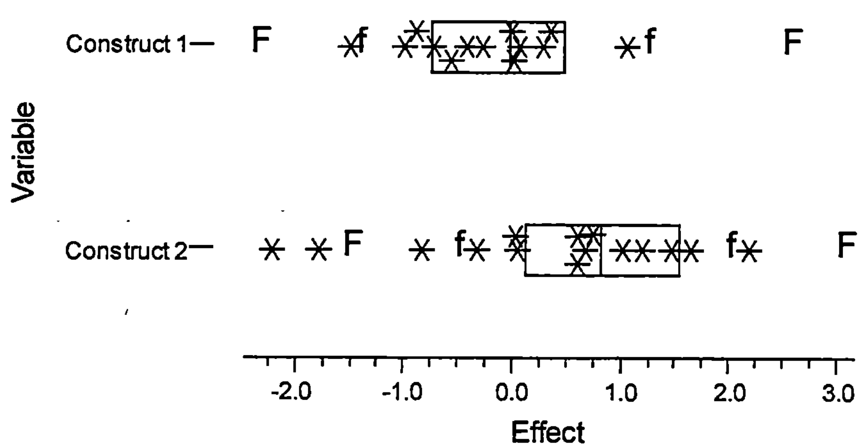


Figure 1. Example of a Schematic Box-and-Whisker Plot for Effect Sizes Corresponding to Two Constructs

fences, marked "F", are another distance of 1.5 times the length of the box beyond the inner fences. Far outliers lie beyond the outer fences.

When reading a schematic box-and-whisker plot, the effect sizes falling in the outlier or far outlier categories are considered with skepticism. They may represent errors in data reporting or some type of miscalculation. When presenting the findings of meta-analysis, these problems should be reported or the outliers should be eliminated from the study and the descriptive statistics recalculated (Glass et al., 1981).

Descriptive Analysis

Summary tables of average effect sizes and their standard deviations are frequently used to record the results of meta-analysis. Similar to the interpretation of data from primary research studies, the meta-analytic researcher must be careful when describing average (mean and median) effect sizes with broad, general categories. The effects of other study characteristics must be reflected in the description in order for the interpretation to be as accurate as possible (Glass et al., 1981).

Glass et al. (1981) assert the integration of a collection of studies is more a descriptive process than an inferential one. Therefore, effect size means and standard deviations are the central focus of Glassian meta-analysis. Effect sizes are grouped according to the research questions at hand. Each value is considered an estimate of the population parameter featured by that group of effect sizes (Glass et al., 1981). In order for the mean value to have statistical significance, the effect sizes must satisfy the condition of homogeneity. Homogeneous effect sizes have one significant element in common – the population parameter in question (Hedges & Olkin, 1985). Therefore, if the group is homogeneous, the mean effect size is determined to be the best estimate of the population parameter. Analysis of the mean becomes the researcher's central focus (Hembree, 1984).

If the values are heterogeneous, the researcher's attention turns to finding reasons for the variability among the effect sizes. The search for the source of the variance begins with the characteristics discovered and classified during the coding phase of meta-analysis. Therefore, the astute meta-analyst must take great care in the coding process and have easy access to the characteristics during all phases of integration.

Analysis of Correlation Coefficients

For correlation coefficients, Glass (Glass et al., 1981) asserts all statistical methods of representing relationships between two variables may be useful in meta-analysis including contingency tables and regression analysis. However, Glass (Glass et al., 1981) states that the most powerful method involves the generation of linear representations on the Pearson product-moment scale. Many significant relationships between study characteristics and findings can be analyzed by evaluating effect magnitudes that are represented by linear correlations.

Glass (Glass et al., 1981) also discusses the use of regression equations in the interpretation of meta-analytic findings from correlation coefficients. Regression equations allow the researcher to calculate estimated effect sizes for specific characteristics or treatment conditions. By setting the independent variables to a specific range of values, the researcher can draw conclusions about certain study elements under different treatment conditions. However, care must be taken in reporting the findings, since one treatment may be superior under certain conditions while another treatment may result in greater benefits under another set of conditions.

Hedges and Olkin (1985) caution against the use of regression analysis for evaluating correlation coefficients. In particular, they state the underlying assumptions necessary for regression analysis are not met since "the variance of a sample correlation is inversely proportional to the sample size of the study", (Hedges & Olkin, 1985, p. 224). Also, a dependent relationship between the variance of the sample correlation and the population correlation exists. Therefore, problems may arise when integrated studies have different sample sizes and different population correlations. Lastly, Hedges and Olkin (1985) state the traditional method of regression analysis does not provide a test of goodness of fit for the regression model.

Other Assessment Techniques

In the assessment and analysis of results, there are many techniques available. Determining the appropriate method depends on the type of data collected and the types of studies involved. Glass (Glass et al., 1981) as well as Hedges and Olkin (1985) describe various methods for analyzing effect sizes and correlation coefficients. These include linear analysis of variance models, multiple linear regression models, logarithmic models, and non-parametric integration models. The reader is invited to examine Statistical Methods for Meta-Analysis by Hedges and Olkin (1985) and Meta-Analysis in Social Research by Glass, McGaw, and Smith (1981) for further details.

The Hunter-Schmidt Approach

As previously mentioned, Hunter and Schmidt (Hunter et al., 1982) and Glass et al. developed their meta-analytic procedures concurrently. However, the researchers worked independently. Many of the concepts developed by Hunter and Schmidt directly correlate with those already covered by the Glass method of meta-analysis. The Hunter and Schmidt method also features solutions to problems involving sampling error, unreliability, and range restrictions (Hunter et al., 1982). These aspects of the Hunter-Schmidt approach are the major differences between the two techniques.

As in the Glass form of meta-analysis, Hunter and Schmidt emphasize the use of effect sizes over p-values. However, Hunter and Schmidt developed formulas for calculating experimental effect sizes after they established techniques for evaluating correlation coefficients (Hunter et al., 1982). The Hunter-Schmidt method defines the numerator of the effect size to be the difference between experimental and control group means.

$$d = \frac{\overline{Y}_E - \overline{Y}_C}{s} \tag{4}$$

Where s is a within-group standard deviation resulting from the pooled variance of experimental and control group data. This equation is similar to equation (1) advocated by Glass.

One difference between the two methods is that the Hunter-Schmidt procedure requires corrections to the mean effect size for problems in instrument unreliability and for restrictions in the range (Hunter et al., 1982). Glass does not discuss the need for these types of corrections. Another significant difference is the Hunter-Schmidt method does not accept the variance of the effect size at face value (Hunter et al., 1982). It tests the variance for statistical flaws including sampling error, range restriction, reliability or validity issues, or computational errors. If necessary, Hunter and Schmidt have developed methods for adjusting, the effect size variance with regard to the first three statistical artifacts mentioned above. The reader is invited to review Meta-Analysis: Cumulating Research Findings Across Studies by Hunter, Schmidt, and Jackson (1982) for more details.

The "File Drawer" Problem

Publication bias has the potential of significantly influencing the results of all types of research reviews. However, since a review can only be based on available research studies, the actual influence of unpublished results is difficult to determine. Nevertheless, the potential effect of unpublished results should be carefully considered by the reviewer. This issue was addressed earlier in this chapter during discussion of the combining p-values method of review. Now the effects of publication bias on the results of a meta-analysis will be described. Rosenthal (1979) conducted extensive research and coined the term "file drawer problem" to describe the situation. The extreme view of the problem is "journals are filled with the 5% of the studies that show Type I errors, while the file drawers back at the lab are filled with the 95% of the studies that show nonsignificant (e.g. $p > .05$) results" (Rosenthal, 1979, p. 638).

With respect to meta-analysis, the reviewer would like an estimate of the number of unpublished, nonsignificant studies that could change significant results into nonsignificant results. In particular, the researcher would like to know the number of file drawer studies that could be incorporated in the calculation of a mean effect size before the significance of the mean would be negatively affected by the null results. If a statistically significant mean effect size fails to be significant after the addition of only a few file drawer studies, the results are not resistant to the "file drawer problem" and should be written up accordingly.

With careful searching techniques, Glass (Glass et al., 1981) asserts the astute meta-analyst can guard against the problem. Two methods are:

1. Requesting unpublished manuscripts from researchers in the field.

2. Conducting extensive searches for dissertations and theses on the topic.

Glass reports that research published in journals is biased toward the researcher's chosen hypotheses. In a comparison of eleven meta-analyses, Glass et al. (1981) determined the findings reported in journals were "one-third standard deviation more disposed toward the favored hypotheses of the investigators than findings reported in theses or dissertations" (p. 67). Therefore, the researcher who fails to include unpublished studies in a meta-analysis may generate misleading results.

Rosenthal (1979, 1991) developed a statistical method, based on probability levels, for determining a quantitative review's tolerance for null results. At the 5% significance level, the number of file drawer studies necessary to cause the probability of a type I error to reach $p = 0.05$ is found by the formula:

$$1.645 = \frac{K\overline{Z}}{\sqrt{K+X}} \qquad (5)$$

where X is the number of file drawer studies, K is the number of studies integrated by the meta-analysis, $\overline{Z}$ is the mean value calculated with standard normal Z values corresponding to the exact p-values from the meta-analysis studies, and 1.645 is the value corresponding to a 5% level of significance for the standard normal distribution. An algebraic simplification of (5) provides the following formula for X.

$$X = \frac{K[K\overline{Z}^2 - 2.706]}{2.706} \qquad (6)$$

Similar formulas can be easily generated for other significance levels. Simply replace 1.645 in formula (5) with the appropriate standard normal distribution value corresponding to the desired level of significance. Likewise, solving for X will yield a formula similar to (6). The size of X describes the magnitude of a meta-analysis' resistance to the file drawer problem. If X is relatively small, the results will be significantly influenced by a small number of file drawer studies.

For a 5% level of significance, Rosenthal (1979, 1991) describes a less rigorous approach for determining the number of file drawer studies needed to influence statistically significant results. If the exact p-values are not available, then the researcher can base an estimate of the number of file drawer studies on S, the number of meta-analysis studies producing statistically significant results and N, the number of meta-analysis studies producing nonsignificant results.

$$X = 19S - N \tag{7}$$

When $p = 0.05$ and the null hypothesis is true, 19 is the ratio of the expected number of nonsignificant results to the expected number of significant results.

Orwin (1983) extended Rosenthal's work to include effect sizes. His rationale for the modification is that exact probability levels are not always reported in primary studies. Hedges and Olkin (1985) support Orwin's method, called the fail-safe N. While the premise is the same as that of Rosenthal's original work, advocates of the fail-safe N state that null results, either new or unpublished, from a variety of sources are the central issue (Brown, 1992; Carson, Schriesheim, Kinicki, 1990). The researcher must assess the stability of the meta-analytic results to the addition of null findings from all possible sources (Carson at al., 1990).

The fail-safe N is the number of studies with null results that would reduce a mean effect size to a negligible level (Hedges & Olkin, 1985). The formula for the fail-safe N is:

$$X = \frac{K(\overline{D} - D_C)}{D_C} \quad (8)$$

where X is the fail-safe N, K is the number of studies integrated by the meta-analysis, $\overline{D}$ is the mean effect size of the meta-analysis studies, and D_C is a criterion value chosen for analysis (Carson et al., 1990). Hedges and Olkin (1985) state D_C is the effect size value which $\overline{D}$ would be reduced to by including X number of studies with null results in the meta-analysis. Therefore, the selection of D_C is crucial to the fail-safe N process. The relationship between the fail-safe N and the mean effect size should be reported with the meta-analysis results. "If the fail-safe N (X) is relatively small in comparison to the number of studies in the meta-analysis (K), then only tenuous conclusions should be drawn, regardless of the magnitude of the effect size" (Carson et al., 1990, p. 239).

It is often difficult to make decisions regarding the meaning of the size of X. The researcher needs to determine whether or not X unpublished studies could actually exist (Rosenthal, 1979, 1991). To make an educated determination, the researcher must have some understanding of the amount of research conducted in the field under consideration. No firm guidelines on the relationship between X and K have been established (Carson et al., 1990). Rosenthal (1979) suggests that a conservative "tolerance level" for X is 5K + 10

(p. 640). This allows for the availability of five times the number of studies involved in the review. If only one study is being reviewed, the +10 sets the minimum number of "file drawer" studies at fifteen.

The Hembree Model

A dissertation written by Hembree in 1984 is an invaluable resource for the researcher conducting an education-based meta-analysis. In first half of his thesis, Hembree provides a description of the processes involved in the integration of studies through a meta-analysis. Hembree's (1984) model is solidly grounded in the work of Glass et al. (1981), Hunter, Schmidt, and Jackson (1982), and Hedges and Olkin (1985). While Hembree's work is a guide and some specific procedures are discussed, Hembree is quick to state "the model is presented less to fix procedures than to describe in one source a statistically sound and orderly process from inception to conclusion" (1984, p. 40). He clearly wants researchers to realize that flexibility is an important characteristic of meta-analysis. Therefore, he does not provide rigorous, specific steps that would be counterproductive to the process as originally defined by Glass.

Hembree's (1984) model outlines the entire meta-analytic procedure from the selection of a topic and establishing research questions to the coding of data and writing the results. He guides the researcher through the process of gathering studies, sorting through the relevant and the irrelevant, and finally selecting the studies necessary for statistical integration. In his tips for coding variables, Hembree (1984) discusses several important concepts. He includes the types of variables imperative for any successful meta-analysis, the difficult but essential process of coding the characteristics of research design, and the importance of a clearly organized coding scheme for accurate data collection.

In his discussion of effect sizes, Hembree (1984) provides the simplest equations which are both practical for the researcher and statistically sound. He discusses the merits of the various approaches provided by Glass, Hunter-Schmidt, and Hedges and Olkin, but endeavors to keep the general process as simple as possible without compromising the statistical integrity. He provides a specific list of assumptions necessary for accurate treatment of effect sizes. In his discussion of data analysis, Hembree (1984) furnishes an easy-to-read flow chart of the general procedures necessary to turn coded data and effect magnitudes into the appropriate confidence intervals. He gives general tips on how to analyze and explain the relationships between the various study characteristics. Hembree's (1984) details on the interpretation of confidence intervals and their relationship to the research questions are clear, concise, and helpful to the novice or experienced researcher interested in meta-analysis.

Summary

While the narrative approach, the vote-counting method, the method of combining p-values, and other review procedures will always be important to scholarly activities, meta-analysis has had a significant impact on the academic world. Glass, Hedges, Olkin, Hunter, Schmidt and many other statistical scholars laid the foundation for the meta-analytic procedures currently in use today. Examples of meta-analysis can be found in a variety of fields. Researchers in the medical community provide their colleagues with significant findings based on the integration of research regarding specific medical treatments and conditions. Perusal of the Internet with any of the common search engines can quickly point the user to several web sites in which company's the central focus is conducting meta-analyses for their customers. The process is continuing to gain statistical integrity. While the meta-analysis will continue to evolve as current and future researchers refine the integration procedures, there is no doubt that it will persist as a significant element of the statistical landscape. The next chapter highlights examples of meta-analysis found in the field of mathematics education.

Chapter III

Review of Related Literature

Over the last century, pedagogical methods in mathematics have been in a gradual yet constant state of change. Since its formation eighty years ago, the National Council of Teachers of Mathematics (NCTM) has been instrumental in initiating change in mathematics classrooms all over this country. In particular, the Curriculum and Evaluation Standards for School Mathematics (1989) was a strong influence on the reform efforts of the last decade. No doubt the successor publication Principles and Standards for School Mathematics (2000) will motivate reform efforts during the 21st century. Educators, students, and researchers are the central participants in every educational movement intent on change. Some changes are well known – like the "back to basics" movement of the 1970s. Other changes, while still influential, are small and not well publicized. Teachers who institute new techniques in teaching mathematical concepts initiate small changes in their students on a daily basis.

One purpose of reviewing literature is to document the significant changes which cover a specific period of time or a particular topic. The central focus of the current study was the calculator's role in changes regarding achievement and attitude in precollege mathematics students. Since the calculator's introduction in the early 1970s, the changes in which it has played a part have been well documented by researchers and educators. The process the current study used to synthesize the most recent calculator initiated changes was meta-analysis. Therefore, the central focus of this literature review is twofold – previous reviews of calculator research and examples of meta-analysis as an evaluation tool in mathematics education.

This chapter begins with discussion of a research review of historical significance. It is a narrative summary of a large number of mathematics education studies written by a mathematics educator interested in change. Highlights of more recent research reviews in mathematics education follow. Particular attention is given to two researchers, Suydam and Dessart, who provided the mathematics education community with essential reading material for educators and administrators alike. Next, a series of investigations featuring calculator studies is discussed. These studies provide an important foundation for the current study. The chapter concludes with examples of meta-analyses in mathematics education. The meta-analyses of calculator research is accorded significant attention since the current study was a continuation of this work.

Reviews of Research in Mathematics Education

A historical example of a narrative review was found in the eighth yearbook of the National Council of Teachers of Mathematics. It was the oldest example uncovered by the current researcher. Benz (1933) reviewed one hundred and thirty-two studies related to the teaching of high school mathematics. At the time, Benz was concerned by the minimal amount of research conducted in the mathematics education field. He wanted to inspire change. His goal was to encourage educators to participate in more research endeavors. To achieve his purpose, Benz created an extensive summary of the research conducted between 1915 and 1932. By his own admission, Benz's review was not comprehensive. His report provided descriptions of studies covering a variety of subjects related to the teaching of high school mathematics. Benz was an early example of an educator with a desire to influence change. He used reviewing techniques to accomplish his goal.

The Work of Suydam and Dessart

Two prominent names in mathematics education are Suydam and Dessart. While their contributions to the field influence both research and pedagogy, their talents in reviewing are significant to the current study. These two experts generated a variety of reviews over a thirty-year time span. They each collaborated with other mathematics educators in several notable instances, and they worked together on a few projects. This section contains highlights of their reviews in mathematics education. Suydam's calculator reviews and her work with the Calculator Information Center at Ohio State University are included later in this chapter.

Early Dessart Reviews

Initial reviews by Dessart were published in the School Science and Mathematics journal during the 1960s. In 1964, Dessart concentrated his reviewing skills on mathematics in secondary education by conducting an extensive integration of research from the first three years of that decade. In the mid-1960s, Dessart collaborated with Paul Burns. They created a series of yearly reviews comprised of all topics in mathematics education. Their series consisted of two documents per year. One focused on research for elementary grades, while secondary grades were the central theme of the companion document (Burns & Dessart, 1965; 1966a; 1966b; Dessart & Burns, 1967). Each publication was divided into several categories. The topics under investigation by researchers and educators during those years are revealed through these reviews. Some of the categories and corresponding examples are listed below.

1. Methods and materials of instruction – The 1964 review of secondary education described mixed results generated by studies featuring discovery learning and programmed instruction (Burns & Dessart, 1966a). Positive and negative results were reported for these controversial methods of instruction. The 1965 elementary education review described a series of reports on Piaget-oriented methods of instruction (Burns & Dessart, 1966b).

2. Development of problem solving skills – The secondary education review of 1965 contained a discussion of potentially confusing elements that appear in the phrasing of word problems (Dessart & Burns, 1967). Informative studies comparing problem solving skills of students in modern and traditional arithmetic programs were highlighted in the elementary education review of the same year (Burns & Dessart, 1966b).

3. Problems of the low achiever in mathematics – In 1965, a featured study discussed the methods for determining the placement of low achieving students in high school mathematics classes (Dessart & Burns, 1967).

4. Achievement and attitude in mathematics – Research featuring pedagogical methods to increase the achievement and attitude levels of students in mathematics appeared in all but one review (Burns & Dessart, 1965, 1966a, 1966b). The 1964 elementary education review highlighted teaching techniques found to significantly influence achievement and attitude changes in students. In particular, students of varying ability levels and diverse socio-economic backgrounds were discussed (Dessart and Burns, 1965).

5. Teacher education programs – The elementary education reviews gave serious consideration to the best approaches for conducting pre-service and in-service teacher education programs (Dessart and Burns, 1965; 1966b).

Burns and Dessart (1965, 1966a, 1966b; Dessart & Burns, 1967) employed the narrative method of review to create these documents. They reported the information in the same manner it was presented by the original researchers.

The Second Handbook of Research on Teaching (Dessart & Frandsen, 1973) contained another monumental review by Dessart. For this work, he collaborated with Henry Frandsen to summarize ten years of empirical studies in secondary mathematics. Some of the highlights of that decade, as reported by Dessart and Frandsen (1973), are listed below.

1. Placing limits on the study of formal logic, increasing the coverage of probability and statistics in mathematics classes, and determining the influence of calculus on the high school curriculum were among the topics of interest to mathematics educators.

2. Discovery learning was a method by which students had the opportunity to discover and test mathematical concepts with guidance, but not direct instruction, from teachers. When compared to traditional instruction, results were slightly favorable for the discovery approach but not statistically significant. Teachers provided anecdotal evidence that students who practiced discovery learning were better equipped to handle problem solving situations and more likely to engage in critical thinking.

3. A controversial topic among mathematics educators throughout the decade was programmed instruction. Research rigorously debated whether or not programmed instruction was superior to teacher-directed instruction. The results were mixed and did not significantly favor one method over the other.

4. Results of the International Study of Achievement in Mathematics reflected unfavorably on the United States. When the mean scores of all nations taking the test were compared, the United States and Sweden received the lowest rankings. There was evidence that American students were aware of the difficulties this posed to our country. Survey data revealed American students understanding of the necessity for strong mathematics skills with regard to the development of our nation.

<u>**More Recent Reviews by Dessart**</u>

The reviews conducted during the later half of Dessart's career provided educators with the same invaluable resources as his initial research summaries written over three decades ago. However, in his later writings, Dessart exhibited a slight change in purpose. One of his intentions in creating the three research summaries discussed here was to encourage change in the mathematics classroom (Dessart & Suydam, 1976, 1983; Dessart, 1989).

In 1989, Dessart completed a review featuring brief summaries of significant research findings from 1987. Dessart gleaned studies from a compilation of research Marilyn Suydam reported in the July 1988 issue of the <u>Journal for Research in Mathematics Education</u> (JRME). Suydam's yearly JRME summaries are discussed later in this chapter. Dessart did not provide statistical details in this review. Instead, he created a reference document full of research-supported teaching methods applicable in mathematics classrooms. He hoped teachers would implement some of the ideas or gain inspiration to conduct their own educational experiments through formal or informal research techniques (Dessart, 1989).

In 1976 and 1983, Dessart collaborated on two projects with Suydam, another skilled reviewer. Their purpose paralleled Dessart's intention for his 1989 review. The "classroom ideas" series provided educators and administrators with concise summaries of current research without statistical details. These two seasoned educators used their reviewing talents to encourage change within the mathematics classroom. The series featured the topics of computational skills, algebra, and geometry. Classroom Ideas from Research on Computational Skills (Dessart & Suydam, 1976) supplied elementary educators with practical findings related to teaching these essential skills. Arithmetic operations with whole and rational numbers were featured. The document provided teachers with ideas for introducing, reinforcing, and maintaining their students' computational skills. It also provided methods for transferring and applying those skills to other mathematical areas.

The successor publication titled Classroom Ideas from Research on Secondary School Mathematics also supplied educators with pedagogical techniques backed by research (Dessart & Suydam, 1983). Dessart investigated the subject of algebra while Suydam concentrated on geometry. The algebra section highlighted three areas of significance for classroom teachers: organizing for instruction, useful teaching methods, and the importance of homework (Dessart & Suydam, 1983). The ideas presented were the cumulative findings from one hundred eighty-eight studies of algebra instruction covering a twelve-year time span.

For the geometry section, Suydam gleaned ideas from ninety-seven research studies. Initially, she examined researchers' perceptions regarding the importance of geometry in the high school curriculum. This examination was followed by an investigation into several important topics including the structure of geometry courses, useful teaching methods, the significance of logic and proofs, and technology's impact on students' understanding of the subject (Dessart & Suydam, 1983).

Yearly Reviews by Suydam

For the last thirty years, the annual reviews by Suydam have been invaluable resources for members of the mathematics education community. They appeared in the July issues of the Journal for Research in Mathematics Education (JRME) from 1970 through 1993. Suydam and several collaborators, including Weaver and Brosnan, participated in this endeavor. On a yearly basis, Suydam organized an extensive bibliographic summary of all available research. Studies were listed according to three basic categories: research summaries, journal-published reports, and dissertation abstracts. Included with the bibliographic information was a brief summary of the study as well as the featured age or grade level (Suydam & Brosnan, 1994).

While each of the annual reviews followed the same general format, several notable changes occurred during the twenty-four year time span. To highlight a significant topic of the time, the reviews in the mid-1970s contained a sub-section devoted to Piagetian-oriented research. In 1977, Suydam and her collaborators recognized the need for an index (Suydam & Weaver, 1978). One appeared at the end of every review for the years 1977 to 1993. The index grouped journal articles and dissertations according to mathematical topics, educational classifications, and student ability levels.

Studies selected for inclusion in the review satisfied two criteria (Suydam

& Weaver, 1972). First, the research involved some type of mathematical variable. Second, the study mentioned a specific implication for instruction or was deemed to have potential influence on the mathematics classroom. Satisfaction of the second guideline was determined by the educational expertise of Suydam and her collaborators.

These annual reviews continue to provide mathematics educators with easy access to listings of years of research. Recent updates to the list have been sorely missed since Suydam's retirement in 1994. While these annual reviews comprised a substantial component of Suydam's career, they were not the only means by which she made an impact on the mathematics classroom. Her work with the Calculator Information Center is featured in the next section.

Reviews of Calculator Use in Mathematics Education

Since the meta-analysis central to this research endeavor serves to integrate the findings of a collection of calculator-based studies, it would be premature to summarize those studies in this review of literature. Therefore, this section is devoted to the findings of previous researchers who tackled this topic. More traditional methods of review like narrative and vote-counting techniques were used in these summaries. In preparation for the current study, only two calculator-related reviews could be found, which presented results applicable to precollege mathematics students, within the 1983 to 2000 time frame. Therefore, reviews conducted before 1983 are also included in the current discussion. The results presented by different researchers are not independent since many studies appeared in more than one review.

The Calculator Research of Suydam

Suydam of Ohio State University contributed several extensive documents to the discussion of the calculator's impact on the mathematics classroom. When her initial reviews were published, the educational relevance of the calculator was a controversial topic. In fact, her first report, Electronic Hand Calculators: The Implications for Pre-College Education, was published at a time when calculators could be found in only twenty-five percent of American classrooms (Suydam, 1976).

Twenty-four studies were evaluated in her first report. Suydam (1976) found most of them were implemented with poor research designs. Many of these initial studies involved small sample sizes and short treatment periods. As a result, most of them were unable to report significant findings. Suydam presented broad generalizations from these studies. Two examples were, "children generally enjoy using calculators" and "calculators may or may not facilitate particular types of achievement" (Suydam, 1976, p. 23). In an effort to assess educators' attitudes toward the use of calculators, Suydam (1976) incorporated survey data in her report. The results were displayed through lists of attitude-related statements. One list advocated the use of the calculator in the classroom and the other attempted to minimize the calculator's place in the study of mathematics.

In an effort to update the information contained in the initial report, Suydam produced five state-of-the-art reviews with data collected between 1978 and 1982 by the Calculator Information Center at Ohio State University. These reviews provided summaries of follow-up research and highlighted classroom innovations which resulted from calculator use. The 1978 report stated that calculators were becoming more prevalent in classrooms and teachers were beginning to design evaluation materials for which the calculator would be neither a help nor a hindrance (Suydam, 1978). The second state-of-the-art review (Suydam, 1979a) reported more widespread use of calculators in classrooms in the United States. In spite of some significant limitations, Suydam (1979a) asserted that the studies conducted in 1979 did not indicate calculator use hindered students' mastery of mathematical concepts.

The second and third reports (Suydam, 1979a, 1980) responded to calculator foes who insisted the device negatively affected students' scores on standardized tests of mathematical achievement. The data uncovered by Suydam revealed two-thirds of the researchers did not allow students to use calculators during testing. However, students with access to calculators during instruction received test scores as high or higher than their non-calculator counterparts. In the 1980 state-of-the-art-review, Suydam reported seventy percent of educators spanning all precollege grade levels were receptive to teaching mathematics with calculators. This fourth review (Suydam, 1981a) highlighted the use of the calculator in the development of problem solving skills. Research revealed, when the calculator was available, that students incorporated a wider variety of problem solving techniques and were more willing to tackle difficult problems. Furthermore, their scores on tests of problem solving ability were not significantly different from students without access to calculators.

The final report summarized one hundred fifty available calculator studies by focusing on the effect of the calculator on three important pedagogical areas: achievement with traditional instruction, achievement with special curriculum materials, and attitude towards mathematics (Suydam, 1982). Suydam presented the results of seventy-five studies on the relationship between student achievement and the use of calculators in traditional instruction. The studies spanned all precollege grade levels with the majority involving grades six through nine. Seventy-nine percent of the studies reported results favorable to calculator use in instruction. On achievement tests following traditional instruction, the students using calculators attained scores that were as high or higher than the students who studied the same material without calculators (Suydam, 1982).

By 1982, thirty-three studies had been conducted on the use of calculators and special curriculum materials in mathematics courses. All of the studies affirm that the calculator had no negative effect on student achievement. Roughly half of the studies described significant differences on achievement tests in favor of the students using calculators (Suydam, 1982).

A note on the relationship between the two treatments is important at this juncture. While the results are interesting, the actual influence of the calculator can not be determined from these studies. The positive effects may be the result of the calculator, the curriculum materials, or a combination of the two. Therefore, these results should be interpreted with caution. The calculator's influence on students' attitudes toward mathematics was also difficult to determine with respect to the studies described in this review. One-sixth of the studies revealed the calculator had a positive effect on student attitudes. However, in the remaining studies, there were no significant attitudinal differences between calculator groups and non-calculator groups (Suydam, 1982).

With Suydam acting in an editorial and supervisory role, the Calculator Information Center produced a series of "information bulletins" to address the concerns of teachers interested in using calculators in mathematics instruction (Suydam 1979b, 1981b). The bulletins provided educators with a tremendous amount of important information. Several examples are listed below.

1. Reviews of research related to the use of calculators in education.
2. A list of guidelines for selecting a calculator for educational use.
3. Recommendations on how to successfully incorporate the calculator in mathematics instruction.
4. Summaries of published articles with detailed suggestions on using the calculator to teach mathematical concepts.

Early Reviews of Calculator-Based Research

The work of Suydam was supplemented by reviews of others interested in the calculator's role in the mathematics classroom. Parkhurst (1979) reviewed thirteen calculator-based studies conducted between 1975 and 1979. Several studies comparing students using calculators with students not using calculators revealed small but significant differences in favor of the calculator group (Parkhurst, 1979). These studies were conducted at the elementary and high school grade levels. The remainder of the studies were unable to report significant differences between the two groups. Therefore, Parkhurst (1979) asserted the calculator was neither a help nor a hindrance to students at the middle school level. Parkhurst commented on the mixed but hopeful results by saying "...this slight positive trend signals possible optimism for use of the calculator in mathematical instruction" (Parkhurst, 1979, p. 8).

Roberts' review (1980) was the most extensive of the reports written before 1983. He synthesized the findings of thirty-four empirical studies spanning all educational divisions. In half the studies involving elementary grades, Roberts reported significant benefits to students' computational skills due to calculator use. This result was given credence by the fact that none of the studies allowed students to use calculators during posttesting. On the other hand, only one elementary study reported calculators had a positive impact on students' understanding of mathematical concepts and attitudes toward mathematics (Roberts, 1980). Interestingly, studies at the junior and high school levels revealed similar results. Six out of eleven reports demonstrated conclusively the calculator was a positive influence on students' computational skills. For conceptual and attitudinal constructs, positive results were reported by only two researchers.

Based on his study, Roberts (1980) called for improvements and further research in three pedagogical areas:

1. Guidelines on the most appropriate time to introduce the calculator to students.

2. The calculator's role in the creation of tests and the determination of testing procedures.

3. Methods of disseminating research reports to ensure classroom teachers ready access to significant material.

One year later, Rabe (1981) reviewed twenty-six documents regarding the use of calculators in the classroom. Several of the research-based studies were also part of the Parkhurst (1979) report. As well as describing the results of calculator-based research, Rabe (1981) discussed the curricular changes necessary to make the calculator an educationally effective tool. For her research summary, Rabe employed a vote-counting approach. Fourteen of the studies revealed significant differences in favor of students using calculators. Only two studies reported significant differences in favor of their non-calculator counterparts. Rabe (1981) also described the positive relationship between students' motivation to learn mathematics and the use of calculators.

As for curricular changes, Rabe (1981) noted when calculators were used, teachers were more selective in the types of paper-and-pencil skills they required their students to master. She also reported on teachers' observations of their calculator and non-calculator students. Students using calculators were more productive with the time they were given to practice problem solving skills and were more likely to assess the reasonableness of their answers before moving on to other problems (Rabe, 1981).

Sigg wrote a review around the same time the Parkhurst and Rabe reports appeared in print. His synthesis incorporated twenty-two research based studies and fifteen curriculum reports (Sigg, 1982). In the first part of the study, Sigg provided a brief synopsis of each document and the original researcher's most significant conclusions. The research was divided into three categories:

1. Calculator effects on computational achievement and attitudes toward mathematics.

2. Calculator effects on problem solving skills.

3. Curricular changes implemented by educators as a result of the expanded availability of calculators in the mathematics classroom.

Along with the narrative summaries, Sigg (1982) employed a vote-counting strategy to further assess the reports. Seven of the thirteen studies which highlighted computational skills yielded significant differences in favor of the students using calculators. Three of the studies revealed improvements in students' attitudes after they had access to calculators. Five out of nine studies indicated the calculator was beneficial to the development of problem solving skills.

Regarding curricular changes, even though many calculator recommendations were available in the early 1980s, most classroom teachers had not implemented the strategies with their own students. To Sigg (1982), it appeared teachers were wary that the possibility of negative calculator effects outweighed the benefits of calculator use. Teachers were not aware of research that described the positive benefits of calculators without compromising students' paper-and-pencil skills. Like Roberts, Sigg (1982) expressed concern about the availability of research for educators in the field.

In a short review of fourteen documents, Neubauer (1982) arrived at conclusions different from those offered by previous reviewers. Neubauer reported results of five research studies and synthesized findings from nine evaluations of standardized test scores. Neubauer's (1982) impression of the research can be summarized in three statements.

1. Calculators should not be used in elementary grades. Students are learning the basics and should focus on developing paper-and-pencil skills.

2. Calculators are ineffective learning tools for students with below average abilities in mathematics.

3. Calculators are useful in helping average or above average students to develop problem solving skills.

More Recent Calculator-Based Reviews

In 1993, Gilchrist conducted a review with the needs of adult students as her focus. Her primary concern was the availability of calculators for students taking the General Educational Development (GED) examination. While this population was not the focus of the current study, her review was relevant due to its inclusion of studies conducted at the precollege level. Gilchrist discussed research spanning a seventeen-year time period. Her report was divided into two parts, the use of the calculator in mathematics instruction and the use of the calculator in standardized testing.

While the results of calculator studies were mixed, in later studies Gilchrist found positive results were becoming more prevalent. Gilchrist (1993) and Neubauer (1982) agreed on one aspect of calculator use. They both found calculators to be successful learning tools for students developing problem solving skills. Gilchrist disagreed with Neubauer on the relationship between calculators and low ability students. In her review, she found calculators were beneficial to students of below average ability. The technological device minimized the computational weaknesses of low ability students and allowed them to focus on developing problem solving skills.

Gilchrist's (1993) report was the only review uncovered by the current researcher to discuss the use of the calculator in standardized testing. The calculator was not beneficial for students with regards to test questions of problem solving skills. Since the calculator has the potential to allow students to answer computational test questions with the push of a button, Gilchrist described the efforts of testing boards to write tests for which the calculator is not needed. Based on changes in standardized testing policies, Gilchrist asserted the doors would open to other possibilities for calculator use in testing and instruction (1993). Due to the prevalence of calculators in the workplace, Gilchrist rationalized the classroom and testing room should follow suit. In her words, educators should "stop preparing students for the past and start preparing them for the future" (Gilchrist, 1993, p. 36).

The final narrative review highlighted the use of one specific type of calculator – the graphing calculator. Penglase and Arnold (1996) reviewed high school and college studies from 1990 to 1995. The results relevant to precollege students are reported here. Penglase and Arnold (1996) searched for answers to two questions:

1. How did the graphing calculator benefit student achievement in mathematics?

2. What kind of learning environment allowed for maximum benefits to be attained?

Based on the evaluated research, the researchers found the answers to those questions to be "elusive and conflicting" (Penglase & Arnold, 1996, p. 59). Because traditional skill-based testing procedures were used to evaluate student achievement, many studies reported inconclusive findings. Penglase and Arnold (1996) agreed with Roberts (1980) and Gilchrist (1993) with regard to calculators and testing procedures. They all expressed a need for new methods to evaluate mathematics students who have been exposed to technology.

Penglase and Arnold (1996) stated research favored the use of the calculator in precalculus courses. However, in most studies, the differences between calculator and non-calculator groups were not significant. In studies featuring students' understanding of the concept of function, the results were more favorable for the calculator but still mixed (Penglase & Arnold, 1996). In general, the researchers were unable to report significant results overwhelmingly in favor of the calculator.

Students' understanding of graphical concepts and their capabilities with spatial visualization skills were two areas that provided conclusive results in favor of graphing technology (Penglase & Arnold, 1996). While students may have struggled with function concepts, calculators helped them make meaningful connections between functions and their graphs. However, most of the studies' original authors emphasized the need for students to spend time evaluating the features of functions with tables and paper-and-pencil techniques. Penglase and Arnold (1996) reported a positive correlation between the development of spatial visualization skills and mathematical achievement. The results were especially significant for members of the female population who had notable difficulties with spatial visualization skills (Penglase & Arnold, 1996).

One interesting result of this review may be due to the incorporation of studies involving college students. Penglase and Arnold (1996) reported the graphing calculator positively influenced students' attitudes toward mathematics. However, some students expressed concern about becoming dependent on calculators and losing their paper-and-pencil skills. This point, as well as the debatable issue of the calculator's role in the classroom, led the researchers to conclude that many pedagogical issues need to be resolved before students will achieve maximum benefits from calculator use in the study of mathematics (Penglase & Arnold, 1996).

Meta-Analyses in Mathematics Education

While the need for the narrative form of review and other review techniques will never disappear, meta-analysis is a statistically significant method of evaluating the overall effect of a collection of research studies. Examples of meta-analysis are becoming more prevalent in educational research. Several examples from the field of mathematics education are described below.

Hembree Meta-Analysis on Effects of Hand-held Calculators

The earliest meta-analysis located by the author and conducted in mathematics education was the work of Hembree (1984), a doctoral student of Dessart. His research endeavor provided researchers interested conducting statistically relevant reviews of educational research with two invaluable assets. First, his study contained a step-by-step guide for using meta-analysis to integrate educational research. The guide is examined in chapter two of the current study. Second, he gave the reader a concise example of a meta-analysis that generated a set of significant results. Two years later, this meta-analysis was published in the Journal for Research in Mathematics Education (Hembree & Dessart, 1986). In 1992, Hembree and Dessart updated the report with data from nine additional studies. Hembree's calculator research is still widely discussed in mathematics education circles. For example, seventy percent of the studies integrated by the current meta-analysis make at least one reference to Hembree's work.

Hembree (1984) considered the effects of hand-held calculators on students' achievement and attitude in the K-12 classroom. He reported results of fifteen research questions which discussed the calculator's effect on three areas of learning:

1. The acquisition, retention, and transfer of operational and problem solving skills.

2. Estimation skills.

3. Attitude, mathematics anxiety, self-concept, motivation to learn, and student perceptions of the value of mathematics.

Hembree's (1984) study involved seventy-nine research studies spanning the years 1969 through 1982. He calculated five hundred twenty-nine effect magnitudes. Most of the studies Hembree included were doctoral dissertations; however, he also incorporated several journal articles, ERIC documents, and an unpublished report. Hembree provided detailed lists of effect sizes, schematic box-and-whisker plots of the values, and mean effect sizes for each research question in his study. A positive effect size revealed the calculator had a positive effect on the set of students in question. While a negative effect size implied the control group performed better than the treatment group in that particular study.

The studies in Hembree's (1984) meta-analysis best described calculator use in grades three through nine. Kindergarten through second grade were significantly underrepresented and grades ten through twelve were only moderately represented. Therefore, the results of Hembree's (1984) meta-analysis were most significant for the upper elementary and middle school grades. Each study involved statistical comparisons of students who used calculators with students who studied the same mathematical material without calculators. The first significant finding of Hembree's study was that the calculator had no significant effect on the students' conceptual knowledge of mathematics. He found this to be true for students in every grade level.

When evaluating computational and problem solving skills, Hembree (1984) separated the effect sizes into two categories. One category contained the studies that did not allow students in the experimental groups to use calculators on tests. This category was used to determine whether or not the calculator had a maintenance effect on students. The other category was comprised of the studies allowing calculator use on tests. Hembree used these studies to determine possible extension effects from calculator use. Further elaboration on the definitions of maintenance and extension can be found in chapter one of the current study.

In the sense of extension, the meta-analysis (Hembree, 1984) found that students of low or average ability experienced improvement in both computational and problem solving skills after using the calculator. For students of high ability, the calculator had no effect on computational skills but provided moderate aid in the development of problem solving skills. In the maintenance sense, the meta-analysis revealed different results based on student ability levels. Average students who used the calculator realized improvements in both computational and problem solving skills. The only exception was the fourth grade. Calculators had a negative effect on the computational skills of fourth grade students. Students in the high and low ability categories received no benefits from calculator use in computational or problem solving skills.

When considering these results, the reader should remember that data revealing no effect does not imply there is a negative effect. This meta-analysis revealed several instances in which the calculator was neither a help nor a hindrance to students' achievement in mathematics.

In general, Hembree (1984) interpreted his findings as encouraging for the role of calculators in the mathematics classroom. The calculator helped students of average ability develop computational and problem solving skills. At the same time, the skills of low and high ability students were not harmed by calculator use. Hembree stated the one negative result in grade four should remind educators that "calculators, though generally beneficial, may not be appropriate for use at all times, in all places, and for all subject matters" (Hembree & Dessart, 1992).

Hembree (1984) also found the calculator to be a positive influence on students' attitudes toward mathematics. This was true for all grade and ability levels. Also, the students in the calculator groups appeared to have more positive self-images regarding their mathematical abilities as compared to their non-calculator counterparts. Due to a lack of sufficient data, Hembree (1984) was unable to provide answers to the research questions related to student motivation to learn mathematics and student perceptions of the societal value of mathematics. Hembree's results regarding the relationship between calculators and students' estimations skills were positive but not statistically significant.

In 1992, Hembree and Dessart extended the results of the original meta-analysis with data from nine additional studies. In terms of student achievement, the new data "either supported or enhanced" the findings presented in the original meta-analysis (Hembree & Dessart, 1992, p. 26). Also, the new data did not negatively affect the original findings on student attitudes. The nine studies provided further support for the calculator's ability to boost students' attitudes toward mathematics. They also emphazised the calculator's impact on students' self-worth in relation to mathematical activities. It should be noted these nine studies were included in the current meta-analysis.

Hembree's (1984) meta-analysis and its update (Hembree & Dessart, 1992) were influential in the relationship between mathematics and technology. Even today, the results of these studies are widely reported in reviews of pedagogical and technology-related research. The work of Hembree and Dessart (Hembree, 1984; Hembree & Dessart, 1986, 1992) will continue to provide a solid foundation for studies regarding classroom use of technology. They represent clear examples of the impact meta-analysis can have on an educational field.

Smith's Update to Hembree's Calculator Meta-Analysis

In 1997, Brian Smith used meta-analysis to research similar questions to those in Hembree's (1984) study. Smith gathered twenty-four studies spanning the years 1984 to 1995. He wanted to determine whether the calculator's influence on students' attitudes and achievement in mathematics had changed since the Hembree study. However, his analysis was not as thorough as Hembree's original work. In particular, he did not gather an exhaustive collection of studies for the eleven-year time period. This fact came to the researcher's attention when preparing for the current meta-analysis. While it may not have been Smith's intention to include every relevant study in his analysis, this was the intention of the current researcher. A recent search for studies revealed several important documents that Smith did not include in his work.

As in Hembree's (1984) work, test results of students using calculators were compared to test results of students who did not use calculators. Smith (1997) generated fifty-four effect sizes from data provided by the studies. In terms of conceptual knowledge, Smith's (1997) study revealed that the calculator had a positive effect on mathematical achievement. This result was significantly different from Hembree's original result and indicated a positive change in the relationship between calculators and conceptual skills. However, Smith's result was based on nine effect sizes while a statistical integration of thirty-nine effect sizes was the foundation for Hembree's conclusions. Therefore, further research with a larger collection of studies is necessary to discriminate between these conflicting results.

While Hembree's (1984) research found student ability levels to be an influential variable in the relationship between achievement and calculator use, Smith (1997) chose to make distinctions by grade level. Smith reported the calculator had positive effects for students in third grade, grades seven through ten, and twelfth grade. Smith's data did not reveal conclusive results for students in kindergarten through second grade. In grades four through six and grade eleven, the calculator had no significant effect on achievement.

Smith (1997) also studied the relationship between the calculator and students' problem solving and computational skills. He found that the calculator had a positive effect on students in all grade levels for both of these areas. In particular, Smith (1997) stated improvements in computational skills increased as students progressed through the precollege grade levels. At the same time, students' paper-and-pencil skills were not hindered by calculator use. Smith and Hembree (1984) conducted data analysis according to different variables (i.e. grade levels as compared to student ability levels). Therefore, comparisons between the results regarding computational and problem solving skills were not attempted by the current researcher. Smith and Hembree agreed on the relationship between the calculator and students' attitudes toward mathematics. Smith's (1997) calculations revealed the calculator was a positive influence on students' attitudes. However, his results were based on only five effect magnitudes and should be considered with caution.

Hembree's Meta-Analyses of Other Educational Constructs

From 1986 to 1992, Hembree used his well-developed techniques in meta-analysis to evaluate several educational constructs. A few of his more recent meta-analyses are summarized below.

Test Anxiety

Hembree (1988) integrated a collection of test anxiety studies from all areas of academia, not just mathematics. He uncovered five hundred sixty-two reports with sufficient information for the research process. The studies were conducted between 1952 and 1986, and spanned all academic levels from kindergarten through college. In each report, two groups of students were compared. One group served as the control group while the other group received treatment for test anxiety. The length of treatment ranged from one to twelve hours.

For the purposes of the study, variations of test anxiety were separated into two categories. Worry referred to a student's concern about his own performance on the test. Emotionality involved involuntary reactions to testing conditions such as accelerated heart rate or perspiration. Treatment conditions were also divided into two categories. Behavioral treatments dealt with the emotionality aspects of test anxiety. Cognitive-behavioral treatments focused primarily on the components of test anxiety related to worry.

Hembree (1988) found a significant relationship between test anxiety and test performance for third grade and above. In particular, as test anxiety levels decreased, students' test performance increased. As an added bonus, students' grade point averages improved when anxiety levels were lowered. With his study, Hembree demonstrated that two widely assumed facts were true. First, females experienced higher levels of test anxiety than their male counterparts. Second, students in the early grades experienced very little test anxiety but anxiety levels increased with each subsequent grade level.

Behavioral and cognitive-behavioral treatments were effective in reducing test anxiety (Hembree, 1988). However, when students were taught study skills without any other form of treatment, changes in anxiety levels were not significant. In several studies, students were evaluated for retention of treatment benefits between three and sixty weeks after treatment. Analysis revealed that the students retained the positive effects of treatment. Test anxiety as well as general anxiety levels remained at or below end-of-treatment levels. In general, Hembree's (1988) analysis revealed that test anxiety was treatable, treatment effects were long lasting, and treatment positively influenced general anxiety difficulties. Behavioral treatments helped to reduce students' anxiety levels in areas outside of testing. Cognitive-behavioral treatments reduced students feelings of tension and anxiety in testing situations (Hembree, 1988).

Mathematics Anxiety

In 1990, Hembree conducted a meta-analysis on the related topic of mathematics anxiety. He integrated the findings of one hundred fifty-one studies spanning all academic levels from third grade through college. The length of treatment for mathematics anxiety ranged from three to twelve hours. The meta-analysis revealed several significant correlations between anxiety and performance in mathematics. First, all students who experienced high levels of mathematics anxiety performed at low levels on tests of mathematics achievement. For males in junior and high school grades, low levels of mathematics anxiety directly correlated to higher performance levels on mathematics tests. However, these results could not be extended to the college level (Hembree, 1990). Two well-known facts were proven by Hembree's work:

1. A significant, positive correlation existed between low mathematics anxiety and positive attitudes towards mathematics.

2. Females exhibited higher levels of mathematics anxiety than males.

Hembree (1990) compared the end-of-treatment anxiety levels of students participating in treatment with students who were not participating in treatment. He found classroom interventions involving special class work, equipment, or materials were not effective in reducing anxiety levels. Psychological treatments such as "systematic desensitization...along with anxiety management training and conditioned inhibition were highly successful in reducing mathematics anxiety levels" (Hembree, 1990, p. 43). When the treatment was effective in reducing anxiety, higher mathematics test scores were an added benefit.

Mathematics Problem Solving Skills

Hembree's (1992) final published report evaluated problem solving skills in mathematics education. He statistically integrated the results of four hundred eighty-seven research studies covering a period of sixty years. The meta-analysis included research across academic levels from kindergarten through college. The length of treatment ranged from five days to the entire school year. Several characteristics common to experienced problem solvers were revealed by Hembree's (1992) research. Effective problem solvers used diagrams, possessed a wide variety of problem solving techniques, and approached questions with particular problem solving strategies in mind.

Hembree revealed that students benefit from accurate visual representations or access to physical objects when tackling problems. Instruction in drawing accurate diagrams and translating English statements into mathematical equations positively affected student performance on problem solving achievement tests. Lastly, Hembree's (1992) research revealed students benefitted from direct instruction in the difficult, but extremely important, task of setting up equations before solving word problems. Hembree's meta-analysis supported the notion that students can be taught problem solving techniques and, therefore, become successful problem solvers.

Meta-Analyses Conducted by Other Educational Researchers

During the last eleven years, other researchers used meta-analytic techniques to integrate findings from a variety of topics in mathematics education. Several of them are described here.

Gender-related Differences in Completing Mathematical Tasks

In 1989, Friedman researched gender issues regarding the completion of mathematical tasks. Friedman integrated ninety-eight studies of students in precollege grade levels over a fifteen-year time span. The mean test scores of males were compared with their female counterparts. The test scores covered three categories: single-subject measures (i.e. computation, algebra, etc.), problem solving measures, and college entrance examinations. In most cases, Friedman used pretest scores to calculate effect sizes. She wanted to compare male and female scores "uninfluenced by intervention" (Friedman, 1989, p. 198). Effect sizes were defined to be the mean difference of average male and female scores divided by a pooled standard deviation. A negative effect size meant the difference favored males.

Friedman's (1989) results revealed the existence of gender differences with regards to the completion of mathematics tasks. In particular, males realized higher performance levels in accelerated and gifted mathematics programs and on college entrance examinations as compared to their female counterparts. However, by comparing her results to earlier research conducted by Maccoby and Jacklin (1974), Friedman concluded the gap between male and female mathematics performance was beginning to narrow. In particular, after 1980, the average effect size for females taking the Scholastic Aptitude Test (SAT) experienced an increase. Friedman concluded "women's performance [was] improving relative to that of men" (Friedman, 1989, p. 205). She stated environmental factors were the most prevalent reason for gender differences. The narrowing of the gap indicated significant changes in environmental influences. As a result, the mathematical capabilities of women were becoming more recognizable (Friedman, 1989).

Another meta-analysis involving gender-related differences was conducted by Wen-Ling Yang in 1997. However, Yang's intentions were quite different from those of Friedman. Yang evaluated data gathered from seventh and eighth grade students in twenty-five countries during the Third International Mathematics and Science Study (TIMSS). Yang chose meta-analysis because she believed a comparison of international data could benefit from the rigors of this statistical approach. She believed features of each country's data that could not be manipulated by the researcher would make other analytical techniques inappropriate for international comparisons. The data was gleaned from a questionnaire asking students to describe their beliefs regarding gender differences in learning mathematics. A 5-point Leikert scale, placing boys at one end of the spectrum and girls at the other, was used. Students were asked questions regarding who they believed was more likely to "be better at mathematics", "solve a difficult mathematics problem", or "have a natural talent for mathematics" (Yang, 1997, p. 6). A composite score for each student's responses to the questions was created. Average scores comprised the data used in the calculations of effect sizes.

In ten of the twenty-five countries, Yang (1997) found both male and female students believed females would be better at completing mathematical tasks. In seven countries, both male and female students perceived males to be better mathematicians. In the remaining eight countries the vote was split. Females believed females were superior in mathematics while males believed in male superiority. Yang reported the perceived gender differences in different countries fell on a fairly even split with females receiving slightly higher values than males. Yang (1997) asserted the results were disjointed and should be considered with caution. However, the work proved meta-analysis is an acceptable method for evaluating data from international studies.

Mathematics Anxiety

Another published example of meta-analysis in mathematics education can be found in the November 1999 issue of the Journal for Research in

Mathematics Education (JRME). Xin Ma (1999) used twenty-six research studies to examine the relationship between mathematics anxiety and mathematics achievement. The differences in the relationship with regards to gender, grade level, ethnicity, and methods of measurement were also evaluated. Ma (1999) reported the existence of a significant relationship between mathematics anxiety and mathematics achievement. In fact, this meta-analysis supports the results of Hembree's (1990) meta-analysis on the same topic.

Both studies revealed improvement in mathematical achievement when mathematics anxiety was reduced. In particular, Ma (1999) reported "such a reduction may be associated with an improvement from the 50^{th} to 71^{st} percentile in mathematics achievement for an average student highly anxious about mathematics" (p. 523). Ma (1999) stated the relationship between anxiety and achievement in mathematics was consistent across gender, grade, and anxiety rating scale levels. Therefore, students of both genders at various grade and anxiety levels experienced higher levels of success in mathematics when their anxiety was reduced.

Summary

The studies discussed in this chapter are examples of educators and researchers documenting change in mathematics education. Reviews of calculator-based research charted the evolution of the calculator's role in the classroom over the last twenty years. Meta-analytic investigations of calculators and other pedagogical influences revealed some interesting findings. Some of the results were conclusive and overwhelming. All of the results provide foundations on which future research will be conducted. For many years to come, mathematics education will continue to be in a gradual yet constant state of change. The change of particular importance to the current researcher is the calculator's effect on mathematics students. The current reform effort in mathematics education led by the National Council of Teachers of Mathematics (NCTM, 2000) calls for technology to continue to play an important role in the classroom. Therefore, the need for other meta-analyses in this area is certain. The current study is a continuation of the work begun over fifteen years ago by Hembree (1984) and extended by Smith (1997) in more recent years.

Chapter IV

Methods

The current study followed the procedures outlined in Hembree's (1984) model for meta-analysis. In a few instances, updated techniques described in more recent publications (Hedges & Olkin, 1985; Hedges et al., 1989; Carson et al., 1990) were incorporated into this study. Differences between the Hembree model and the updated procedures will be noted where appropriate.

The first phase of this project was to establish a research topic. With the foundation outlined in chapter three, the effects of calculator use in the K-12 classroom was chosen for analysis. The research questions, listed in chapter one, were then developed. Since this study was a continuation of the work of Hembree, the questions were similar to those researched by Hembree. Once the topic and research questions were defined, the meta-analysis proceeded through four basic steps:

1. Studies relevant to the topic were identified.

2. Studies were selected based on their ability to satisfy a set of necessary assumptions.

3. The properties and outcomes of the studies were coded.

4. A statistical analysis was performed on the data.

The details of these steps are outlined below.

Limitations

1. The results of this meta-analysis are directly dependent on the data provided by the original researchers. Therefore, the research findings reflect exactly what was provided in the original documents.

2. While the researcher made every effort to find all available studies, both published and unpublished, it is unlikely total success was realized. Therefore, the validity of this meta-analysis is dependent on the degree to which the collection of studies is representative of the entire population of existing studies. To allow for further assessment of the validity issue, a fail-safe N was calculated for the appropriate research questions. The process is described in detail later in this chapter.

Delimitations

The general purpose of this meta-analysis is to answer the question: How does the calculator affect students' achievement and attitude in the precollege mathematics classroom? In particular, has the effect of the calculator changed in the sixteen years following the original research conducted by Hembree? Two significant limitations are found in the definition of the student population and the types of calculators used in the original studies:

1. The population was established as students in K-12 mathematics classrooms.

2. Hand-held calculators, including basic, scientific, and graphing calculators, were the technological devices under scrutiny. This was appropriate since, aside from computers, hand-held calculators are the devices most likely to be found in present-day and future classrooms.

Identification and Collection of Studies

The initial search for studies involved a perusal of the Educational Resources Information Center (ERIC) and Dissertation Abstracts International (DAI) computer databases. Descriptors like "calculator", "graphing calculator", "hand-held calculator", "mathematics education", "mathematics instruction", and "research" were used to locate an initial list of abstracts from which other search techniques could be implemented. Three other techniques were used to locate citations and abstracts:

1. A manual search of the annual bibliographies in the Journal for Research in Mathematics Education was conducted.

2. A manual search of the appropriate volumes of Dissertation Abstracts International was conducted.

3. As a study was evaluated for inclusion in the meta-analysis, its accompanying bibliography was scanned for other inclusion possibilities.

During these initial searches, studies were selected if they fit the following criteria:

1. The study featured the use of a hand-held calculator.

2. The study involved students in a mainstream K-12 classroom.

Following Hembree's (1984) model, the selected studies met five other requirements for assurance that the prerequisites of meta-analysis were satisfied:

1. The experimental studies provided data necessary for the calculation of effect sizes. Means and standard deviations were used most frequently, but other statistics like gain scores or summary data from an ANOVA table were also used in effect size calculations.

2. The correlation studies provided Pearson's product-moment correlation or other correlation statistics necessary for transformation into Pearson's product-moment correlation.

3. The data used in the configuration of dependent variables were numerical values on a continuous scale.

4. Each study sample size contained a minimum of ten subjects, except for studies that used whole classes as units of analysis.

5. If the results of a study appeared in more than one research report, the more complete report was included in the meta-analysis.

6. No report was rejected due to a flawed design.

All of the studies found in the initial search were reviewed based on the aforementioned requirements. Some studies were eliminated immediately by reading their abstracts. All other studies were analyzed thoroughly by the researcher. The University of Tennessee at Chattanooga library provided copies of journal articles and ERIC documents. Dissertations were borrowed from university libraries participating in interlibrary loan services. Studies that could not be gathered in this manner were purchased from University Microfilm International. Unpublished documents were requested directly from the authors. Electronic mail was used to make the requests and the documents were received through this medium as well.

Coding the Studies

The information in each study was coded to allow the data to be easily accessed during the evaluation phase of the meta-analysis. The various characteristics of the studies were independent variables. The effect sizes calculated from data regarding the outcome constructs of achievement and attitude were dependent variables.

Independent Variables

Each independent variable was classified as either categorical or continuous. Categorical variables were used to classify entities into two or more sections. Continuous variables were those for which attributes were measured in a numerical fashion. All of the independent variables, except one, were the same as those initially established by Hembree (1984). A category for type of calculator was added to the current meta-analysis.

The categorical variables for the current study were:

1. Form of publication – Journal article, ERIC document, dissertation, or unpublished report.
2. Type of measuring instrument – Standardized, teacher designed with reliability information provided, or teacher designed without reliability information.
3. Grade level – K through 12.
4. Student ability level – Mixed, low (below average), average, or high (above average).
5. Student socio-economic status – Mixed, low, middle, or upper.
6. Subject matter – Mathematical subject studied during treatment.
7. Experimenter bias – Much or little. (A judgement of the experimenter's direct involvement in the study.)
8. Curriculum – Traditional methods except for the calculator or special methods including materials developed specifically for use with the calculator.
9. Calculator use – Functional use, essentially for computation or pedagogical use, essentially an aid in teaching or learning concepts.
10. Calculator availability – Ratio of one calculator to the number of students who use it (i.e. 1/2)
11. Ethnicity – Predominant ethnic or cultural group involved in the experiment.

12. <u>Gender-related differences in calculator effects</u> – None, results favored females, or results favored males.

13. <u>Type of calculator</u> – Basic, scientific, or graphing.

The <u>continuous variables</u> for the current study were:

14. <u>Year of report publication</u> – 1983 to 2000.

15. <u>Length of calculator treatment</u> – Number of school days.

16. <u>Time of calculator use per treatment session</u> – Number of minutes.

17. <u>Amount of teacher training with the calculator</u> – Number of hours.

18. <u>Teacher experience</u> – Number of years.

19. <u>Retention period</u> – Number of weeks. (This only applied to the studies which examined student retention.)

20. <u>Research design rating</u> – A value on the scale of 1 to 3.

The method used to calculate research design ratings was originally described in Hembree's (1984) method. Each study was assessed according to eight criteria: problem definition, population description, sampling procedures, error control, test instruments, data analysis, conclusions, and evaluation of the overall report. The guidelines used to evaluate each criterion are described in Appendix A. Six out of eight criteria were analyzed on a scale of one to three. Sampling and error control were considered to be more important than the other six, therefore, they were rated on a scale of one to six. The numerical rating was calculated by adding the total number of points and dividing by ten.

Statistical Assumptions

As originally presented within the Hembree (1984) model, assumptions necessary for statistical treatment of the data were applied to the current study:

1. All data used in effect size calculations fulfilled the requirements of a statistical t-test.

2. The instruments (i.e. test scores and attitude surveys) of the various studies used to measure the same construct were linearly equatable.

3. The use of large-sample statistical approximations was appropriate and well-founded.

4. The collection of studies was a probabilistic sample and contained all of the studies the researcher was able to unearth through extensive search techniques.

Dependent Variables

The constructs of achievement and attitude were featured within the collection of studies. For the current meta-analysis, dependent variables were defined as calculator effects on the achievement and attitude constructs. In particular, the dependent variables were:

1. Calculator effects on <u>achievement</u> related to the acquisition, retention, and transfer of operational and problem solving skills in terms of maintenance and extension. (These terms were defined in chapter one.)

2. Calculator effects on attitude concerned students' attitude toward mathematics, anxiety toward mathematics, attitude toward the use of calculators in mathematics, self-concept with respect to mathematical abilities, motivation to learn mathematics, attitude toward mathematics teachers, and impressions regarding the value of mathematics in society.

Some studies provided data on both constructs. As a result, more than one dependent variable was available for analysis. Dependent variables were represented numerically. The values, called effect sizes, were calculated from the data provided by the outcome measures of the studies. An effect size is a numerical representation of the "extent to which post-treatment outcomes for a calculator sample differed from the outcomes for a non-calculator sample" (Hembree, 1984). Effect sizes were calculated differently for experimental and correlational studies. The equations for effect sizes and correlation coefficients described below are based on those found in the Hembree (1984) model and the writings of Hedges (Hedges and Olkin, 1985; Hedges et al., 1989)

Effect Size Equations for Data from Experimental Studies

For experimental studies, the effect size is defined to be the difference in experimental and control group means divided by a pooled standard deviation. In the formula below, $\overline{Y}_E$ and $\overline{Y}_C$ are the respective experimental and control group sample means and S_P is the within-group standard deviation.

$$ES = \frac{\overline{Y}_E - \overline{Y}_C}{S_P} \tag{1}$$

The pooled variance, S_P^2, requires the experimental and control group sample sizes, n_E and n_C, and the experimental and control group standard deviations, s_E and s_C.

$$S_P^2 = \frac{(n_E - 1)s_E^2 + (n_C - 1)s_C^2}{n_E + n_C - 2} \qquad (2)$$

If sample means are not reported in the study, but a t-test or two-group F-test is used to determine an outcome construct's significance level, the equations listed below are used to calculate effect sizes.

$$ES = t\sqrt{\frac{1}{n_E} + \frac{1}{n_C}} \qquad (3)$$

$$ES = \sqrt{F}\sqrt{\frac{1}{n_E} + \frac{1}{n_C}} \qquad (4)$$

The sign of the resulting effect size is determined by considering which group generated the higher test results. The sign is "+" if the test results favored the experimental group using calculators. The sign is "-" if the test results favored the control group.

When a study reports data from a one-way analysis of variance (ANOVA) or analysis of covariance (ANCOVA), the within-group variance, s_W^2, is a good estimate of the population variance (Hembree, 1984). In these cases, s_W is used in the denominator of effect size equation (1) creating the following equation.

$$ES = \frac{\overline{Y}_E - \overline{Y}_C}{S_W} \qquad (5)$$

When ANCOVA statistics are reported, $\overline{Y}_E$ and $\overline{Y}_C$ are adjusted group means. The aforementioned equations were used to calculate effect sizes for the study.

Effect Magnitudes from Correlational Studies

In correlational studies, effect magnitudes are defined as the Pearson's product-moment correlation. The current meta-analysis did not include any studies reporting correlational statistics other than Pearson's r. Therefore, conversions to Pearsons' r were not needed. However, they are readily available in the writings of Glass (Glass et al., 1981).

Correction for Distribution Bias and Measurement Errors

As originally described by Hembree (1984), "each raw effect size is a sample statistic which estimates the underlying population effect size δ" (p. 132). Hedges and Olkin (1985) proved this estimate is biased based on the number of degrees of freedom, $n_E + n_C - 2$, for S_P^2 in equation (2). The exact values for the bias correction can be found in the writings of Hedges and Olkin (1985, p. 80). However, the approximation correction factor J(m) where $m = n_E + n_C - 2$ is accurate enough for the purposes of meta-analysis.

$$J(m) \approx 1 - \frac{3}{4m-1} \qquad (6)$$

Measurement error is a factor of instrument reliability, r_{YY}. Hembree states the error can be corrected by multiplying the effect size by the reciprocal square root of the instrument reliability. Frequently, studies fail to report the reliability data for their tests and survey instruments. In these cases, a value of $r_{YY} = 1$ is used.

In the current meta-analysis, each raw effect size generated through equations (1), (3), (4), or (5) was corrected for distribution bias and measurement errors with the following equation. This resulted in an error-free effect size value, g.

$$g = \frac{ES \cdot J(m)}{\sqrt{r_{YY}}} \tag{7}$$

Transformation of Correlation Coefficients

Correlation coefficients do not need to be corrected for bias. However, Hembree (1984) recommended transformations to Fisher's z values for the purpose of simplifying the data analysis process. The following equation is used for the transformation.

$$z = \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right) \tag{8}$$

For the current study, only two correlation coefficients were obtained from reported data. Since statistical calculations were not conducted on this small sample, transformations to Fisher's z values were unnecessary.

Recording Dependent and Independent Variables

For each study, the independent variables and the data required for calculations of effect sizes were recorded on a coding sheet (see Figure 2). The

Study ID No.______________ Construct______________

Author(s)____________________________ Treatment______________

Categorical Independent Variables

ID	Variable	Measure
X_1	Publication Form	
X_2	Instrument Type	
X_3	Calculator Type	
X_4	Grade Level	
X_5	Ability Level	
X_6	SES	
X_7	Subject Matter	
X_8	Experimenter Bias	
X_9	Curriculum	
X_{10}	Calculator Use	
X_{11}	Calc. Availability	
X_{12}	Ethnicity	
X_{13}	Gender-differences	

Continuous Independent Variables

ID	Variable	Measure
X_{14}	Publication Year	
X_{15}	Study Duration	
X_{16}	Time of Calc. Use	
X_{17}	Teacher Training	
X_{18}	Teacher Experience	
X_{19}	Retention Period	
X_{20}	Res. Design Rating	

Research Design Rating: X =

1. Problem Definition =
2. Population Description =
3. Sampling Prodedures =
4. Error Control =
5. Instruments =
6. Data Analysis =
7. Conclusion =
8. Report =

ES or r determination

$\bar{Y}_E$	$\bar{Y}_C$	S_E^2	S_C^2	S_P	t	n_E	n_C	raw ES or r

For use in analysis

r_{yy}	c_m	g or z
	Average g for dependent data	

Figure 2. Effect Size Calculation Coding Sheet

coding sheet was designed by using the example provided by Hembree (1984) as a guide. A separate form was used for each independent effect size, set of dependent effect sizes, or Pearson's product-moment correlation. All corrections for bias in effect sizes and correlational transformations were recorded on the coding sheets. The data collection phase of the meta-analysis was complete when all possible coding sheets contained the appropriate information regarding the independent variables and effect size calculations.

Data Dependence

Most studies provided enough data for more than one effect size to be calculated. Within one study, the same experimental and control groups were used to generate effect sizes. Therefore, the effect sizes for a single study were not independent. In most cases, different effect sizes corresponded to different outcome constructs (i.e. dependent variables). Since each outcome construct was analyzed separately, data dependence was not a significant problem. However, there were several occasions in which the experimental and control groups from one study provided data for more than one effect size for the same outcome construct. When this happened, the dependent effect sizes were averaged and the average g was incorporated in the data analysis.

Data Analysis

For each research question, effect sizes were calculated by the methods described above. By definition, a set of effect sizes is homogeneous if each effect size in the set is an estimate of the population effect size, δ (Hedges & Olkin, 1985). For a homogeneous set of effect sizes, the mean of the set is the best estimate of the population effect size, δ, and the amount of variance in the set is a result of sampling error (Hembree, 1984). Therefore, the central focus of data analysis was to test sets of effect sizes for homogeneity and to generate the appropriate effect size means.

For heterogeneous sets of effect sizes, the variance is larger than the amount expected from sampling error. The mean of the set is not a good estimate of the population effect size. Hembree (1984) describes two possible causes for heterogeneity:

1. Outlier data points may be the cause. By deleting the outliers, a homogeneous set of effect sizes may be created.

2. A study characteristic (i.e. independent variable) may cause heterogeneity. Separating the effect sizes into subsets according to their relation to the confounding study characteristic may result in one or more homogeneous subsets of effect sizes.

The statistics described below were calculated using two computer programs. MetaWin 2.0 is a statistical software program designed to perform all calculations necessary for a meta-analysis. It was used for the tests of homogeneity, the calculation of weighted means, and the corresponding confidence intervals. It was also used for calculation of fail-safe N values. MetaWin 2.0 does not contain a function for generating standardized residuals of effect sizes. Therefore, those calculations were performed with researcher defined equations in Minitab 12.0. All schematic box-and-whisker plots were generated in Minitab 12.0 as well.

Test for Homogeneity

The test for homogeneity used in the current study is the same as the one described in the Hembree (1984) model. For $g_1, g_2, \ldots, g_k$, a set of unbiased effect sizes, the i^{th} effect size, g_i, estimates a δ_i and the test for homogeneity determines whether or not all of the δ_i's are equal. In particular, $\delta_i = \delta$ for $1 \le i \le k$. If they are equal, then the set is homogeneous and the mean of the set is the best estimate of δ. The steps to determine homogeneity are:

1. Null and alternative hypotheses are established:

$H_0 : \delta_i = \delta$ for $i = 1, 2, \ldots, k$

H_A: At least one δ_i is different from the others

2. The test statistic, H_T is calculated with one of the following formulas:

For <u>experimental</u> effect sizes, g_i's, the test statistic, H_T, and the estimate of the population variance, $\sigma_i^2(\delta_i)$, are listed below:

$$H_T = \sum_{i=1}^{k} \frac{g_i^2}{\sigma_i^2(g_i)} - \frac{\left[\sum_{i=1}^{k} \frac{g_i^2}{\sigma_i^2(g_i)}\right]^2}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2(g_i)}} \tag{9}$$

$$\sigma_i^2(g_i) = \frac{n_{E_i} + n_{C_i}}{n_{E_i} n_{C_i}} + \frac{g_i^2}{2(n_{E_i} + n_{C_i})} \tag{10}$$

For correlational effect magnitudes, z_i's, the test statistic, H_T, is listed below:

$$H_T = \sum_{i=1}^{k}(n_i - 3)z_i^2 - \frac{\left[\sum_{i=1}^{k}(n_1 - 3)z_i\right]^2}{\sum_{i=1}^{k}(n_i - 3)} \tag{11}$$

In general, H_T has a χ^2 distribution with k – 1 degrees of freedom. Therefore, H_T is evaluated by comparing it to the critical value of χ^2_{k-1} at a pre-established significance level. In the case of the current study, a 5% level of significance was used.

3. Comparison of H_T with the critical value χ^2_{k-1} is the factor that determines whether or not a set of effect sizes is homogeneous.

   a. If $H_T < \chi^2_{k-1}$, then H_0 is not rejected. The set of effect sizes is determined to be homogeneous and the mean of the set and corresponding confidence interval are calculated.

   b. If $H_T \geq \chi^2_{k-1}$, then H_0 is rejected. The set of effect sizes is determined to be heterogeneous and a search for the cause of the heterogeneity is conducted.

These steps were followed to test sets of effect sizes for homogeneity.

Effect Size Means and Confidence Intervals

For each set of homogeneous effect sizes, the population effect size, δ, is estimated with a weighted mean of the unbiased effect sizes. Hembree (1984) stated "since estimates from studies with large sample sizes will be more precise than estimates from smaller samples, more weight should be accorded the more precise estimates" (p. 81). The equations used to calculate consistent estimates of the population effect size, δ, and related confidence intervals are listed below.

1. For <u>experimental</u> effect sizes, g_i's, the weighted mean, $\overline{g}_W$, and an estimation of the population variance, $\sigma^2(\delta)$, are calculated with the following formulas:

$$\overline{g}_W = \frac{\sum_{i=1}^{k} \frac{g_i}{\sigma_1^2(g_i)}}{\sum_{i=1}^{k} \frac{1}{\sigma_1^2(g_i)}} \qquad (12)$$

$$\sigma^2(\overline{g}_W) = \frac{1}{\sum_{i=1}^{k} \frac{1}{\sigma_i^2(g_i)}} \qquad (13)$$

The $(100-\alpha)$% confidence interval for δ at the significance level α is calculated with the weighted mean, $\overline{g}_W$, the estimate of population variance, $\sigma^2(\delta)$, and $z_{\alpha/2}$ obtained from the normal distribution.

$$\overline{g}_W - z_{\alpha/2}\,\sigma(\overline{g}_W) \le \delta \le \overline{g}_W + z_{\alpha/2}\,\sigma(\overline{g}_W) \qquad (14)$$

2. For <u>correlational</u> effect sizes, z_i's, the weighted mean, $\overline{z}_W$, and an estimation of the population variance, $\sigma^2(\delta)$, are calculated with the following formulas:

$$\overline{z}_W = \frac{\sum_{i=1}^{k}(n_i - 3)\,z_i}{\sum_{i=1}^{k}(n_i - 3)} \qquad (15)$$

$$\sigma^2(\bar{z}_W) = \frac{1}{\sum_{i=1}^{k}(n_i - 3)} \tag{16}$$

The $(100 - \alpha)$% confidence interval for δ at the significance level α is calculated with the weighted mean, $\bar{z}_W$, the estimate of population variance, $\sigma^2(\delta)$, and $z_{\alpha/2}$ obtained from the normal distribution.

$$\bar{z}_W - z_{\alpha/2}\,\sigma(\bar{z}_W) \leq \delta \leq \bar{z}_W + z_{\alpha/2}\,\sigma(\bar{z}_W) \tag{17}$$

Conversion from a weighted mean of Fisher's z values to a weighted mean of r, the Pearson's product-moment correlation, is conducted with the following formula:

$$\bar{r}_W = \tanh(\bar{z}_W) = \frac{e^{2\bar{z}_W} - 1}{e^{2\bar{z}_W} + 1} \tag{18}$$

The corresponding confidence interval for the population correlation coefficient, ρ, is calculated with a similar conversion.

$$\tanh(\bar{z}_W - z_{\alpha/2}\,\sigma(\bar{z}_W)) \leq \rho \leq \tanh(\bar{z}_W + z_{\alpha/2}\,\sigma(\bar{z}_W)) \tag{19}$$

The aforementioned equations were used to generate all weighted, means, variance estimates, and confidence intervals in the current meta-analysis. Once these values were calculated for every set or subset of homogeneous effect sizes and fail-safe N values were calculated for all statistically significant mean effect sizes, the data analysis phase was complete.

The Fail-safe N

As described by Hembree (1984), a mean effect size is statistically significant when the confidence interval generated for the mean effect size does not contain zero. Under these circumstances, a mean effect size is significantly different from zero. A fail-safe N value is a quantitative description of the validity of the results used to generate a statistically significant mean effect (Hedges & Olkin, 1985). The fail-safe N provides the number of studies reporting null results that will negatively affect the significance of the meta-analytic results (Carson et al., 1990). In particular, if the studies described by the fail-safe N were integrated with the existing studies, the mean effect size would be altered in such a way that its confidence interval would contain zero.

While the fail-safe N is a significant part of the work of Rosenthal (1979, 1991), Orwin's (1983) version of the fail-safe N directly applies to effect sizes. The process described below follows methods outlined by Orwin (1983), Hedges and Olkin (1985), and Carson et al. (1990) with a few changes in notation to correspond with the notation already established in this chapter.

The fail-safe N is calculated with the number of studies integrated by the meta-analysis, K, the mean effect size of the studies, $\overline{g}_W$, and a criterion value chosen for analysis, g_c.

$$N = \frac{K(\overline{g}_W - g_c)}{g_c} \qquad (20)$$

Based on the writings of Hedges and Olkin (1985), g_c is the effect size value to which $\overline{g}_W$ would be reduced by including N studies with null results in the calculation of the mean effect size.

For the current study, a fail-safe N was calculated for each statistically significant mean effect size. The value g_c was chosen to be the largest effect size with the same characteristics as $\overline{g}_W$ except for one important feature – the confidence interval for g_c contained zero. In terms of similarity, g_c and $\overline{g}_W$ had the same variance. Fail-safe N values were reported with the results of data analysis.

Search for Outliers

In the current study, outliers were found by two different methods. The first method was the one described by Hembree (1984) in his model for meta-analysis. The second method was found in the writings of Hedges and Olkin (1985). For each heterogenious set of effect sizes, the outliers were removed and the question of homogeneity was re-visited.

1. Effect sizes lying outside the inner fence of a schematic box-and-whisker plot were considered outliers. Further explanation of the plot and location of outliers can be found in chapter two (p. 31 & 32).

2. Effect sizes with standardized residuals larger than 3.00 significantly affect the homogeneity of a set of effect sizes. Therefore, these were also considered outliers.

The process involved in generating standardized residuals for a set of effect sizes is outlined below. It directly follows the method described by Hedges and Olkin (1985) with a few changes in notation to correspond with the notation already established in this chapter.

For a set of unbiased effect sizes, $g_1, g_2, \ldots, g_k$, and the weighted mean of the set, $\overline{g}_W$, a residual $g_i - \overline{g}_W$ is generated for each g_i for $1 \le i \le k$. Since $\overline{g}_W$ is created with the entire set of g_i's, the residuals are dependent. Therefore, the actual residual used in the calculations is the difference between g_i and a pooled estimate of δ that does not include the ith effect size. The pooled estimate is essentially the weighted mean of the g_i's without the ith effect size.

Let TW_i denote the sum of the reciprocals of the effect size variances with the ith value omitted and TWG_i denote the sum of the weighted unbiased effect sizes with the ith value omitted. The appropriate formulas for these values and the weighted mean, g_{Wi}, with the ith effect size omitted are listed below:

$$TW_i = \left[\sum_{i=1}^{k} \frac{1}{\sigma^2(g_i)}\right] - \frac{1}{\sigma^2(g_i)} \tag{21}$$

$$TWG_i = \left[\sum_{i=1}^{k} \frac{g_i}{\sigma^2(g_i)}\right] - \frac{g_i}{\sigma^2(g_i)} \tag{22}$$

$$\overline{g}_{Wi} = \frac{TWG_i}{TW_i} \tag{23}$$

Therefore, the residual of the ith effect size is $e_i = g_i - \overline{g}_{Wi}$. This residual "reflects the discrepancy between the ith estimate of effect size and a composite of the other observations" (Hedges and Olkin, 1985, p. 254).

One difficulty with e_i is that the variance is not constant for the entire set of effect sizes for which the search for outliers is being conducted. If e_i is created for an effect size from a study with a large sample size, then the variance of e_i

will usually be small. Similarly, if e_i is created for an effect size from a study with a small sample size, then the variance of e_i will usually be large. As a result, it is difficult to discern when e_i is truly large (Hedges and Olkin, 1985). A remedy for this situation is to standardize e_i with its variance. A good estimate for the variance of e_i, $\sigma^2(e_i)$, is calculated with TW_i from equation (17) and $\sigma^2(g_i)$ from equation (10).

$$\sigma^2(e_i) = \frac{1}{TW_i} + \sigma^2(g_i) \tag{24}$$

The i^{th} standardize residual, $\tilde{e}_i$, is calculated with e_i and its standard deviation.

$$\tilde{e}_i = \frac{e_i}{\sigma(e_i)} \tag{25}$$

Hedges and Olkin (1985) assert if a set of standardized residuals is generated for a set of effect sizes corresponding to one population effect size, then the set of standardized residuals has a distribution that is approximated by the normal distribution. Therefore, standardized residuals larger than 2.00 are found in homogeneous sets of effect sizes about 5% of the time. Standardized residuals larger than 3.00 are found even less frequently.

In the current study, each heterogeneous set of effect sizes was first analyzed for outliers by evaluation of a schematic box-and-whisker plot. If no outliers could be found, then standardized residuals were calculated and those larger than 3.00 were removed. The test for homogeneity was then conducted on the smaller set of effect sizes. If the set of effect sizes without the residual outliers was homogeneous, the appropriate means and standard deviations were calculated. If the set of effect sizes was still heterogeneous, a search for a confounding independent variable causing the heterogeneity was conducted.

Summary of Data Analysis Procedures

Outlined below is a list of procedures used in the current study. The concepts and equations involved in these procedures were discussed in this chapter. The entire set of effect sizes gathered during data collection was divided according to the research questions outlined in chapter one. To inhibit data dependence difficulties, the effect sizes corresponding to one research question were analyzed separately from the other values. The steps outlined below were followed for each set of effect sizes.

1. The effect sizes were graphed using a schematic box-and-whisker plot and the graphical outliers were identified.
2. The test for homogeneity was conducted without the outliers from step 1.
3. If the set was homogeneous, the following steps were taken.
    a. The effect size mean and corresponding confidence interval were calculated.
    b. The confidence interval was evaluated for statistical significance. If the confidence interval did not contain zero, the mean was determined to be statistically significant. This fact provided the statistical proof necessary for stating appropriate conclusions.

c. A fail-safe N value was calculated for each statistically significant mean effect size.

4. If the set was found to be heterogeneous, a search for residual outliers was conducted.

5. If no residual outliers could be found or the set was still heterogeneous after the removal of residual outliers, a search was conducted for homogeneous subsets related to a study characteristic. The effect sizes were grouped according to the confounding dependent variable and tests for homogeneity were conducted on each subset.

6. Effect size means and confidence intervals were calculated for the homogeneous subsets.

7. For heterogeneous subsets, further grouping by another confounding dependent variable was attempted.

8. If subsets existed in which no cause for heterogeneity could be found or the subsets were too small for further subgrouping according to confounding variables, the means were calculated and presented for descriptive purposes only.

Chapter V

Results

The purpose of this study was to determine the effects of calculators on students in K-12 mathematics classrooms. Data collection resulted in a set of unbiased effect sizes and an extensive list of study characteristics. The next task was to analyze the plethora of gathered data. This chapter describes the results of that task. General characteristics of the data are presented first, followed by the findings of each research question.

Features of the Data

Through the initial search, 83 studies were uncovered through the broadly defined category of calculator-based research in the K-12 classroom. All studies conducted between 1983 and June of 2000, which could be located by this researcher, were included. After evaluating the studies according to criteria necessary for meta-analysis, 30 studies were eliminated. The criteria involved in this process are explained in chapter four. Fifty-three studies remained from which data for meta-analysis was gleaned. The studies are listed in Appendix B. As required by meta-analysis, the outcome data of each study existed on a continuous, numerical scale. No report was rejected due to insufficient information. Whenever appropriate, missing information was obtained directly from the original author. Traditional and electronic mail were the means by which the researcher made the necessary inquiries. The researcher was able to obtain all essential information for effect size calculations and other aspects of data analysis. The final collection of studies included 34 dissertations, three master's theses, eight journal articles, one project report, five ERIC documents, and two unpublished reports.

A total of 307 effect magnitudes were calculated. Two values were Pearson's product moment correlation coefficients. The remaining 305 were effect sizes calculated according to the experimental data methods described in chapter four. They were generated from studies with quasi-experimental research designs in which the treatment group used the calculator during instruction while the control group received the same instruction without access to calculators. All effect sizes are listed in Appendix C. The values range from -1.3260 to 2.0341. Almost every study provided more than one effect size. One study provided the most with 18 values, while nine studies provided only one effect size. The researcher calculated all effect sizes and coded all study characteristics into the appropriate independent variables. Therefore, analysis of inter-rater reliability was not necessary.

The findings of a meta-analysis are based solely on data provided by the included studies; therefore, characteristics of the data are an invaluable component of the process. The coded independent variables for each report integrated within current study are listed in Appendix D. While it is hoped the results of this study will benefit present and future classrooms engaged in calculator use, the results are a direct reflection of the coded study characteristics and calculated effect sizes.

Distribution of Studies by Year of Publication

Publication years span the time frame of 1983 to June of 2000. This range was chosen with the Hembree study in mind. Hembree's (1984) calculator meta-analysis was conducted with all locatable studies through the year 1982. Since the current meta-analysis was an update to Hembree's work, the researcher included studies from 1983 to the present. Figure 3 illustrates the distribution. The two unpublished reports were accepted for publication in 2000. Therefore, it is the year under which they appear in the graph. The year providing the most reports was 1991. All years in the range were represented except 1997.

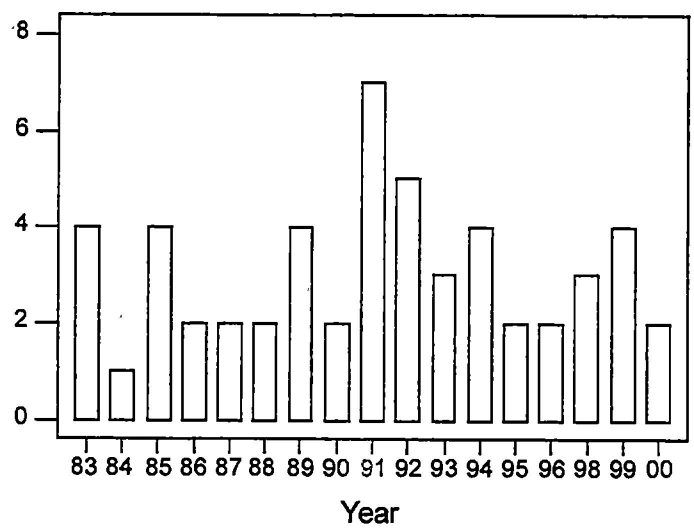


Figure 3. Distribution of Studies by Publication Year

Distribution of Studies by Grade Level

Mathematics students in mainstream precollege classrooms were the focus of the included studies. Roughly two-thirds of the reports featured more than one grade level. Hence, the grade level count exceeded fifty-three. The early elementary grades were the least represented grades within these studies. Kindergarten and first grade were each featured in one study and second grade students did not participate in any calculator-based research. Nearly seventy percent of the studies involved at least one of grades eight through twelve. The distribution of studies by grade level appears in Figure 4. Based on the spread, inferences drawn from this meta-analysis are best applied to mathematics students in higher grades, specifically eighth through twelfth grades. This is significantly different from the Hembree (1984) study in which grades three through nine provided the majority of the data.

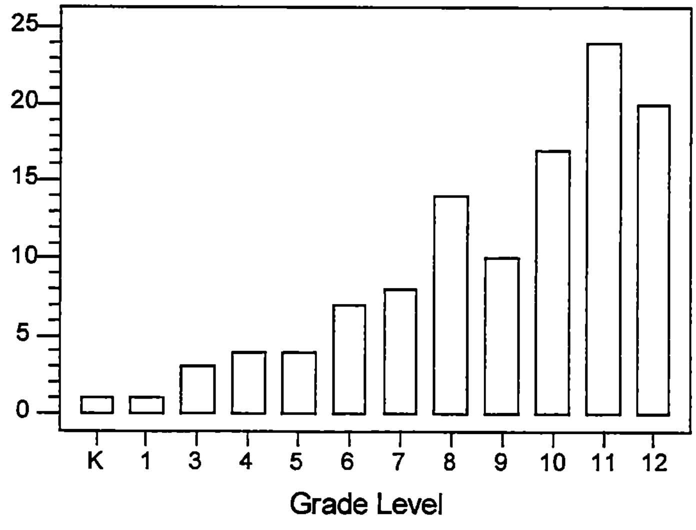


Figure 4. Distribution of Studies by Grade Level

Continuous Independent Variables

The length of calculator treatment ranged from zero to 650 days (i.e. 3½ school years). Studies that featured an exam and no calculator treatment were represented by zero treatment days. In particular, there were seven studies in which no pre-exam treatment took place. The duration of the treatment phase exceeded 30 days for nearly sixty percent of the studies. Only three studies evaluated students after a pre-determined retention period. The length of the retention period ranged from two to twelve weeks.

Based on the method of calculation, a research design rating assumed a value on the scale of one to three units. The criteria are listed in Appendix A. The actual range of the calculated values was 1.6 to 2.7. The mean research design rating was 2.0679. The median value was 2.1. Ten studies received the median rating. The range of average ratings was determined to be 1.8 to 2.3. Seventy-seven percent of the studies fell within the range and therefore received an average research design rating. Six studies incorporated better than average research designs.

Distribution of Effect Sizes by Educational Construct

Appendix C displays the unbiased effect sizes for each study. The two correlation coefficients evaluated the attitude construct. Due to the small number of values, statistical analysis was not attempted. Seventeen effect sizes were generated from the use of the calculator with special curriculum materials.

Students in the experimental groups received two forms of treatment while the control group received no treatment and traditional mathematics instruction. Thirteen effect sizes measured students' achievement levels and four values measured students' attitudes toward mathematics.

The remaining 288 effect sizes were generated by studies participating in traditional methods of instruction. Studies in which the experimental group used calculators while the control group had no access to calculators were the source for these values. Forty-three effect sizes measured students' attitudes toward mathematics. The achievement construct was represented by the remaining 245 values. Achievement was divided into three categories – acquisition, retention, and transfer of mathematical skills. The retention category contained thirteen effect sizes. Transfer of skills was represented by only three effect sizes. With 229 effect sizes, skills acquisition was well represented and yielded the most complete results during data analysis.

There were a total of 49 effect magnitudes related to the attitude construct. The total included values resulting from special calculator curriculum materials and the Pearson's product moment correlation coefficients. Figure 5 contains the distribution of effect sizes by attitude sub-construct. The calculator's effect on student attitudes toward mathematics generated the largest number of effect magnitudes. The remaining sub-construct categories contained six or fewer values. The most underrepresented categories were student anxiety toward mathematics, student motivation to increase mathematical knowledge,

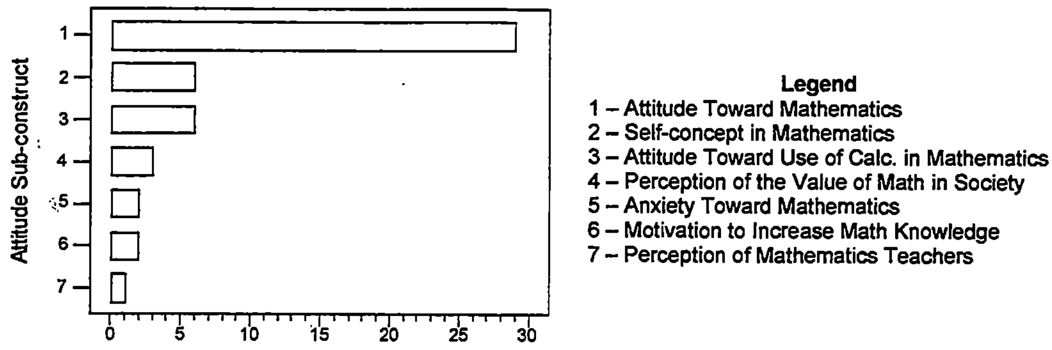


Figure 5. Distribution Effect Sizes by Attitude Sub-Construct

and student perceptions of mathematics teachers. While none of the categories contained an overwhelmingly large number of values, the attitude toward mathematics sub-construct had the most solid foundation for data analysis.

Calculator Use By the Experimental Groups

In his study, Hembree (1984) described two different roles for the calculator. Functional use implied the calculator was available for computation, drill and practice, and checking paper-and-pencil work. For pedagogical use, the calculator was an essential element in the teaching and learning of mathematics. In Hembree's (1984) study, the calculator predominantly assumed a functional role. Hembree described only a few studies in which calculators were an integral part of the learning process. The earliest study unearthed by the current researcher in which the calculator had a pedagogical purpose was published in 1985. In fact, two-thirds of the studies integrated by the current meta-analysis involved an active teaching and learning role for the calculator. Based on these findings, it appears the role of the calculator has changed since the mid-1980's.

Presentation of Findings

To study the effect of the calculator on student achievement and attitude in the mathematics classroom, 305 effect sizes were available for statistical analysis. For the succeeding discussion, the effects were divided according to the research questions outlined in chapter one. The methods of data analysis described in chapter four were applied to each set of effect sizes. All calculations were conducted with the software package MetaWin 2.0. All graphs, including schematic box-and-whisker plots, were generated in Minitab 12.0. Every test for homogeneity was conducted at the 5% level of significance. For the appropriate effect size means, 95% confidence intervals were generated.

An analysis summary is provided for each research question. All effect sizes involved in the initial stage of analysis are listed in ascending order at the top of the analysis summary. Each effect size satisfying the definition of outlier is identified with an "o" superscript. All tests for homogeneity were conducted without outliers. The summary contains the test results directly below the list of effect sizes. If the set was homogeneous, the mean effect size, $\overline{g}_W$, and corresponding confidence interval were calculated. If the confidence interval did not contain zero, the mean value was considered significantly different from zero. In the analysis summary, a "*" superscript highlights the confidence intervals corresponding to statistically significant mean effect sizes at the 5% level of significance. A fail-safe N value accompanies these confidence intervals.

If the test for homogeneity in the initial stage of analysis resulted in a heterogeneous set of effect sizes, the data was analyzed for homogeneous subsets. Data was partitioned according to significant independent variables (i.e. study characteristics) and outliers were removed. The subsets generated through the second stage analysis were either heterogeneous or homogeneous. The analysis summary displays a list of effect sizes for each subset. Results of the test for homogeneity, the mean effect size, and corresponding confidence interval accompany each homogeneous subset.

The heterogeneous subsets required a third stage of analysis. Data was partitioned according to another significant study characteristic and analysis followed the same format as described for the second stage. If at any point in data analysis a moderating variable was not available or a subset was too small for further partitioning, the mean of the heterogeneous subset was calculated for descriptive purposes. At most, three stages of analysis were conducted for each research question.

A schematic box-and-whisker plot of the initial set of effect sizes accompanies the discussion of each research question. Inner fences, which are 1½ box lengths, are marked "f". Outer fences are defined to be three times the length of the box. However, in the current study, outer fences were not necessary as none of the effect sizes were large enough to be graphed a distance of three box lengths beyond the box endpoints.

The results of data analysis for each research question are presented below. The first six research questions discuss the calculator's effect on the achievement construct. This construct was divided into three categories which represented the acquisition, retention, and transfer of mathematical skills. There are two research questions for each of the three aspects of achievement – one for operational skills and one for problem solving skills. Each achievement research question was evaluated for maintenance and extension effects. Maintenance effects are those involving paper-and-pencil posttests. Extension effects result from the experimental group having access to calculators during posttesting. Each maintenance and extension subcategory was evaluated in two different ways – analysis of effect sizes generated from all types of calculators and analysis of effect sizes generated from graphing calculators alone. This allowed the researcher to compare and contrast the effects of general calculators with the effects of graphing calculators.

Seven research questions are devoted to the attitude construct. The data for these questions was generated from student use of all types of calculators. The research question regarding students' attitudes toward mathematics provided a sufficient amount of data for a separate analysis of graphing calculator effects. One research question features students' estimation skills after calculator use. However, the lack of sufficient information made meaningful analysis impossible. The remaining two research questions could not be scrutinized through the medium of meta-analysis. One question highlights gender differences resulting from calculator use. The other question considers the influence of special calculator curriculum on students using calculators.

Research Question #1

What are the effects of calculators on the acquisition of composite operational skills?

a. What are the effects of calculators on the acquisition of computational skills?

b. What are the effects of calculators on the acquisition of conceptual skills?

Composite Operational Skills – Maintenance Results

All Calculator Types

Thirty-two effect sizes applied to the maintenance aspect of this research question. As the box-and-whisker plot in Figure 6 indicates there were no outliers. The effect sizes are listed at the top of Figure 7. During the initial stage

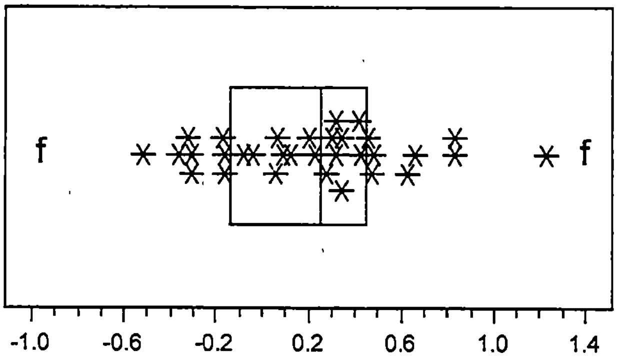


Figure 6. Box-and-Whisker Plot: Achievement in Acquiring Composite Operational Skills – Maintenance Effect Sizes

Stage 1: **Composite Operational Effect Sizes**

-0.5203	-0.3070	-0.0838	0.0891	0.2743	0.3397	0.4549	0.6589
-0.3672	-0.1675	-0.0381	0.1257	0.2994	0.3418	0.4717	0.8294
-0.3220	-0.1651	0.0551	0.2036	0.3149	0.4137	0.4839	0.8331
-0.3080	-0.1588	0.0694	0.2298	0.3181	0.4243	0.6284	1.2274
$H_T = 98.31 > \chi^2_{31} = 44.99$ Heterogeneous							

Stage 2: **Ability Groups**

Low Ability	Mixed Ability	High Ability
-0.0838 0.4137 0.4717 0.4839 0.6284	-0.3220 0.0551 0.2743 0.3418 -0.3070 0.0694 0.2994 0.4549 -0.1675 0.0891 0.3149 0.6589 -0.1651 0.2036 0.3181 0.8294 -0.1588 0.2298 0.3397 0.8331 -0.0381	-0.5203 -0.3672 -0.3080 0.1257 0.4243 1.2274°
$H_T = 7.01 < \chi^2_4 = 9.49$ Homogeneous $\bar{g}_W = 0.2973$ (0.1081, 0.4865)* N = 3	$H_T = 61.55 > \chi^2_{20} = 31.41$ Heterogeneous	$H_T = 8.23 < \chi^2_4 = 9.49$ Homogeneous $\bar{g}_W = -0.2333$ (-0.4760, 0.0095)

Stage 3: **Mixed Ability Group – Educational Divisions**

K – 5 Grades	6 – 8 Grades	9 – 12 Grades
-0.1675 0.2036 -0.1588 0.2298 -0.0381 0.6589 0.0694	-0.3220 0.3181 0.3418 0.8331	-0.3070 0.2994 -0.1651 0.3149 0.0551 0.3397 0.0891 0.4549 0.2743 0.8294°
$H_T = 12.52 < \chi^2_6 = 12.59$ Homogeneous $\bar{g}_W = 0.0699$ (-0.0206, 0.1604)	$H_T = 19.61 > \chi^2_3 = 7.81$ Hetrogeneous $\bar{g}_W = 0.3852$	$H_T = 12.80 < \chi^2_8 = 15.51$ Homogeneous $\bar{g}_W = 0.0790$ (0.0374, 0.1206)* N = 8

o : outlier , * : $\bar{g}_W$ significantly different from zero

Figure 7. Analysis Summary: Achievement in Acquiring Composite Operational Skills – Maintenance Effect Sizes

of analysis, the test for homogeneity revealed the group was heterogeneous (H_T = 98.31 > χ^2_{31} = 44.99).

A significant relationship between the effect of the calculator and student ability level was discovered in the second stage of analysis. The effect sizes were partitioned into three categories – low, mixed, and high ability. The low ability subset, containing effect sizes generated from students of below average ability, was homogeneous (H_T = 7.01 < χ^2_4 = 9.49). The mean effect of 0.2973 was statistically significant since the 95% confidence interval (0.1081, 0.4865) does not contain zero. The fail-safe N for the low ability group was N = 3. Therefore, if null results from three studies were included in this low ability subset, the mean effect size would no longer be statistically significant.

The high ability subset featured students of above average ability. Analysis of this subset revealed an outlier. After its removal, the remaining set of effect sizes was homogeneous (H_T = 8.23 < χ^2_4 = 9.49) with a mean effect of -0.2333. This value was not statistically significant since the confidence interval (-0.4760, 0.0095) includes zero. Effect sizes for the mixed ability group were gleaned from studies in which students of low or high ability were not separated from the rest of the class during data collection. Therefore, the mixed ability group represents the range of abilities found in a typical classroom. The subset was heterogeneous (H_T = 61.55 > χ^2_{20} = 31.41) and thus required a third stage of analysis.

The moderating variable for the mixed ability effect sizes was determined to be the separation of grade levels into educational divisions. The elementary division contained kindergarten through fifth grade. Sixth through eighth grades were featured by the middle school division. The high school division consisted of grades nine through twelve. The elementary and high school mixed ability subsets were homogeneous. The elementary subset generated a non-significant mean effect of 0.0699. An outlier was removed from the high school subset resulting in a mean effect of 0.0790. With a confidence interval of (0.0374, 0.1206), this value is significantly different from zero. A fail-safe N was calculated and resulted in a value of N = 8. Therefore, if the null results of eight studies were integrated with the existing studies, the mean effect size would be reduced to a value that would be statistically insignificant.

The middle school subset was heterogeneous. With such a small set of values, further partitioning by another study characteristic was not worthwhile. The mean effect of 0.3852 is provided as a descriptive statistic.

Graphing Calculator Only

For the acquisition of composite operational skills in the maintenance sense, ten effect sizes resulted from studies of graphing calculator use. Figure 8 contains the analysis summary of these values. After the removal of an outlier, a homogeneous ($H_T = 15.06 < \chi_8^2 = 15.51$) set of effects remained. The mean effect was 0.1825. Since the 95% confidence interval (0.0408, 0.3243) does not contain zero, this value is statistically significant. The fail-safe N value was N = 3. The null results of three studies would negatively affect the mean effect size.

Stage 1: **Composite Operational Effect Sizes**

-0.3672°	-0.1651	0.3149	0.8294
-0.3080	0.2743	0.3397	
-0.3070	0.2994	0.4549	

$H_T = 15.06 < \chi^2_8 = 15.51$
Homogeneous
$\bar{g}_W = 0.1825$
(0.0408, 0.3243)*
N = 3

o : outlier , * : $\bar{g}_W$ significantly different from zero

Figure 8. Analysis Summary: Achievement in Acquiring Composite Operational Skills – Maintenance Effect Sizes – Graphing Calculators

Composite Operational Skills – Extension Results

All Calculator Types

Twenty-seven effect sizes regarding the acquisition of composite operational skills were integrated by the first stage of analysis. The box-and-whisker plot in Figure 9 displays the spread of the data. A clustering of values exists at the lower end of the box. An outlier, -0.5223, lies to the left of the lower fence resulting in its removal before the test for homogeneity was conducted. The test indicated the set of effects was heterogeneous (H_T = 80.38 > χ^2_{25} = 37.65).

The second stage of analysis produced two homogeneous subsets. Grade levels separated according to educational divisions had an influence on homogeneity. Only one effect size was generated from a grade in the elementary division. Therefore, a separate subset for this single effect size was

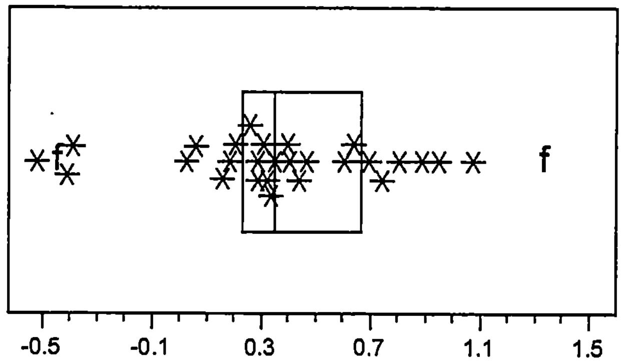


Figure 9. Box-and-Whisker Plot: Achievement in Acquiring Composite Operational Skills – Extension Effect Sizes

not practical. The elementary value and the effects from grades 6 – 8 were combined into an elementary and middle school educational division. As depicted in Figure 10, after the removal of an outlier, the remaining effect sizes were homogeneous (H_T = 12.98 < χ^2_7 = 14.07) with a statistically significant mean of 0.3835. The fail-safe N was determined to be N = 12. Therefore, the null results of twelve studies could be added to the existing data before the mean effect size would fail the test of statistical significance. The remaining subset, which contained grades 9 – 12, was homogeneous (H_T = 24.92 < χ^2_{15} = 25.00) following the removal of an outlier. The mean effect for the upper grades, $\overline{g}_W$= 0.2811, was statistically significant. The fail-safe N calculation resulted in a value of N = 139. Hence, it would take the null results of 139 studies to negatively affect the mean effect size.

<u>Graphing Calculator Only</u>

Twelve graphing calculator effects were available for analysis. They are listed at the top of Figure 11. The set remained heterogeneous (H_T = 21.83 >

Stage 1: **Composite Operational Effect Sizes**

-0.5223°	0.0599	0.2598	0.3221	0.4044	0.6443	0.8880
-0.4067	0.1463	0.2852	0.3359	0.4415	0.6976	0.9554
-0.3868	0.1813	0.2911	0.3478	0.4715	0.7439	1.0747
0.0248	0.2037	0.3107	0.3980	0.6087	0.8046	

$H_T = 80.38 > \chi^2_{25} = 37.65$
Heterogeneous

Stage 2: **Educational Divisions**

K – 8 Grades		9 – 12 Grades			
-0.4067	0.6087	-0.3868	0.2598	0.3980	0.8880
0.2037	0.6443	0.0248	0.2852	0.4415	1.0747°
0.3359	0.8046	0.0599	0.2911	0.4715	
0.3478	0.9554°	0.1463	0.3107	0.6976	
0.4044		0.1813	0.3221	0.7439	

K – 8 Grades	9 – 12 Grades
$H_T = 12.98 < \chi^2_7 = 14.07$	$H_T = 24.92 < \chi^2_{15} = 25.00$
Homogeneous	Homogeneous
$\bar{g}_W = 0.3835$	$\bar{g}_W = 0.2811$
(0.2281, 0.5389)*	(0.2476, 0.3146)*
N = 12	N = 139

o : outlier , * : $\bar{g}_W$ significantly different from zero

Figure 10. Analysis Summary: Achievement in Acquiring Composite Operational Skills – Extension Effect Sizes

Stage 1: **Composite Operational Effect Sizes**

-0.4067	0.3107	0.4415	0.8046
0.0248	0.3221	0.4715	0.8880
0.1813	0.4044	0.7439	1.0747°
$H_T = 21.83 > \chi^2_{10} = 18.31$ Heterogeneous			

Stage 2: **Educational Divisions**

K – 8 Grades	9 – 12 Grades	
-0.4067	0.0248	0.4415
0.4044	0.1813	0.4715
0.8046	0.3107	0.7439
	0.3221	0.8880
$H_T = 9.81 > \chi^2_2 = 5.99$ Heterogeneous $\bar{g}_W = 0.3520$	$H_T = 12.02 < \chi^2_7 = 14.07$ Homogeneous $\bar{g}_W = 0.3376$ (0.1779, 0.4974)* N = 9	

o : outlier , * : $\bar{g}_W$ significantly different from zero

Figure 11. Analysis Summary: Achievement in Acquiring Composite Operational Skills – Extension Effect Sizes – Graphing Calculators

χ^2_{10} = 18.31) after the removal of an outlier. Similar to the data for this question generated from the use of all types of calculators, the second stage of analysis revealed homogeneity was affected by grade levels according to educational divisions. Three effect sizes were generated from studies involving pre-high school students. These values were heterogeneous (H_T = 9.81 > χ^2_2 = 5.99) with a descriptive mean of 0.3520. The remaining eight values were derived from studies featuring high school students. The subset was homogeneous (H_T = 12.02 < χ^2_7 = 14.07) with a significant mean of 0.3376. The fail-safe N = 9 revealed the mean would be affected by nine studies with null results.

a. What are the effects of calculators on the acquisition of computational skills?

<u>**Computational Skills – Maintenance Results**</u>

<u>All Calculator Types</u>

The acquisition of computational skills in the maintenance sense provided 37 effect sizes for the first stage of analysis. The graph in Figure 12 describes a

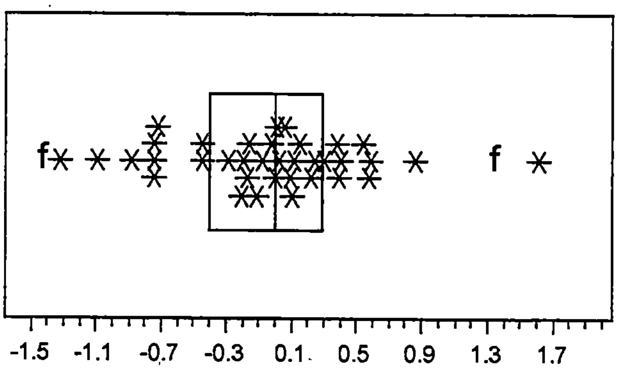


Figure 12. Box-and-Whisker Plot: Achievement in Acquiring Computational Skills – Maintenance Effect Sizes

fairly normal distribution with one outlier. The outlier was removed but the test for homogeneity revealed a heterogeneous set (H_T = 126.60 > χ^2_{35} = 49.80). This result led to a second stage of analysis in which student ability level was discovered to be a moderating variable. The data was partitioned into high, mixed, and low ability subsets. The low (H_T = 9.03 < χ^2_6 = 12.59) and mixed (H_T = 25.36 < χ^2_{16} = 27.14) ability subsets were both homogeneous. As displayed in Figure 13, the low ability group had a statistically significant mean of 0.2978. The fail-safe N calculation resulted in a value of N = 10. Therefore, the mean effect size would be negatively affected by null results from ten additional studies. The mixed ability group's mean of 0.0838 was not significant since the confidence interval (-0.0088, 0.1764) contains zero. The high ability group was heterogeneous (H_T = 28.38 > χ^2_9 = 16.92) so a third stage of analysis was attempted. Unfortunately, statistical analysis was unable to uncover a significant relationship between calculator effects with respect to high ability students and a study characteristic. The descriptive mean for this subset was -0.3067.

Graphing Calculator Only

Only four maintenance effect sizes regarding the acquisition of computational skills involved the graphing calculator. They are listed in Figure 14. This small set was homogeneous (H_T = 6.15 < χ^2_3 = 7.81) in the first phase of analysis. The mean effect of -0.2670 was significant since the confidence interval (-0.5207, -0.0132) does not contain zero. The fail-safe N value, N = 2, revealed the mean would be significantly affected by two studies with null results.

Stage 1: **Computational Effect Sizes**

-1.3260	-0.7208	-0.1720	-0.0012	0.0969	0.2424	0.5399
-1.0938	-0.4437	-0.1632	0.0104	0.0989	0.3012	0.5742
-0.8811	-0.4416	-0.1189	0.0125	0.1213	0.3752	0.5948
-0.7443	-0.2884	-0.0849	0.0220	0.1470	0.3897	0.8652
-0.7443	-0.2050	-0.0198	0.0622	0.2242	0.4023	1.6093[o]
-0.7424	-0.1974					

$H_T = 126.60 > \chi^2_{35} = 49.80$
Heterogeneous

Stage 2: **Ability Groups**

Low Ability		Mixed Ability			High Ability	
-1.0938[o]	0.0969	-0.7443	-0.0849	0.1470	-1.3260	-0.2050
-0.8811	0.4023	-0.7424	-0.0198	0.2242	-0.7443	0.0220
-0.2884	0.5742	-0.1974	0.0104	0.2424	-0.7208	0.0989
-0.0012	0.5918	-0.1720	0.0125	0.3012	-0.4437	0.3897
		-0.1632	0.0622	0.3752	-0.4416	0.5399
		-0.1189	0.1213	0.8652[o]		
$H_T = 9.03 < \chi^2_6 = 12.59$ Homogeneous $\bar{g}_W = 0.2978$ (0.0781, 0.5176)* N = 10		$H_T = 25.36 < \chi^2_{16} = 27.14$ Homogeneous $\bar{g}_W = 0.0838$ (-0.0088, 0.1764)			$H_T = 28.38 > \chi^2_9 = 16.92$ Heterogeneous $\bar{g}_W = -0.3067$	

o : outlier , * : $\bar{g}_W$ significantly different from zero

Figure 13. Analysis Summary: Achievement in Acquiring Computational Skills – Maintenance Effect Sizes

Stage 1: **Computational Effect Sizes**

-0.7443	-0.4416	-0.2050	0.1213
$H_T = 6.15 < \chi^2_3 = 7.81$ Homogeneous $\bar{g}_W = -0.2670$ (-0.5207, -0.0132)* N = 2			

o : outlier , * : $\bar{g}_W$ significantly different from zero

Figure 14. Analysis Summary: Achievement in Acquiring Computational Skills – Maintenance Effect Sizes – Graphing Calculators

Computational Skills – Extension Results

All Calculator Types

The acquisition of computational skills in the sense of extension provided 22 effect sizes for analysis. The distribution of the data is portrayed in the box-and-whisker plot in Figure 15. Even after the removal of an outier, the set was heterogeneous ($H_T = 65.35 > \chi^2_{20} = 31.41$). Second stage analysis revealed a significant relationship between effect size and student ability level. Effect sizes generated from studies involving students of low ability were separated from the high and mixed ability values. The subset was homogeneous ($H_T = 8.43 < \chi^2_6 = 12.59$) with a significant mean effect of 0.5139. While this mean value was fairly large, the fail-safe N calculation of N = 8 meant the value would be negatively affected by eight studies with null results. Based on statistical evaluation and a small number of values, creating a subset for the high ability effect sizes was not

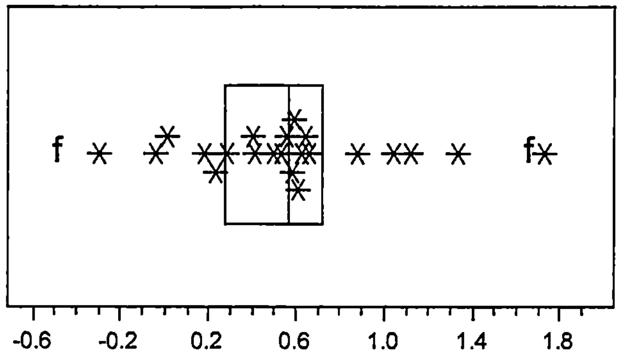


Figure 15. Box-and-Whisker Plot: Achievement in Acquiring Computational Skills – Extension Effect Sizes

practical. Therefore, as displayed in Figure 16, the test for homogeneity was administered to a subset of mixed ability effects which included a few high ability values. This subset contained two outliers. After their removal, the subset was homogeneous (H_T = 16.29 < χ^2_{11} = 19.68). The mean effect size, 0.3210, was statistically significant. The fail-safe N of N = 20 was fairly large for this mean value. Therefore, it would take the null results of twenty studies to negatively affect the statistically significant mean generated with the provided data.

Graphing Calculator Only

The four extension effect sizes regarding the acquistion of computational skills are listed in Figure 17. In spite of the small number of values, 1.7305 satisfied the definition of outlier and was removed before the test of homogeneity was conducted. The remaining three values were homogeneous (H_T = 5.57 < χ^2_2 = 5.99). However the mean effect, 0.0915, was not statistically significant.

Stage 1: **Computational Effect Sizes**

-0.2973	0.1774	0.4028	0.5350	0.5957	0.6436	1.0413	1.7305^{o}
-0.0382	0.2342	0.4143	0.5581	0.6094	0.6637	1.1210	
0.0101	0.2862	0.4969	0.5796	0.6239	0.8780	1.3372	

$H_T = 65.35 > \chi^2_{20} = 31.41$
Heterogeneous

Stage 2: **Ability Groups**

Low Ability		Mixed Ability			
0.0101	0.6436	-0.2973	0.2862	0.5581	1.0413^{o}
0.4028	0.8780	-0.0382	0.4143	0.5957	1.3372^{o}
0.5796	1.1210	0.1774	0.4969	0.6239	
0.6094		0.2342	0.5350	0.6637	
$H_T = 8.43 < \chi^2_6 = 12.59$ Homogeneous $\bar{g}_W = 0.5139$ (0.2649, 0.7628)* N = 8		$H_T = 16.29 < \chi^2_{11} = 19.68$ Homogeneous $\bar{g}_W = 0.3210$ (0.1826, 0.4594)* N = 20			

o : outlier , * : $\bar{g}_W$ significantly different from zero

Figure 16. Analysis Summary: Achievement in Acquiring Computational Skills – Extension Effect Sizes

Stage 1: **Computational Effect Sizes**

-0.2973	0.2862	0.5957	1.7305°
$H_T = 5.57 < \chi_2^2 = 5.99$ Homogeneous $\bar{g}_W = 0.0915$ (-0.2243, 0.4073)			

o : outlier , * : $\bar{g}_W$ significantly different from zero

Figure 17. Analysis Summary: Achievement in Acquiring Computational Skills – Extension Effect Sizes – Graphing Calculators

b. What are the effects of calculators on the acquisition of conceptual skills?

Conceptual Skills – Maintenance Results

All Calculator Types

For the acquisition of conceptual skills, eighteen maintenance effect sizes were available for statistical evaluation. The values are listed at the top of Figure 18. As portrayed in the box-and-whisker plot in Figure 19, two effect sizes were plotted to the right of the upper fence. The sixteen effect sizes lying within the fences were homogeneous ($H_T = 20.77 < \chi_{15}^2 = 25.00$). The confidence interval (-0.0256, 0.1379) contains zero. Therefore, the mean effect size, 0.0562, was not statistically significant.

Graphing Calculator Only

With regards to the graphing calculator, five effect sizes were available to address this question in the maintenance sense. The values, listed in Figure 20,

Stage 1: **Conceptual Effect Sizes**

-0.4214	-0.2154	-0.1490	0.1100	0.1916	0.4272
-0.4135	-0.1956	0.0198	0.1189	0.2309	0.8898°
-0.3609	-0.1920	0.0236	0.1906	0.3365	0.9066°

$H_T = 20.77 < \chi^2_{15} = 25.00$
Homogeneous
$\bar{g}_W = 0.0562$
(-0.0256, 0.1379)

o : outlier , * : $\bar{g}_W$ significantly different from zero

Figure 18. Analysis Summary: Achievement in Acquiring Conceptual Skills – Maintenance Effect Sizes

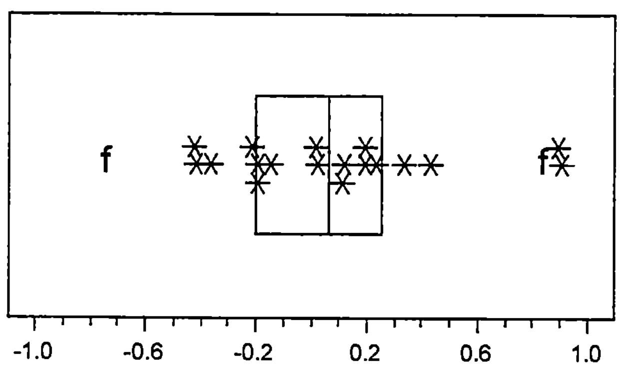


Figure 19. Box-and-Whisker Plot: Achievement in Acquiring Conceptual Skills – Maintenance Effect Sizes

Stage 1: **Conceptual Effect Sizes**

-0.4214	-0.4135	-0.2154	-0.1920	0.4272
$H_T = 5.99 < \chi^2_4 = 9.49$ Homogeneous $\bar{g}_W = 0.0277$ (-0.2362, 0.2915)				

o : outlier , * : $\bar{g}_W$ significantly different from zero

Figure 20. Analysis Summary: Achievement in Acquiring Conceptual Skills – Maintenance Effect Sizes – Graphing Calculators

were homogeneous ($H_T = 5.99 < \chi^2_4 = 9.49$). The mean effect size was 0.0277. However, this value was not significant since zero exists in the confidence interval (-0.2362, 0.2915).

Conceptual Skills – Extension Results

All Calculator Types

There were nineteen effect sizes regarding the extension aspect of the acquisition of conceptual skills. One effect size was extremely large and quickly determined to be an outlier. As Figure 21 reveals, the remaining eighteen effects were heterogeneous ($H_T = 54.92 > \chi^2_{17} = 27.59$). The spread of the data favored the lower half of the box. This is graphically represented by Figure 22. Upon further analysis, the relationship between effect size and student ability level was found to be significant. Since only one effect size was available to represent the low ability group, a separate category could not be established. Therefore, this value was included with the mixed ability effect sizes. After removal of an outlier,

Stage 1: **Conceptual Effect Sizes**

-0.1198	0.1374	0.2859	0.6294	0.8518
-0.0651	0.2059	0.3041	0.6463	1.3664
0.0123	0.2315	0.3301	0.7084	2.0341°
0.0454	0.2627	0.5673	0.7209	

$H_T = 54.92 > \chi^2_{17} = 27.59$
Heterogeneous

Stage 2: **Ability Groups**

Mixed Ability			High Ability	
-0.1198	0.1374	0.3301	0.2859	0.6463
-0.0651	0.2059	0.6294	0.3041	0.7084
0.0123	0.2315	0.7209	0.5673	1.3664°
0.0454	0.2627	0.8518°		
$H_T = 16.54 < \chi^2_{10} = 18.31$ Homogeneous $\bar{g}_W = 0.1553$ (0.0674, 0.2432)* N = 8			$H_T = 3.44 < \chi^2_{4} = 9.49$ Homogeneous $\bar{g}_W = 0.3809$ (0.2126, 0.5492)* N = 6	

o : outlier , * : $\bar{g}_W$ significantly different from zero

Figure 21. Analysis Summary: Achievement in Acquiring Conceptual Skills – Extension Effect Sizes

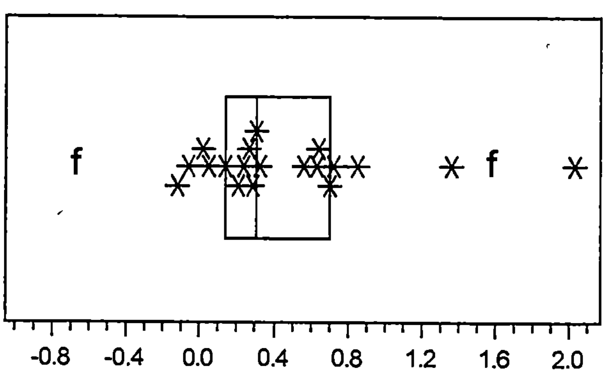


Figure 22. Box-and-Whisker Plot: Achievement in Acquiring Conceptual Skills – Extension Effect Sizes

the subset was homogeneous (H_T = 16.54 < χ^2_{10} = 18.31) with a statistically significant mean of 0.1553. Since the fail-safe N = 8, this value would be drawn out of significance by the inclusion of null results from eight studies. The high ability subset also contained an outlier. Once it was removed, the data was homogeneous (H_T = 3.44 < χ^2_4 = 9.49) and the a statistically significant mean effect size, 0.3809, was calculated. The fail-safe N calculation resulted in a value of N = 6. Therefore, the null results of six studies featuring the calculator's effect on students' conceptual skills in the sense of extension would negatively affect the mean effect size.

Graphing Calculator Only

Eleven of the original nineteen effect sizes generated for this research question resulted from the use of graphing calculators. These values were used to determine the effect of the graphing calculator on the acquisition of conceptual skills in the sense of extension. The same large effect size satisfying the definition of outlier in the above analysis was an outlier in this smaller set of effects. However, once it was removed, the set was homogeneous (H_T = 16.43 < χ^2_9 = 16.92) with a statistically significant mean effect of 0.4806 and a fail-safe N of N = 21. This fairly large N value means the mean effect size will maintain its statistically significant status until the null results of twenty-one studies have been integrated with the existing data. The results of the analysis are summarized in Figure 23.

Stage 1: **Conceptual Effect Sizes**

-0.0651	0.3041	0.6463	1.3664
0.0454	0.5673	0.7084	2.0341°
0.2059	0.6294	0.7209	
$H_T = 16.43 < \chi_9^2 = 16.92$ Homogeneous $\overline{g}_W = 0.4806$ (0.2993, 0.6620)* N = 21			

o : outlier , * : $\overline{g}_W$ significantly different from zero

Figure 23. Analysis Summary: Achievement in Acquiring Conceptual Skills – Extension Effect Sizes – Graphing Calculators

Research Question #2

What are the effects of calculators on the acquisition of composite problem solving skills?

a. What are the effects of calculators on the acquisition of problem solving productivity skills?

b. What are the effects of calculators on the acquisition of problem solving selectivity skills?

Composite Problem Solving Skills – Maintenance Results

All Calculator Types

Twenty-one effect sizes were available for analysis of composite problem solving skills in the maintenance sense. A box-and-whisker plot of the distribution is displayed in Figure 24. The □ represents an outlier found through the calculation of standardized residuals. The details of this method are provided

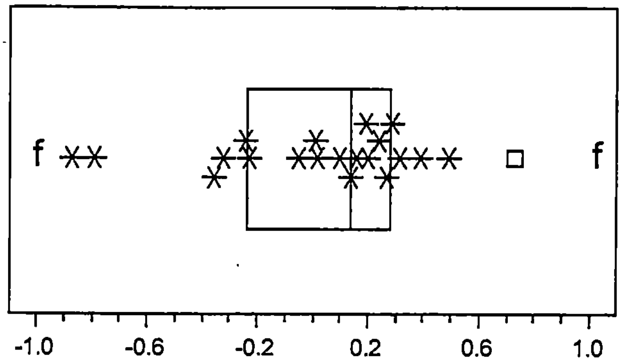


Figure 24. Box-and-Whisker Plot: Achievement in Acquiring Problem Solving Skills – Maintenance Effect Sizes

in chapter four. Once the outlier was removed, the remaining set of effects was homogeneous ($H_T = 26.71 < \chi^2_{19} = 30.14$) with a mean of 0.1160. The confidence interval, (0.0129, 0.2191), does not contain zero. Therefore, the mean was statistically significant. A fail-safe N of N = 11 resulted from the appropriate calculations. This number implies the fairly small mean value would be statistically insignificant with the addition of null results from eleven similar research studies. The results are summarized in Figure 25.

Graphing Calculator Only

None of the effect sizes for this research question were calculated from studies in which students used graphing calculators. Therefore, graphing calculator analysis could not be performed.

Stage 1: **Composite Problem Solving Effect Sizes**

-0.8724	-0.3225	-0.0481	0.0962	0.1940	0.2701	0.3877
-0.7880	-0.2425	0.0123	0.1374	0.1999	0.2886	0.4898
-0.3551	-0.2286	0.0142	0.1582	0.2433	0.3158	0.7275°

$H_T = 26.71 < \chi^2_{19} = 30.14$
Homogeneous
$\bar{g}_W = 0.1160$
(0.0129, 0.2191)*
N = 11

o : outlier , * : $\bar{g}_W$ significantly different from zero

Figure 25. Analysis Summary: Achievement in Acquiring Problem Solving Skills – Maintenance Effect Sizes

Composite Problem Solving Skills – Extension Results

All Calculator Types

In the sense of extension, twenty-nine effect sizes were generated from studies of composite problem solving skills. The box-and-whisker plot in Figure 26 reveals the outlier removed during the initial phase of analysis. The remaining twenty-eight effect sizes were heterogeneous ($H_T = 84.90 > \chi^2_{27} = 40.11$). During the second stage of analysis, the moderating variable was determined to be student ability level. With the small number of effects resulting from studies of high ability students, the creation of a high ability subset was not worthwhile. Therefore, the mixed ability subset contained a few high ability effect sizes. The subset was heterogeneous ($H_T = 61.11 > \chi^2_{18} = 28.87$) and required a third stage of analysis. After an outlier was removed, the low ability group was

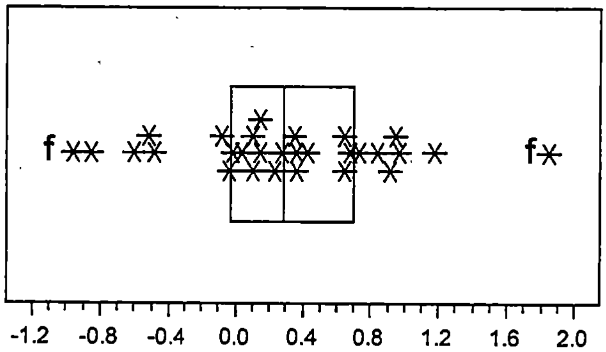


Figure 26. Box-and-Whisker Plot: Achievement in Acquiring Composite Problem Solving Skills – Extension Effect Sizes

homogeneous ($H_T = 13.91 < \chi^2_7 = 14.07$). However, the mean effect, -0.1367, was not significant.

A significant relationship between effect size and the educational divisions of grade levels was found through the third stage of analysis. The values were divided into two categories, one containing effect sizes from the elementary division and the other containing the effect sizes from middle and high school divisions. As revealed in Figure 27, the elementary subset was heterogeneous ($H_T = 24.57 > \chi^2_6 = 12.59$). Further analysis was attempted, but evidence did not exist of a relationship between elementary effect sizes and another independent variable. The mean effect of 0.1020 is provided as a descriptive statistic. The subset containing middle and high school grades was homogeneous ($H_T = 12.51 < \chi^2_{10} = 18.31$) after the removal of an outlier. The confidence interval, (0.0717, 0.3214), proves the mean effect of 0.1965 was statistically significant. The fail-safe N calculation was conducted and resulted in a value of N = 10. Therefore,

Stage 1: **Composite Problem Solving Effect Sizes**

-0.9622	-0.0773	0.1059	0.3527	0.6482	0.9433
-0.8548	-0.0440	0.1430	0.3583	0.6769	0.9732
-0.6017	-0.0130	0.1453	0.3637	0.7293	1.1795
-0.5136	0.0329	0.2290	0.4241	0.8363	1.8421^{o}
-0.4814	0.1026	0.2846	0.6481	0.9138	

$H_T = 84.90 > \chi^2_{27} = 40.11$
Heterogeneous

Stage 2: **Ability Groups**

Low Ability

-0.9622	0.2290
-0.6017	0.3527
-0.4814	0.7293
-0.0130	1.1795^{o}
0.1059	

$H_T = 13.91 < \chi^2_{7} = 14.07$
Homogeneous
$\bar{g}_W = -0.1367$
(-0.3799, 0.1066)

Mixed Ability

-0.8548	0.1026	0.3637	0.8363
-0.5136	0.1430	0.4241	0.9138
-0.0773	0.1453	0.6481	0.9433
-0.0440	0.2846	0.6482	0.9732
0.0329	0.3583	0.6769	

$H_T = 61.11 > \chi^2_{18} = 28.87$
Heterogeneous

Stage 3: Average and High Abilities Group – Educational Institution Levels

K – 5 Grades

-0.8548	0.8363
-0.5136	0.9138
0.2846	0.9433
0.6482	

$H_T = 24.57 > \chi^2_{6} = 12.59$
Heterogeneous
$\bar{g}_W = 0.1020$

6 – 12 Grades

-0.0773	0.1430	0.4241
-0.0440	0.1453	0.6481
0.0329	0.3583	0.6769
0.1026	0.3637	0.9732^{o}

$H_T = 12.51 < \chi^2_{10} = 18.31$
Homogeneous
$\bar{g}_W = 0.1965$
(0.0717, 0.3214)*
N = 10

o : outlier , * : $\bar{g}_W$ significantly different from zero

Figure 27: Analysis Summary: Achievement in Acquiring Composite Problem Solving Skills – Extension Effect Sizes

the addition of null results from ten studies would negatively affect the mean effect size.

Graphing Calculator Only

Only one effect size was generated from an extension-based, graphing calculator study. Therefore, graphing calculator effects on composite problem solving skills could not be determined through meta-analysis.

a. What are the effects of calculators on the acquisition of problem solving productivity skills?

Problem Solving Productivity Skills – Maintenance Results

All Calculator Types & Graphing Calculator Only

This question could not be addressed in the maintenance sense, since no effect sizes were available for analysis.

Problem Solving Productivity Skills – Extension Results

All Calculator Types

Problem solving productivity skills in the sense of extension were represented by three effect sizes. They are listed at the top of Figure 28. With such a small number of values, a box-and-whisker plot was not produced. The set was homogeneous ($H_T = 1.42 < \chi_2^2 = 5.99$) with a statistically significant mean effect of 0.2339. The fail-safe N calculation was conducted and resulted in a value of N = 4. Hence, four studies with null results would alter the statistically significant mean and result in the generation of an insignificant value.

Stage 1: **Problem Solving Productivity Effect Sizes**

0.1939	0.2729	0.3554
	$H_T = 1.42 < \chi_2^2 = 5.99$ Homogeneous $\bar{g}_W = 0.2339$ (0.1373, 0.3305)* N = 4	

o : outlier , * : $\bar{g}_W$ significantly different from zero

Figure 28. Analysis Summary: Achievement in Acquiring Problem Solving Productivity Skills – Extension Effect Sizes

Graphing Calculator Only

All three values available for this question were generated from studies which involved basic and scientific calculators. Therefore, graphing calculator analysis could not be conducted.

b. What are the effects of calculators on the acquisition of problem solving selectivity skills?

Problem Solving Selectivity Skills – Maintenance Results

All Calculator Types

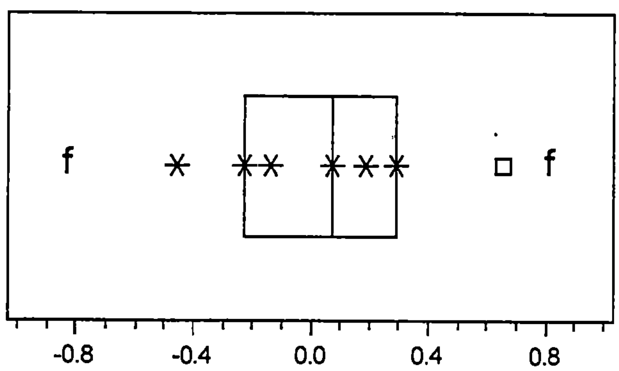
Figure 29 contains the seven effect sizes regarding the acquisition of problem solving selectivity skills in the maintenance sense. One outlier was found with the Hedges and Olkin (1985) standardized residual method. It is represented by the □ in Figure 30. After this outlier was removed, the set was homogeneous ($H_T = 3.42 < \chi_5^2 = 11.07$). Based on the confidence interval (-0.0507, 0.2114), the mean effect size, 0.0803, was not statistically significant.

Stage 1: **Problem Solving Selectivity Effect Sizes**

-0.4554	-0.1369	0.2902	0.6522°
-0.2261	0.0705	0.1872	
$H_T = 3.42 < \chi^2_5 = 11.07$ Homogeneous $\bar{g}_W = 0.0803$ (-0.0507, 0.2114)			

o : outlier , * : $\bar{g}_W$ significantly different from zero

Figure 29. Analysis Summary: Achievement in Acquiring Problem Solving Selectivity Skills – Maintenance Effect Sizes



□ : outlier by residual method

Figure 30. Box-and-Whisker Plot: Achievement in Acquiring Problem Solving Selectivity Skills – Maintenance Effect Sizes

Graphing Calculator Only

No graphing calculator effect sizes were available regarding this question. Therefore, a statistical evaluation could not be performed.

Problem Solving Selectivity Skills – Extension Results

All Calculator Types

The fourteen effect sizes used to analyze the relationship between problem solving selectivity skills and calculators are listed at the top of Figure 31. The data is graphically displayed in Figure 32. The set was homogeneous ($H_T = 9.30 < \chi^2_{13} = 22.36$) during the first phase of analysis. The confidence interval (0.0426, 0.3257) does not contain zero. Therefore, the mean effect size of 0.1841 was statistically significant. The fail-safe N for this data was N = 4. With the inclusion of null results from four studies, the statistically significant mean effect size would be converted to an insignificant value.

Stage 1: **Problem Solving Selectivity Effect Sizes**

-0.2876	-0.1546	-0.0796	0.1565	0.2838	0.3904	0.5877
-0.2290	-0.0911	0.0780	0.2046	0.2952	0.4581	0.6527

$H_T = 9.30 < \chi^2_{13} = 22.36$
Homogeneous
$\bar{g}_W = 0.1841$
(0.0426, 0.3257)*
N = 4

o : outlier , ∗ : $\bar{g}_W$ significantly different from zero

Figure 31. Analysis Summary: Achievement in Acquiring Problem Solving Selectivity Skills – Extension Effect Sizes

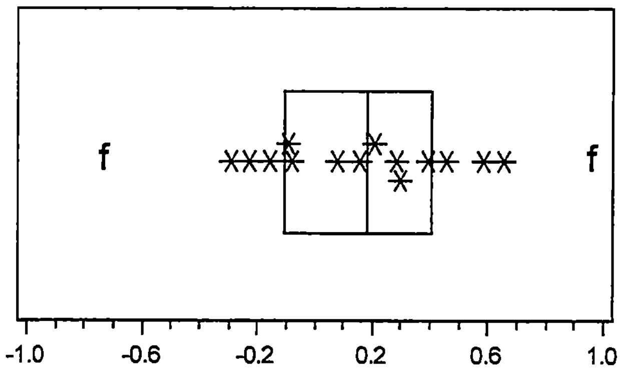


Figure 32. Box-and-Whisker Plot: Achievement in Acquiring Problem Solving Selectivity Skills – Extension Effect Sizes

Graphing Calculator Only

Only two effect sizes were available for statistical analysis with regard to the graphing calculator perspective. Therefore, analysis was not attempted.

Research Question #3

What are the effects of calculators on the retention of operational skills?

Operational Skills – Maintenance Results

All Calculator Types

Four effect sizes were available to evaluate the retention of operational skills in the maintenance sense. They are listed at the top of Figure 33. Due to the size of this data set, a box-and-whisker plot was not created. The first stage of analysis revealed the set was homogeneous ($H_T = 0.11 < \chi^2_3 = 7.81$). The mean effect size of -0.1381 was not statistically significant since the confidence interval (-0.3902, 0.1139) contains zero.

Stage 1: **Operational Effect Sizes**

-0.2297	-0.1446	-0.1376	-0.0973
$H_T = 0.11 < \chi_3^2 = 7.81$ Homogeneous $\overline{g}_W = -0.1381$ (-0.3902, 0.1139)			

o : outlier , * : $\overline{g}_W$ significantly different from zero

Figure 33. Analysis Summary: Achievement in Retaining Operational Skills – Maintenance Effect Sizes

Graphing Calculator Only

Due to a lack of sufficient data, the effect of the graphing calculator on the retention of operational skills in the maintenance sense could not be evaluated through meta-analysis.

Operational Skills – Extension Results

All Calculator Types

The relationship between calculators and the retention of operational skills was represented by five effect sizes. They are listed at the top of Figure 34. A box-and-whisker plot was unnecessary for this small set of effects. None of the effect sizes satisfied the definition of outlier. The test for homogeneity revealed the data to be heterogeneous (H_T = 27.63 > χ_4^2 = 9.49). No moderating variable to explain the heterogeneity could be found. The mean effect of 0.3881 is provided as a descriptive statistic.

Stage 1: **Operational Effect Sizes**

-0.4035	0.2959	0.5245	0.8194	1.3153
$H_T = 27.63 > \chi^2_4 = 9.49$ Heterogeneous $\bar{g}_W = 0.3881$				

o : outlier , * : $\bar{g}_W$ significantly different from zero

Figure 34. Analysis Summary: Achievement in Retaining Operational Skills – Extension Effect Sizes

<u>Graphing Calculator Only</u>

Two effect sizes were available to address this research question from the graphing calculator perspective. Therefore, graphing calculator effects on the retention of operational skills in the sense of extension could not be statistically analyzed.

Research Question #4

What are the effects of calculators on the retention of problem solving skills?

<u>Problem Solving Skills – Maintenance Results</u>

<u>All Calculator Types & Graphing Calculator Only</u>

No effect sizes regarding the retention of problem solving skills were available for analysis. Therefore, the maintenance aspect of this research question could not be evaluated.

Problem Solving Skills – Extension Results

All Calculator Types

Four effect sizes participated in the analysis of problem solving skill retention. They are listed in Figure 35. No outliers existed in this small data set. A box-and-whisker plot was not generated for this small set of values. The test for homogeneity revealed the set was heterogeneous ($H_T = 33.99 > \chi^2_3 = 5.99$). Further analysis was unable to locate moderating variables to explain the heterogeneity. The mean, -0.0583, is provided as a descriptive statistic.

Graphing Calculator Only

Due to insufficient data, the effect of the graphing calculator on the retention of problem solving skills in the sense of extension could not be evaluated.

Stage 1: **Problem Solving Effect Sizes**

-1.2295	0.2460	0.3195	0.6863
$H_T = 33.99 > \chi^2_3 = 5.99$ Heterogeneous $\overline{g}_W = -0.0583$			

o : outlier , ∗ : $\overline{g}_W$ significantly different from zero

Figure 35. Analysis Summary: Achievement in Retaining Problem Solving Skills – Extension Effect Sizes

Research Question #5

What are the effects of calculators on the transfer of operational skills?

Operational Skills – Maintenance Results

All Calculator Types

None of the studies provided data regarding the transfer of operational skills. Therefore, the maintenance aspect of this question could not be analyzed.

Operational Skills – Extension Results

All Calculator Types

Only two effect sizes were available regarding the transfer of composite operational skills in the sense of extension. As a result, statistical analysis was not attempted.

Research Question #6

What are the effects of calculators on the transfer of problem solving skills?

Problem Solving Skills – Maintenance Results

All Calculator Types

Since transfer of problem solving skills was not represented through the effect size medium, statistical analysis could not be conducted.

Problem Solving Skills – Extension Results

All Calculator Types

Only one effect size was generated from a study of the transfer of problem solving skills. Thus, meta-analysis could not be performed on the extension aspect of this research question.

Research Question #7

What are the effects of calculators on students' estimation skills?

Due to insufficient data, no effect sizes related to this question could be addressed.

Research Question #8

What are the effects of calculators on students' attitude toward mathematics?

All Calculator Types

Twenty-three values addressed the effects of calculators on students' attitude toward mathematics. The box-and-whisker plot in Figure 36 displays the data. The set contained no outliers. During the initial stage of analysis, the set was heterogeneous ($H_T = 74.54 > \chi^2_{22} = 33.92$). As a result, a second stage of analysis was conducted. A significant relationship between calculator effect sizes and educational divisions was determined. The effects were partitioned into two subsets. As displayed in Figure 37, one subset contained the

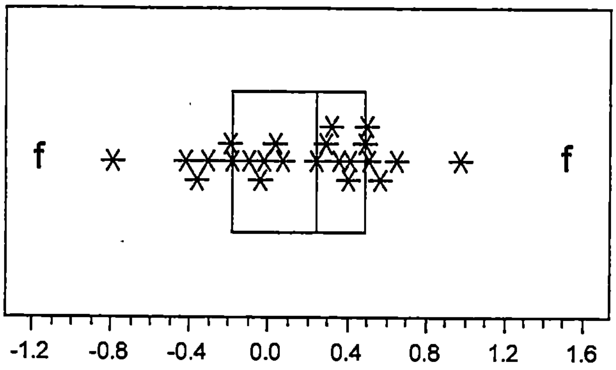


Figure 36: Box-and-Whisker Plot: Attitude Toward Mathematics

elementary and middle school divisions while the other subset contained the high school division. Two outliers were removed from the elementary and middle school subset. The remaining effect sizes were homogeneous ($H_T = 18.81 < \chi^2_{12} = 21.03$) with a mean value of 0.0481. The high school division was also found to be homogeneous ($H_T = 12.29 < \chi^2_6 = 12.59$) after the removal of one outlier. The mean effect for this group was 0.1052. Neither mean value was statistically significant since each of the corresponding confidence intervals contain zero.

Graphing Calculator Only

Seven effect sizes were generated from studies emphasizing the use of the graphing calculator. These are listed in Figure 38. The box-and-whisker plot in Figure 39 portrays the spread of the data. The set was homogeneous ($H_T = 10.66 < \chi^2_6 = 12.59$) with a statistically significant mean effect size of 0.3821. The fail-safe N was determined to be N = 8. Therefore, the null results of eight additional studies would result in a non-significant mean effect size.

Stage 1: **Attitude Toward Mathematics Effect Sizes**

-0.7940	-0.1945	-0.0256	0.2874	0.4119	0.5698
-0.4216	-0.1834	0.0296	0.3216	0.4882	0.6497
-0.3657	-0.0965	0.0732	0.3539	0.5028	0.9727
-0.3114	-0.0447	0.2403	0.4023	0.5074	
$H_T = 74.54 > \chi^2_{22} = 33.92$ Heterogeneous					

Stage 2: **Educational Divisions**

K – 8 Grades				9 – 12 Grades	
-0.7940°	-0.0447	0.2874	0.5028	-0.4216	0.4023
-0.3657	-0.0256	0.3216	0.5074	-0.3114	0.4882
-0.1945	0.0296	0.3539	0.9727°	-0.1834	0.5698
-0.0965	0.2403	0.4119		0.0732	0.6497°
$H_T = 18.81 < \chi^2_{12} = 21.03$ Homogeneous $\bar{g}_W = 0.0481$ (-0.0399, 0.1362)				$H_T = 12.29 < \chi^2_6 = 12.59$ Homogeneous $\bar{g}_W = 0.1052$ (-0.0588, 0.2691)	

o : outlier , * : $\bar{g}_W$ significantly different from zero

Figure 37: Analysis Summary: Attitude Toward Mathematics

Stage 1: **Attitude Toward Mathematics Effect Sizes**

-0.4216	-0.3114	0.3216	0.4023	0.4882	0.5698	0.6497
$H_T = 10.66 < \chi^2_6 = 12.59$ Homogeneous $\bar{g}_W = 0.3821$ (0.2049, 0.5993)* N = 8						

o : outlier , * : $\bar{g}_W$ significantly different from zero

Figure 38. Analysis Summary: Attitude Toward Mathematics – Graphing Calculator Effect Sizes

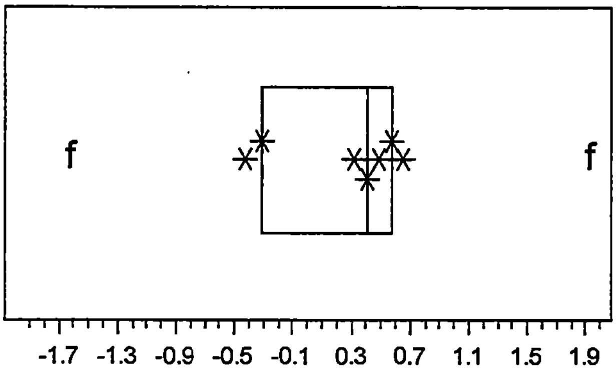


Figure 39. Box-and-Whisker Plot: Attitude Toward Mathematics – Graphing Calculator Effect Sizes

Research Question #9

What are the effects of calculators on students' attitude toward the use of the calculator in mathematics?

Six values were available to evaluate the relationship between calculator effects and students' attitude toward the use of calculators in mathematics. They are listed in Figure 40. As the box-and-whisker plot in Figure 41 reveals, the data contained no outliers. The small set was found to be heterogeneous ($H_T = 28.86 > \chi_5^2 = 11.07$). In spite of thorough analysis, no significant relationship between effect sizes and independent variables could be found. Therefore, the mean effect of -0.0784 is presented as a descriptive statistic.

Research Question #10

What are the effects of calculators on students' anxiety toward mathematics?

Only two effect sizes were available regarding students' anxiety toward mathematics. Therefore, statistical analysis was not attempted.

Stage 1: Attitude Toward Use of Calculators in Mathematics Effect Sizes

-0.7924	-0.4385	-0.4030	-0.3718	0.0388	0.7207
$H_T = 28.86 > \chi_5^2 = 11.07$ Heterogeneous $\bar{g}_W = -0.0784$					

o : outlier , * : $\bar{g}_W$ significantly different from zero

Figure 40. Analysis Summary: Attitude Toward Use of Calculators in Mathematics

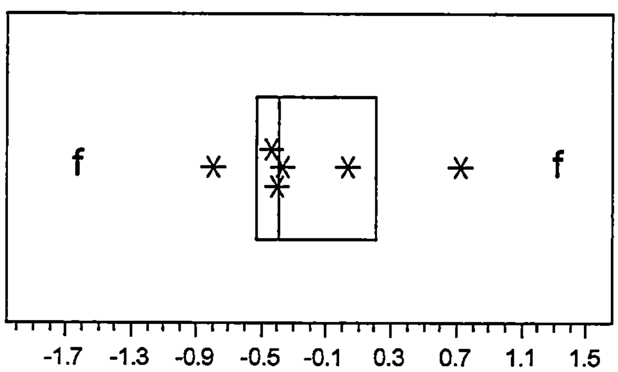


Figure 41. Box-and-Whisker Plot: Attitude Toward Use of Calculators in Mathematics

Research Question #11

What are the effects of calculators on students' self-concept in mathematics?

Six effect sizes addressed the relationship between self-concept in mathematics and calculator use. They are listed in Figure 42. In spite of the small size of this data set, one value was significantly larger than the others and satisfied the definition of outlier. This is portrayed in the box-and-whisker plot in Figure 43. After the removal of the outlier, the test for homogeneity was conducted. The data was homogeneous ($H_T = 4.33 < \chi_4^2 = 9.49$) with a non-significant mean effect of 0.0473.

Research Question #12

What are the effects of calculators on students' motivation to learn mathematics?

Data collection yielded two effect sizes for this research question. Because there was not enough data for meta-analysis calculations, the question could not be addressed.

Stage 1: **Self-Concept in Mathematics Effect Sizes**

-0.3582	-0.1546	0.0440	0.2214	0.3070	1.7462°
$H_T = 4.33 < \chi_4^2 = 9.49$ Homogeneous $\bar{g}_W = 0.0473$ (-0.0580, 0.1526)					

o : outlier , ∗ : $\bar{g}_W$ significantly different from zero

Figure 42. Analysis Summary: Self-Concept in Mathematics

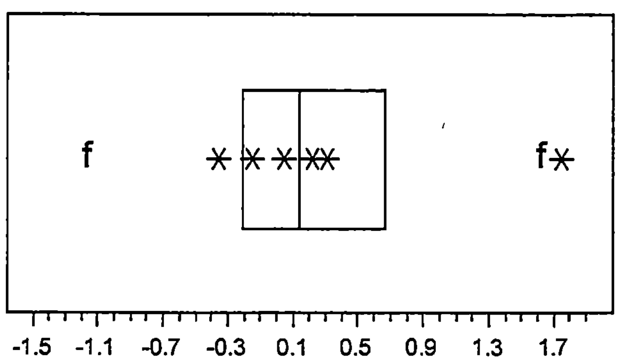


Figure 43. Box-and-Whisker Plot: Self-Concept in Mathematics

Research Question #13

What are the effects of calculators on students' attitude toward mathematics teachers?

One effect size was generated from a study which assessed the effect of calculators on students' attitude toward mathematics teachers. Therefore, statistical analysis could not be conducted.

Research Question #14

What are the effects of calculators on how students perceive the value of mathematics in society?

Three studies generated effect sizes for this research question. The values are listed in Figure 44. Due to the small size of this set, a box-and-whisker plot was not produced. The set was homogeneous ($H_T = 1.68 < \chi_2^2 = 5.99$) with a non-significant mean effect of -0.0372.

Stage 1: **Value of Mathematics as a Subject Effect Sizes**

-0.1157	-0.0871	0.3470
	$H_T = 1.68 < \chi_2^2 = 5.99$ Homogeneous $\overline{g}_W = -0.0372$ (-0.2678, 0.1935)	

o : outlier , ∗ : $\overline{g}_W$ significantly different from zero

Figure 44. Analysis Summary: Value of Mathematics as a Subject

The fourteen previous research questions were evaluated with inferential statistics methods. The remaining research questions contain confounding variables that are not conducive to the same type of evaluation. Descriptive remarks based on observational analysis of the data will be provided.

Research Question #15

Are the effects of calculators on achievement and attitude different for male and female students?

Five studies (see Appendix D) investigated gender-related differences of the calculator's effect on achievement and attitude. A vote-counting method was used to gather data. Three studies reported results favoring males and the other two studies reported results favoring females. In totality, these studies revealed the effects of calculators on achievement and attitude were slightly better for males than females. However, significant differences between males and females could not be determined with this descriptive technique evaluating such a small number of studies.

Research Question #16

What are the effects of calculators on achievement and attitude when special calculator curricula are involved?

Seventeen effect sizes resulted from calculator use with special curriculum materials. They were separated into five categories. The effect sizes and their corresponding means are listed in Figure 45. All categories contained a small number of effects. The composite problem solving skills category included seven effect sizes with the largest mean value of 0.7107. Only one effect size was available to represent the composite operational skills category. The smallest mean effect, 0.0011, belonged to computational skills in the sense of extension. Since this value was based on only two effect sizes, the mean has only minor influence on the evaluation of the results from this research question. The computational skills category, in the maintenance sense, was only slightly larger with three effect sizes. However, the mean of the data, 0.2659, was quite large. Attitude toward mathematics was represented by four effect sizes with a mean of 0.3122. The mean values for all categories were fairly large with the exception of 0.0011. For this observational analysis, composite problem solving skills in the sense of extension represented by the largest value, 0.7107, was the most significant finding.

Summary and Discussion

The "Presentation of Findings" described above contains a detailed explanation of the procedures used in analysis and the outcomes for each of the sixteen research questions outlined in chapter one. A summary of findings will now be presented in anticipation of drawing conclusions from the data.

The first six research questions featured calculator effects on achievement in basic operational skills and problem solving skills. In particular, the discussion

Effect Sizes Resulting from Special Curriculum Materials

Category	Effect Sizes		Mean Effect
Composite Operational Skills – Extension	1.0065		
Computational Skills – Extension	-0.0279	0.1474	0.0011
Computational Skills – Maintenance	-0.0468 0.1594	0.9864	0.2659
Composite Problem Solving Skills – Extension	-0.3483 0.0250 0.1795 0.2561	1.0820 1.0916 1.5208	0.7107
Attitude Toward Mathematics	0.0404 0.3238	0.3635 0.4246	0.3122

Figure 45. Analysis Summary: Achievement in Skills Acquisition and Attitude Toward Mathematics – Effects with Special Curriculum Materials

of these questions describes results related to the acquisition, retention, and transfer of skills. Summaries of the findings from these six research questions are presented in Figures 46 – 48. Figures 46 and 47 refer to all calculator types. Figure 48 summarizes the effects of graphing calculators. Statistically significant mean effect sizes are labeled with "*" superscripts. Values without stars are not significant since their confidence intervals contain zero. Mean effects resulting from heterogeneous data are descriptive statistics. Therefore, inferential explanations cannot be provided. The effects are identified with "+" superscripts.

Acquisition of Skills

RQ1 and RQ2 analyzed the acquisition of basic operational and problem solving skills. These questions considered calculator effects in terms of maintenance and extension.

1. <u>Maintenance Effects – All Calculator Types</u>

For composite operational skills, two significant effects were discovered. In particular, low ability students and mixed ability students in grades 9-12 generated positive mean effect sizes. Therefore, the basic skills of students in these groups improved as a result of calculator treatment. The results for high school, mixed ability students were most resistant to the inclusion of studies with null results since the fail-safe N value was N = 8. The low ability results could be more easily influenced with only three studies necessary to alter the statistical significance of the mean effect size. High ability students from all grade levels

Mean Effect Sizes Regarding Basic Operational Skills

Skill Type	Acquisition of Operational Skills	
	Maintenance	Extension
Composite Operational Skills	**Ability Level** Low: 0.2973* High: -0.2333 Mixed: (three values below) **Educational Divisions** K-5: 0.0699 6-8: 0.3852+ 9-12: 0.0790*	**Educational Divisions** K-8: 0.3835* 9-12: 0.2811*
Computation Skills	**Ability Level** Low: 0.2978* Mixed: 0.0838 High: -0.3067+	**Ability Level** Low: 0.5139* Mixed: 0.3210*
Conceptual Skills	0.0562	**Ability Level** Mixed: 0.1553* High: 0.3809*

Skill Type	Retention of Operational Skills		Transfer of Operational Skills	
	Maintenance	Extension	Maintenance	Extension
Composite Operational Skills	-0.1381	0.3881+	–	–

* : mean effect significantly different from zero, + : descriptive statistic

Figure 46. Summary: Achievement Effects Regarding Acquisition, Retention, and Transfer of Basic Operational Skills

Mean Effect Sizes Regarding Problem Solving Skills

Skill Type	**Acquisition of Problem Solving Skills**	
	Maintenance	Extension
Composite Problem Solving Skills	0.1160*	**Ability Level** Low: -0.1367 Mixed: (two values below) **Educational Divisions** K-5: 0.1020+ 6-12: 0.1965*
Productivity Skills	–	0.2339*
Selectivity Skills	0.0803	0.1841*

Skill Type	**Retention of Problem Solving Skills**		**Transfer of Problem Solving Skills**	
	Maintenance	Extension	Maintenance	Extension
Composite Problem Solving Skills	–	-0.0583+	–	–

* : mean effect significantly different from zero, + : descriptive statistic

Figure 47. Summary: Achievement Effects Regarding Acquisition, Retention, and Transfer of Problem Solving Skills

Mean Effect Sizes Regarding Basic Operational Skills

Skill Type	Acquisition of Operational Skills	
	Maintenance	Extension
Composite Operational Skills	0.1825*	**Educational Divisions** K-8: 0.3520+ 9-12: 0.3376*
Computation Skills	-0.2670*	0.0915
Conceptual Skills	0.0277	0.4806*

* : mean effect significantly different from zero, + : descriptive statistic

Figure 48. Summary: Achievement Effects Regarding Acquisition of Basic Operational Skills – Graphing Calculators

and mixed ability students in grades K-5 generated non-significant mean effect sizes. Thus, the basic skills of the calculator groups were statistically similar to the basic skills of their control group counterparts. The mixed ability students in grades 6-8 produced a positive mean effect size, 0.3852, but the value was descriptive and was not proved significant through inferential statistics.

With respect to computational skills, only one significant effect was generated. The mean effect size for the low ability group was 0.2978. Therefore, low ability students realized improvement in their paper-and-pencil skills as a result of participating in calculator treatment. With N = 10, this mean effect size was fairly resistant to the addition of studies with null results. The mixed ability group generated a positive but non-significant effect. Hence, it can be assumed there were no differences in the paper-and-pencil skills of calculator and non-calculator students in mixed ability classrooms. A descriptive statistic was calculated for students of high ability. The value was negative which leads to the possibility calculator use inhibited the growth of paper-and-pencil skills in high ability students.

For conceptual skills, a non-significant mean effect of 0.0562 was produced for all ability levels and educational divisions. Therefore, the calculator neither helped nor hindered students' development of skills necessary to understand mathematical concepts.

Effects regarding the acquisition of composite problem solving skills were quite different from the results for operational skills described above. A significant mean effect of 0.1160 was produced for all ability levels and educational divisions. Therefore, the paper-and-pencil problem solving skills of students using calculators significantly improved from treatment. Since the inclusion of eleven studies with null results would be required to draw the mean value out of significance, this data was fairly resistant to the file drawer problem. There are no problem solving productivity results to discuss since the studies did not provide sufficient data for statistical analysis. Problem solving selectivity skills were represented by a non-significant mean effect of 0.0803. Therefore, no significant differences existed between treatment and control groups with regard to the selection of appropriate processes for problem solving.

These results are similar to those reported by Hembree (1984) with two notable exceptions. Hembree was unable to provide significant effects with regard to students in low ability groups. The current meta-analysis reported significant positive effects for students of low ability in composite operational and computational skills. Therefore, a trend may be developing in which low ability students benefit from calculator use in the acquisition of basic skills. Hembree (1984) reported mixed results for students with respect to the acquisition of problem solving skills. In particular, students in mixed ability classes realized improvement in paper-and-pencil problem solving skills, while students of high or low ability were neither helped nor hindered. The current study revealed improved paper-and-pencil problem solving skills for students of all ability levels.

2. Maintenance Effects – Graphing Calculator Only

When the graphing calculator was the only treatment device under analysis, two significant effects were produced. The mean effect size for composite operational skills was 0.1825. This value revealed significant improvement in the basic skills of students using the calculator as compared to their non-calculator counterparts. However, the small fail-safe N value meant the significant mean effect could become non-significant with the addition of null results from three studies. The mean effect size for computational skills was -0.2670. Therefore, there is evidence that the graphing calculator had a negative influence on students' computational skills. With a fail-safe N value of two, this effect could easily be drawn out of significance with null results from two studies.

A non-significant effect was generated for conceptual skills acquisition. Therefore, the graphing calculator had no effect on the skills necessary for conceptual understanding of mathematics.

3. <u>Extension Effects – All Calculator Types</u>

While it was necessary to group the data according to educational divisions to generate statistically significant results, the basic skills of students in all divisions were improved when the calculator was an integral part of testing. This was true for computational and conceptual skills as well, although the method of partitioning the data was slightly different. The values were grouped according to student ability level. Students of all ability levels engaged in calculator treatment produced higher test scores than their non-calculator counterparts. These results confirm that calculator use during testing will improve student test scores, especially in basic mathematical skills. The fail-safe N values for the questions regarding students' basic skills ranged from six to 139. With N = 6, the results related to the conceptual skills of high ability students were most vulnerable to the file drawer problem. It would require 139 studies with null results to alter the significance of the composite operational results of high school students.

Problem solving results followed a similar positive trend with only one non-significant effect reported. The effect sizes from students of low ability were unable to produce a statistically significant result. Therefore, low ability students using calculators performed in the same fashion as low ability students without access to calculators. The problem solving results for classes of mixed and high ability students were better. Students in grades 6-12 generated a statistically significant positive mean effect. Students in grades K-5 also generated a positive mean effect, but it was descriptive instead of inferential. Therefore, mixed and high ability classes across all grade levels benefited, at least moderately, from calculator use with respect to the acquisition of problem solving skills. With calculator use, the students realized marked improvement in productivity and selectivity skills. When compared with students who did not use calculators, students in treatment groups were able to solve more problems and make better decisions with regard to selecting methods for generating solutions. These results lend further credence to Hembree's (1984) study. The problem solving results generated lower fail-safe N values than the basic operational skills results mentioned above. In the sense of extension, the fail-safe N values for problem solving skills ranged from four to ten. The composite problem solving results were most resistant to the file drawer problem with N = 10. Problem solving productivity and selectivity results could be significantly influenced by four studies reporting null results.

4. Extension Effects – Graphing Calculator Only

All of the effects were positive but only half of them were statistically significant when the graphing calculator was allowed during testing. The basic skills of high school students and the conceptual skills of all students experienced improvement from calculator use. These results were proven with statistically significant mean effect sizes. The results for each of these areas were generated with data from six studies. Therefore, the fail-safe N values of nine and 21 revealed the results to be resistant to the file drawer problem. With respect to the basic operational skills of high school students, 1.5 times the number of studies gathered would be needed to change the significance of the results. The fail-safe N value for conceptual skills reflects the need for 3.5 times as many studies as were used to generate the original data. The mean effect size of basic skills for elementary and middle school students was positive, but not inferential. While the value was not statistically significant, paper-and-pencil skills of all students were represented by a positive mean effect size. Therefore, when the graphing calculator was used during testing, students' basic skills and conceptual skills realized at least moderate improvement. Students' computational skills were neither helped nor hindered by graphing calculator use. Only two effect sizes were available for which the graphing calculator was involved in the acquisition of problem solving skills. Therefore, no results could be reported.

Retention of Skills

RQ3 and RQ4 considered the retention of operational and problem solving skills. In terms of maintenance, a non-significant mean effect was generated for operational skills. Hence, retention posttest scores for students involved in calculator treatment were no different from scores for students not involved in calculator treatment. Only descriptive mean effects were available for the retention of operational and problem solving skills in terms of the extension aspect of these questions. The mean effect for operational skills was a large, positive value but the mean effect for problem solving skills was a small, negative value. Therefore, students' abilities to retain basic skills may have improved from using calculators during testing. At the same time, students' abilities to retain problem solving skills were most likely not improved and may have been slightly harmed from calculator use during testing.

These results are different than those reported by Hembree (1984). In the sense of extension, Hembree listed positive results for calculator use with regards to operational and problem solving skills. While the current results were slightly positive, they were not statistically significant.

Transfer of Skills

RQ5 and RQ6 were designed to analyze calculator effects on the transfer of operational and problem solving skills to other mathematical areas. The studies integrated by meta-analysis did not produce sufficient data for a meaningful evaluation. Therefore, there are no transfer results to report.

Estimation Skills

RQ7 was established to assess calculator effects on students' estimation skills. None of the studies integrated by meta-analysis reported data on students' estimation abilities. Therefore, statistical analysis was not possible.

Student Attitudes

RQ8 through RQ14 analyzed calculator effects on different aspects of the attitude construct.

1. <u>All Calculator Types</u>

The results regarding students' attitudes were represented by four non-significant effects and one descriptive effect. The mean effect sizes representing students' attitudes toward mathematics were not statistically significant. Therefore, the calculator had no significant effect on students' attitude toward mathematics. The categories of self-concept in mathematics and value of mathematics in society also produced non-significant mean effect sizes. Therefore, students' mathematical self-concepts were neither helped nor hindered by calculator use. Similarly, students' perceptions about the value of mathematics in society were not influenced by calculator use. The descriptive effect related to students' attitudes toward the use of calculators in mathematics. The mean value of -0.0784 revealed the possibility of a slightly negative trend in this area. While the result was not supported by inferential statistics, students' attitudes may have been slightly negative about the role of the calculator in the mathematics classroom.

RQ10, RQ12, and RQ13 were not evaluated with the data produced by the current meta-analysis. Therefore, students' anxiety toward mathematics, motivation to learn mathematics, and attitudes toward mathematics teachers as a result of calculator use could not be assessed. The results reported are quite different than those presented by Hembree (1984). Hembree reported a statistically significant effects for students' attitude toward mathematics and students' self-concept in mathematics. The current study was unable to replicate the findings. The current study reported a non-significant mean effect regarding students' perceptions of the value of mathematics in society. Hembree's (1984) study did not contain sufficient data to analyze this attitude sub-construct.

2. Graphing Calculator Only

Only one attitude research question yielded significant results with regards to the graphing calculator. Students' attitudes toward mathematics were greatly improved after using graphing calculators during mathematics instruction. This was significantly different than the result generated for all types of calculators. A fail-safe N value of 8 was calculated for this data. Since six studies were involved in the data analysis, it would take the null results of twice as many studies to influence the statistical significance of this result.

Descriptively Analyzed Research Questions

RQ15 and RQ16 were addressed with non-inferential methods. There were no remarkable differences between calculator effects for male and female students. While only descriptive, the mean effects resulting from special curriculum materials were all positive and three out of four were fairly large. Composite problem solving skills in the sense of extension achieved the most significant improvement from calculator use. In particular, students realized higher scores on tests of paper-and-pencil skills after instruction with the calculator and special curriculum. Finally, students' attitudes toward mathematics benefited from the dual use of calculators and special curriculum materials.

Keeping in mind these values were descriptive while the results of most of the data from traditional instruction were inferential, the mean effects resulting from special calculator instruction were larger than their traditional instruction counterparts. Therefore, the combination of the calculator and specially created curriculum materials had a better effect on student achievement and attitude than the combination of the calculator and traditional instruction.

Chapter VI

Conclusions and Recommendations

The purpose of this study was to determine the effects of the calculator on precollege mathematics students. Meta-analysis was the method used to investigate calculator effects. In particular, a statistical integration of the 307 effect sizes, partitioned according to the research questions, was conducted. The conclusions are based on the findings of this analysis. Since 53 studies were gathered as a result of exhaustive search techniques, the collection was considered a representative, probabilistic sample adequate for meta-analysis. Generalizations of the findings reported in chapter five are presented below.

Conclusions

1. Students in grades K-12 maintain their paper-and-pencil mathematics skills after participation in traditional instruction with calculators. This is true for students of all ability levels and applies to all types of calculators.

2. The basic operational skills, with paper-and-pencil, of low ability students in all grades can improve as a result of calculator use during traditional instruction. This is also true for high school classes of mixed ability students.

3. The combination of calculators and traditional instruction can foster development of the computational skills of low ability students.

4. When calculators are used during traditional mathematics instruction and calculations are conducted with paper-and-pencil, the problem solving skills of students in all grades and all ability levels can improve.

5. With regard to basic operational skills, the scores for tests in which calculators are allowed will be higher than paper-and-pencil test scores. This is true for all grades and all ability levels.

6. Middle and high school classes containing mixed ability students can experience improvement in problem solving skills when calculators are an integral part of instruction and testing. This is also true for middle and high school classes containing high ability students.

7. Calculator use in instruction and testing fosters the development of computational skills of students of all ability levels and in all grades.

8. When calculators are a significant element of learning and evaluation, the skills necessary for understanding mathematical concepts are improved through the pairing of calculators and traditional instruction. This is true for all grades and all ability levels and applies to all types of calculators.

9. Students' abilities to select appropriate processes for use during problem solving improve when calculators are part of all aspects of the learning process. This is true for all grades and all ability levels.

10. Problem solving computations are more accurate as a result of the calculator being an integral part of mathematics instruction and evaluation. This is true for all grades and all ability levels.

11. When the graphing calculator is a significant element in all aspects of high school mathematics classes, the basic operational skills of students can improve.

12. Students who use graphing calculators during mathematics instruction will have better attitudes toward the subject than their non-calculator counterparts. This is true for all grades and all ability levels.

13. Curriculum designed specifically for instruction with calculators can enhance student achievement in operational and problem solving skills. This is especially true when the calculator is a significant element in all aspects of the learning process, including evaluation. However, further statistical analysis of these types of studies is necessary before more significant conclusions can be reported.

Recommendations for Classroom Usage

Recommendations for calculator use in mathematics classrooms include:

1. Calculators should be used in all precollege mathematics classrooms. Based on the grade distribution of the studies in this meta-analysis, length of calculator availability during instruction should increase with each increasing grade level.

2. Based on the limited research featuring the early grades, calculator use should be restricted to experimentation and recreation. In kindergarten

through second grade, basic computational skills are the featured construct in mathematics instruction and calculator use is unnecessary.

3. Calculators should be used during instruction of problem solving skills in middle and high school (i.e. grades six through twelve) mathematics courses. This may result in increased success with word problems as well as more positive attitudes toward mathematics, in general, and word problems in particular.

4. Calculators should be available during evaluations of middle and high school students' problem solving skills and their understanding of mathematical concepts. This recommendation is based on the following:

a. The overwhelming results reported in the current meta-analysis.

b. The fact-based opinions of other reviewers and educational experts regarding the inconsistencies that occur when tests are given without calculators after instruction has taken place with calculators. These ideas were discussed in chapter three of this study.

5. Teachers should design lessons which integrate calculator-based explorations of word problems and mathematical concepts with regular instruction, especially in middle and high school mathematics classrooms.

6. The NCTM (2000; 1989) has outlined suggestions for including technology in the mathematics curriculum. These suggestions should be incorporated in mathematics classrooms at all grade levels.

Recommendations for Future Research

Recommendations for future research include:

1. While nearly half of the studies featured the graphing calculator, the number of effect sizes available from those studies was relatively small. Therefore, the graphing calculator aspect of this study should be replicated when a larger sample of research studies is available.

2. The computer is another technological device advocated by NCTM. A study similar to the current meta-analysis should be conducted with computer-based research.

3. Only a select few studies researched the calculator's role in the retention and transfer of operational skills. Since we are becoming a more technological society, research should be conducted on retaining skills after instruction with calculators. Also, further research is needed regarding the transfer of skills to other mathematical subjects and to areas outside of mathematics.

4. The studies featuring the graphing calculator primarily focused on the acquisition of basic operational skills. In particular, only one effect size represented the relationship between the graphing calculator and student achievement in problem solving skills. Therefore, future research should include studies of graphing calculator use in the development of problem solving skills.

5. Most of the graphing calculator studies featured grades nine through twelve. It needs to be determined whether or not the graphing calculator

has a place in elementary and middle school classrooms. In particular, the graphing calculator's role in the elementary and middle school grades should be investigated.

6. As technology becomes more prevalent in the classroom and the latest NCTM recommendations involving technology are implemented, further study is needed in the methods used to prepare teachers to use calculators and other technological devices effectively.

Summary

This meta-analysis was an extension of similar work conducted by Hembree in 1984 and updated by Hembree and Dessart in 1992. The current study agreed with many of the results reported in the original meta-analysis (Hembree, 1984; Hembree & Dessart, 1992). Therefore, it appears that the use of calculators during the last fifteen years has not hindered student learning in the mathematics classroom. In fact, the results reflect that the calculator has been a positive learning tool for students of various grade and ability levels. The current study also reveals several areas of improvement since Hembree and Dessart's (1992) report. In particular, the operational skills of low ability students and the problem solving skills of all students have improved from calculator use.

As more calculator-based studies are conducted and the technological recommendations of the NCTM (2000) are implemented in K-12 classrooms, meta-analysis is an appropriate medium for determining if this positive trend continues. Based on the work of Hembree (1984) and Dessart (Hembree & Dessart, 1992) and the results of the current meta-analysis, the calculator has an important role to play in the mathematics classroom. Future research should include defining the calculator's role and determining the grade and ability levels in which the calculator can be most beneficial to precollege mathematics students.

References

References

Benz, H. E. (1933). A summary of some scientific investigations of the teaching of high school mathematics. In W. D. Reeve (Ed.), The teaching of mathematics in the secondary school (pp. 14-54). New York, NY: National Council of Teachers of Mathematics.

Burns, P. C., & Dessart, D. J. (1965). A summary of investigations relating to mathematics in elementary education: 1964. School Science and Mathematics, 65 (9), 779-790.

Burns, P. C., & Dessart, D. J. (1966a). A summary of investigations relating to mathematics in elementary education: 1965. School Science and Mathematics, 66 (9), 838-851.

Burns, P. C., & Dessart, D. J. (1966b). A summary of investigations relating to mathematics in secondary education: 1964. School Science and Mathematics, 66 (1), 73-80.

Carson, K. P., Schriesheim, C. A., & Kinicki, A. J. (1990). The usefulness of the "fail-safe" statistic in meta-analysis. Educational and Psychological Measurement, 50 (2), 233-243.

Cooper, H. M., & Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. Psychological Bulletin, 87, 442-449.

Dessart, D. J. (1964). Mathematics in the secondary school. Review of Educational Research, 34, 298-312.

Dessart, D. J. (1989). A review and synthesis of research in mathematics education reported during 1987. Columbus, OH: Clearinghouse for Science, Mathematics, and Environmental Education.

Dessart, D. J., & Burns, P. C. (1967). A summary of investigations relating to mathematics in secondary education: 1965. School Science and Mathematics, 67 (2), 135-144.

Dessart, D. J., & Frandsen, H. (1973). Research on teaching secondary school mathematics. In R. M. W. Travers (Ed.), Second handbook of research on teaching (pp. 1177-1195). Chicago, IL: American Educational Research Association.

Dessart, D. J., & Suydam, M. N. (1976). Classroom ideas from research on computational skills. Reston, VA: National Council of Teachers of Mathematics.

Dessart, D. J., & Suydam, M. N. (1983). Classroom ideas from research on secondary school mathematics. Reston, VA: National Council of Teachers of Mathematics.

Friedman, L. (1989). Mathematics and the gender gap: A meta-analysis of recent studies on sex differences in mathematical tasks. Review of Educational Research, 59 (2), 185-213.

Gilchrist, M. E. (1993). Calculator use in mathematics instruction and standardized testing: An adult education inquiry. Dayton, VA: Virginia Adult Educators Research Network. (ERIC Document Reproduction Service No. ED 372 919)

Glass, G. V. (1978). Integrating findings: The meta-analysis of research. Review of Research in Education, 5, 351-379.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). Meta-analysis in social research. Beverly Hills, CA: Sage.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. Journal of Educational Statistics, 6 (2), 107-128.

Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. San Diego, CA: Academic Press.

Hedges, L. V., Shymansky, J. A., & Woodworth, G. (1989). A practical guide to modern methods of meta-analysis. Washington, DC: National Science Teachers Association. (ERIC Document Reproduction Service No. ED 309 952)

Hembree, R. (1984). Model for meta-analysis of research in education, with a demonstration in mathematics education: Effects of hand-held calculators (Doctoral dissertation, University of Tennessee, 1984). Dissertation Abstracts International, 45, 3087A.

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. Review of Educational Research, 58 (1), 47-77.

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. Journal for Research in Mathematics Education, 21 (1), 33-46.

Hembree, R. (1992). Experiments and relational studies in problem solving: A meta-analysis. Journal for Research in Mathematics Education, 23 (4), 242-273.

Hembree, R. & Dessart, D. J. (1986). Effects of hand-held calculators in precollege mathematics education: A meta-analysis. Journal for Research in Mathematics Education, 17 (2), 83-99.

Hembree, R., & Dessart, D. J. (1992). Research on calculators in mathematics education. In J. T. Fey (Ed.), Calculators in mathematics education (pp. 23-32). Reston, VA: National Council of Teachers of Mathematics.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). Meta-analysis: Cumulating findings across studies. Beverly Hills, CA: Sage.

Light, R. J., & Pillemer, D. B. (1984). Summing up. Cambridge, MA: Harvard University Press.

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. Journal for Research in Mathematics Education, 30 (5), 520-540.

Maccoby, E. E., & Jacklin, C. N. (1974). The psychology of sex differences. Stanford, CA: University Press.

National Council of Teachers of Mathematics. (1974). NCTM Board approves policy statement on the use of minicalculators in the mathematics classroom. NCTM Newsletter, 11, 3.

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

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

Neubauer, S. G. (1982). The use of hand-held calculators in schools: A review. South Bend, IN: Indiana University. (ERIC Document Reproduction Service No. ED 220 272)

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. Journal of Educational Statistics, 8, 157-159.

Parkhurst, S. (1979). Hand-held calculators in the classroom: A review of the research. (ERIC Document Reproduction Service No. ED 200 416)

Penglase, M., & Arnold, S. (1996). The graphics calculator in mathematics education: A critical review of recent research. Mathematics Education Research Journal, 8 (1), 58-90.

Rabe, R. M. (1981). Calculators in the mathematics curriculum: Effects and changes. South Bend, IN: Indiana University. (ERIC Document Reproduction Service No. ED 204 178)

Roberts, D. M. (1980). The impact of electronic calculators on educational performance. Review of Educational Research, 50 (1), 71-88.

Rosenthal, R. (1978). Combining results of independent studies. Psychological Bulletin, 85 (1), 185-193.

Rosenthal, R. (1991). Meta-analytic procedures for social research (Rev. ed.). Newbury Park, CA: Sage.

Sigg, P. O. (1982). The hand-held calculator: Effects on mathematical abilities and implications for curriculum change. South Bend, IN: Indiana University School of Education. (ERIC Document Reproduction Service No. ED 218 147)

Smith, B. A. (1997). A meta-analysis of outcomes from the use of calculators in mathematics education. (Doctoral dissertation, Texas A & M University, 1996). Dissertation Abstracts International, 58, 0787A.

Suydam, M. N. (1976). Electronic hand calculators: The implications for pre-college education. Final Report. Abbreviated Version. Washington, DC: National Science Foundation. (ERIC Document Reproduction Service No. ED 127 206)

Suydam, M. N. (1978). State-of-the-art review on calculators: Their use in education. Columbus, OH: Calculator Information Center. (ERIC Document Reproduction Service No. ED 167 426)

Suydam, M. N. (1979a). The use of calculators in pre-college education: A state-of-the-art review. Columbus, OH: Calculator Information Center. (ERIC Document Reproduction Service No. ED 171 573)

Suydam, M. N. (Ed.) (1979b). Information bulletins from the calculator information center, bulletins 1-7. Columbus, OH: Calculator Information Center. (ERIC Document Reproduction Service No. ED 171 574)

Suydam, M. N. (1980). Using calculators in pre-college education: Third annual state-of-the-art review. Columbus, OH: Calculator Information Center. (ERIC Document Reproduction Service No. ED 206 454)

Suydam, M. N. (1981a). The use of calculators in pre-college education: Fourth annual state-of-the-art review. Columbus, OH: Calculator Information Center. (ERIC Document Reproduction Service No. ED 206 454)

Suydam, M. N. (Ed.) (1981b). Information bulletins from the calculator information center, bulletins 8-11. Columbus, OH: Calculator Information Center. (ERIC Document Reproduction Service No. ED 206 453)

Suydam, M. N. (1982). The use of calculators in pre-college education: Fifth annual state-of-the-art review. Columbus, OH: Calculator Information Center. (ERIC Document Reproduction Service No. ED 220 273)

Suydam, M. N., & Brosnan, P. A. (1994). Research on mathematics education reported in 1993. Journal for Research in Mathematics Education, 25 (4), 375-434.

Suydam, M. N., & Weaver, J. F. (1972). Research on mathematics education (K-12) reported in 1971. Journal for Research in Mathematics Education, 3 (4), 196-232.

Suydam, M. N., & Weaver, J. F. (1978). Research on mathematics education reported in 1977. Journal for Research in Mathematics Education, 9 (4), 242-318.

Tukey, J. W. (1977). Exploratory data analysis. Reading, MA: Addison-Wesley.

Yang, W. (1997). Validity issues in cross-national relational analyses. A meta-analytic approach to perceived gender differences on mathematics learning. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 409 331)

Appendices

Appendix A

Criteria for a Research Design Rating

Criteria for a Research Design Rating

A research design rating was tabulated for each study included in the meta-analysis. The method outlined here closely resembles the method outlined by Hembree (1984). Each study was assessed according to eight criteria: problem definition, population description, sampling procedures, error control, test instruments, data analysis, conclusions, and evaluation of the overall report. For the appropriate level attained under each criterion, the heirarchical point value listed in parentheses was applied to the study's rating. The numerical rating was calculated by adding the total number of points obtained from the eight categories and dividing by ten.

1. Statement of the problem
   a. Clear hypothesis (3)
   b. No hypothesis but clear research questions (2)
   c. Hypothesis or research questions are confusing (1).
2. Description of the population under study
   a. Thorough description (3)
   b. Partial description (2)
   c. Minimal description (1)
3. Sampling procedures used
   a. Fully random sample (6)
   b. Existing population with randomized students, teachers, and classes (5)

c. Existing population with randomized students, but no other elements of a random sample (4)

d. Existing classes with randomized groups and teachers, but no other elements of a random sample (3)

e. Existing classes with randomized groups, but no elements of a random sample (2)

f. Not random because the sample contains existing classes (1)

4. Methods used to control for error

a. Pretest-Posttest Control Group design (PPCG), with analysis of covariance (ANCOVA), control for pretest-treatment interaction, and low experimenter bias (6)

b. PPCG with ANCOVA, with control for pre-test treatment interaction, and high experimenter bias (5)

c. PPCG with ANCOVA, without control for interaction pre-test treatment interaction (4)

d. PPCG without ANCOVA, with control pre-test treatment interaction, and low experimenter bias (3)

e. PPCG without ANCOVA, with control for pre-test treatment interaction, and high experimenter bias (2)

f. PPCG without ANCOVA, without control for pre-test treatment interaction(1)

5. Test instruments used

a. Standardized (3)

b. Teacher designed with reliability information provided (2)

c. Teacher designed without reliability information (1)

6. Methods of data analysis

a. Appropriate methods with full disclosure of the results (3)

b. Appropriate methods but missing data (2)

c. Inappropriate methods or a large amount of missing data (1)

7. Researcher's conclusions

a. Appropriate and related to the hypotheses (3)

b. Appropriate but unclear or muddled (2)

c. Inappropriate (1)

8. Does the report allow a reader to critically examine the evidence?

a. Yes (3)

b. More or less (2)

c. No (1)

Appendix B

Bibliography of Studies Included in the Meta-Analysis

Bibliography of Studies Included in the Meta-Analysis

1. Abuloum, K. (1996). Graphing calculators: Teachers perceptions, training, and attitude (Doctoral dissertation, University of Nebraska, 1996). Dissertation Abstracts International, 57, 1063A.

2. Aldridge, W. B. (1991). The effects of utilizing calculators and a mathematics curriculum stressing problem solving techniques (Doctoral dissertation, Wayne State University, 1990). Dissertation Abstracts International, 54, 0404A. (University Microfilms No. AAC91-18846)

3. Ansley, T., Spratt, K., & Forsyth, R. (1989). The effects of using calculators to reduce the computational burden on a standardized test of mathematics problem solving. Educational and Psychological Measurement, 49 (1), 277-286.

4. Bartos, J. J. (1986). Mathematics achievement and the use of calculators for middle elementary grade children (Doctoral dissertation, United States International University, 1986). Dissertation Abstracts International, 47, 1227A. (University Microfilms No. AAC86-14912)

5. Bitter, G., & Hatfield, M. (1992). The calculator project: Assessing school-wide impact of calculator integration. In G. Bright, H. Waxman, & S. Williams (Eds.), Impact of calculators on mathematics instruction (pp. 49-65). Lanham, MD: University Press of America.

6. Boudreau, J. (1999). A treatment on middle school students to increase student achievement on geometry skills (Master's thesis, California State University, 1999). Masters Abstracts International, 37, 1297A.

7. Bridgeman, B., Harvey, A., & Braswell, J. (1995). Effects of calculator use on scores on a test of mathematical reasoning. Journal of Educational Measurement, 32 (4), 323-340.

8. Burnett, C. M. (1985). The effectiveness of using a hand-held calculator as an instructional aid in the teaching of fraction to decimal conversion to sixth-grade students (Doctoral dissertation, Boston University, 1985). Dissertation Abstracts International, 46, 2174A.

9. Chandler, P. A. (1992). The effect of the graphing calculator on high school students' mathematical achievement (Doctoral dissertation, University of Houston, 1992). Dissertation Abstracts International, 53, 3832A.

10. Colefield, R. P. (1985). The effect of the use of electronic calculators versus hand computation on achievement in computational skills and achievement in problem-solving abilities of remedial middle school students in selected business mathematics topics (Doctoral dissertation, New York University, 1985). Dissertation Abstracts International, 46, 2168A.

11. Devantier, A. (1992). The impact of graphing calculators on the understanding of functions and their graphs (Master's thesis, Central Michigan University, 1992). Masters Abstracts International, 31, 535A.

12. Drottar, J. (1999). An analysis of the effect of the graphing calculator on student performance in algebra II (Doctoral dissertation, Boston College, 1999). Dissertation Abstracts International, 60, 56A.

13. Edens, H. S. (1983). Effects of the use of calculators on mathematics achievement of first grade students (Doctoral dissertation, University of Virginia, 1983). Dissertation Abstracts International, 43, 3248A.

14. Ellerman, T. (1998). A study of calculator usage on the mathematics achievement of seventh and eighth grade students and on attitudes of students and teachers (Doctoral dissertation, Louisiana Tech University, 1998). Dissertation Abstracts International, 59, 1101A.

15. Fleener, M. (1988). Using the computer as an instructional tool to aid in mathematical problem solving (Doctoral dissertation, University of North Carolina, 1988). Dissertation Abstracts International, 50, 0611A.

16. Frick, F. A. (1988). A study of the effect of the utilization of calculators and a mathematics curriculum stressing problem-solving techniques on student learning (Doctoral dissertation, University of Connecticut, 1988). Dissertation Abstracts International, 49, 2104A.

17. Giamati, C. M. (1991). The effect of graphing calculator use on students' understanding of variations on a family of equations and the transformations of their graphs (Doctoral dissertation, University of Michigan, 1991). Dissertation Abstracts International, 52, 0103A. (University Microfilms No. AAC91-16100)

18. Glover, M. (1992). The effect of the hand-held calculator on the computation and problem solving achievement of students with learning disabilities (Doctoral dissertation, State University of New York at Buffalo, 1991). Dissertation Abstracts International, 52, 3888A.

19. Graham, A., & Thomas, M. (in press). Building a versatile understanding of algebraic variables with a graphic calculator. Educational Studies in Mathematics.

20. Graham, A., & Thomas, M. (2000). A graphic calculator approach to understanding algebraic variables. Manuscript in preparation, TIME 2000 International Conference, Auckland, New Zealand.

21. Hall, M. K. (1992). Impact of the graphing calculator on the instruction of trigonometric functions in precalculus classes (Doctoral dissertation, Baylor University, 1992). Dissertation Abstracts International, 54, 0451A.

22. Harskamp, E., Suhre, C., & Van Streun, A. (1998). The graphics calculator in mathematics education: An experiment in the Netherlands. Hiroshima Journal of Mathematics Education, 6, 13-31.

23. Heath, R. D. (1987). The effects of calculators and computers on problem solving ability, computational ability, and attitude towards mathematics (Doctoral dissertation, Northern Arizona University, 1987). Dissertation Abstracts International, 48, 1102A.

24. Hersberger, J. (1983). The effects of a problem-solving oriented mathematics program on gifted fifth-grade students (Doctoral dissertation, Purdue Univeristy, 1983). Dissertation Abstracts International, 44, 1715A. (University Microfilms No. AAC83-23998)

25. Humphries, J. H. (1986). The effectiveness of the Little Professor calculator as a supplement to kindergarten mathematics program (Doctoral dissertation, East Texas State University, 1986). Dissertation Abstracts International, 47, 2014A.

26. Kelly, M. G. (1984). The effect of the use of the hand-held calculator on the development of problem-solving strategies (Doctoral dissertation, Utah State University, 1984). Dissertation Abstracts International, 46, 2169A.

27. King, M. (1998). The effects of frequency of technology use on high school students' mathematics and science achievement (Doctoral dissertation, University of Houston, 1998). Dissertation Abstracts International, 59, 0778A.

28. Langbort, C. R. (1983). An investigation of the ability of fourth grade children to solve word problems using hand-held calculators (Doctoral dissertation, University of California, 1983). Dissertation Abstracts International, 43, 2914A.

29. Lawrence, I., & Dorans, N. (1994). Optional use of calculators on a mathematical test: Effect on item difficulty and score equating. (Report No. ETS-RR-94-40). Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 382 656)

30. Lesmeister, L. (1997). The effect of graphing calculators on secondary mathematics achievement (Master's thesis, University of Houston Clear Lake, 1996). Masters Abstracts International, 35, 0039A.

31. Lewis, J., & Hoover, H. (1983). The effect on pupil performance of using hand-held calculators on standardized mathematics achievement tests. Paper presented at the annual meeting of the National Council on Measurement in Education, Los Angeles, CA. (ERIC Document Reproduction Service No. ED 204 152)

32. Lim, C. (1992). Effects of using calculators on computational ability on non-college bound students, their attitudes toward mathematics and achievements on unit tests in probability and statistics (Doctoral dissertation, Temple University, 1991). Dissertation Abstracts International, 52, 2450A.

33. Liu, S. T. (1993). Effects of teaching calculator use and problem solving strategies on mathematics performance and attitude of fifth grade Taiwanese male and female students (Doctoral dissertation, Memphis State University, 1993). Dissertation Abstracts International, 54, 4354A.

34. Long, V. M., Reys, B., & Osterlind, S. J. (1989). Using calculators on achievement tests. Mathematics Teacher, 82 (5), 318-325.

35. Loyd, B. (1991). Mathematics test performance: The effects of item type and calculator use. Applied Measurement in Education, 4 (1), 11-22.

36. Magee, E. F. (1985). The use of the minicalculator as an instructional tool in applying consumer mathematics concepts (Doctoral dissertation, University of Virginia, 1985). Dissertation Abstracts International, 47, 2104A.

37. Merckling, W. (1999). Relationship(s) between the perceptual preferences of secondary school students and their achievement in functions using a graphing calculator (Doctoral dissertation, St. John's University, 1999). Dissertation Abstracts International, 60, 371A.

38. Morgan, R., & Stevens, J. (1991). Experimental study of the effects of calculator use on the advanced placement calculus examinations. (Report No. ETS-RR-91-5). Princeton, NJ: Educational Testing Service. (ERIC Document Reproduction Service No. ED 392 816)

39. Ottinger, T. P. (1994). Conceptual and procedural learning in first-year algebra using graphing calculators and computers (Doctoral dissertation, Georgia State University, 1993). Dissertation Abstracts International, 54, 2934A.

40. Rich, S. B. (1991). The effect of the use of graphing calculators on the learning of function concepts in precalculus mathematics (Doctoral dissertation, University of Iowa, 1990). Dissertation Abstracts International, 52, 0835A.

41. Riley, A. G. (1992). Effects of integrating TI Math Explorer calculators into the grades 3 through 6 mathematics curriculum on achievement and attitude (Doctoral dissertation, Texas A&M University, 1992). Dissertation Abstracts International, 54, 1661A.

42. Rodgers, K. (1996). The effects on achievement, retention of mathematical knowledge, and attitudes toward mathematics as a result of supplementing the traditional algebra II curriculum with graphing calculator activities (Doctoral dissertation, Southern Illinois University, 1995). Dissertation Abstracts International, 57, 0091A.

43. Ruthven, K. (1990). The influence of graphic calculator use on translation from graphic to symbolic forms. Educational Studies in Mathematics, 21 (5), 431-450.

44. Scott, B. A. (1994). The effect of graphing calculators in algebra II classrooms: A study comparing achievement, attitude, and confidence (Doctoral dissertation, University of North Texas, 1994). Dissertation Abstracts International, 55, 2755A.

45. Siskind, T. (1994). The effect of calculator use on mathematics achievement for rural high school students. Rural Educator, 16 (2), 1-4.

46. Stacey, K., & Groves, S. (1994). Calculators in primary mathematics. Paper presented at the 72nd meeting of the National Council of Teachers of Mathematics, Indianapolis, IN. (ERIC Document Reproduction Service No. ED 373 963)

47. Starr, S. (1989). The effects of including calculators on the problem-solving instruction for low-income sixth-grade students (Master's thesis, Mercer University, 1989). (ERIC Document Reproduction Service No. ED 309 070)

48. Szetela, W., & Super, D. (1987). Calculators and instruction in problem solving in grade 7. Journal for Research in Mathematics Education, 18 (3), 215-229.

49. Upshaw, J. T. (1994). The effect of the calculator-based, graph-exploration method of instruction on advanced placement calculus achievement (Doctoral dissertation, University of South Carolina, 1993). Dissertation Abstracts International, 54, 4023A.

50. Vazquez, J. L. (1991). The effect of the calculator on student achievement in graphing linear functions (Doctoral dissertation, University of Florida, 1991). Dissertation Abstracts International, 51, 3660A. (University Microfilms No. AAC91-06488)

51. Weber, T. (1999). Graphing technology and its effect on solving inequalities (Doctoral dissertation, Wayne State University, 1998). Dissertation Abstracts International, 60, 88A. (University Microfilms No. AAC99-15749)

52. Whisenant, M. A. (1989). The effects of the use of calculators in algebra I classes on basic skills maintenance and algebra achievement (Doctoral dissertation, University of North Texas, 1989). Dissertation Abstracts International, 54, 0450A.

53. Wilkins, C. (1995). The effect of the graphing calculator on student achievement in factoring quadratic equations (Doctoral dissertation, Mississippi State University, 1995). Dissertation Abstracts International, 56, 2159A.

Appendix C

Results of Data Collection – Dependent Variables

Table C-1: Effect Sizes by Study Number
Dependent Variables Related to Student Achievement in the Acquisition of Operational Skills
Equivalent Instruction Except for Calculator Use

	Composite				Computation				Concept			
Study #	Maintenance		Extension		Maintenance		Extension		Maintenance		Extension	
1	.8294											
2	.4839											
4					.2242		-.0382					
5	.3418	.8331	.2037	.9554	.3012	.8652	.2342	1.0413	.3365	.9066	.2315	.8518
7			.2911									
8					-1.3260	-.7443						
					.0989	.5399						
9	.3149											
10							.0101	.8780				
11			.1813	.4715								
12			.3221	.4415			.2862	.5957			-.0651	.2059
13	-.1588				-.0849				-.3609			
14											.0123	.2627
											.2859	.3301
15							.6239					
17	-.3672				-.4416				-.2154			
18					-1.0938	-.8811	.4028	.6094				
					.0969	-.2884	.6436	1.1210				
19											.3041	.5673
											.6463	.7084
											1.3664	

Table C-1 Continued

	Composite				Computation				Concept			
Study #	Maintenance		Extension		Maintenance		Extension		Maintenance		Extension	
20											.6294	.7209
21	-.1651											
22			1.0747									
23	-.5203	.3181			.3752	.3897			.0236	.1100		
	.4243	.4717			.5918				.1189			
24	1.2274				-.4437	1.6093						
25	.1257	.6284										
	.6589											
26									.8898			
27	.0551	.0891										
28					-.0012	.0104	.4143	.4969				
					.0220		.5796					
29			.1463	.2598								
31			.3478	.6087			.5581	1.3372			-.1198	.1374
32	-.0838	.4137			.4023							
33					-.7208	-.1632						
					.5742							
34			.2852	.6443			.6637	.1774				
35			.3980				.5350					
37	.2994	.3397										
38			.0599									
39	.2743				.1213				.4272			
40			.3107									

Table C-1 Continued

	Composite		Computation		Concept	
Study #	Maintenance	Extension	Maintenance	Extension	Maintenance	Extension
41	-.3220 -.1675		-.1974 -.1189		-.1956 -.1490	
	.0694 .2298		-.0198 .1470		.1916 .2235	
42		.7439	-.7443			
43		.8880		1.7305		.0454
44		.0248				
45		.6976				
46	-.0381 .2036		-.1720 .2424		.0198 .1906	
49	-.3080		-.2050		-.4214 -.4135	
					-.1920	
50		-.4067 .4044				
51	-.3070 .4549					
52		-.5223 -.3868	-.7424 .0125			
		.3359	.0622			
53		.8046		-.2973		2.0341

Table C-2: Effect Sizes by Study Number
Dependent Variables Related to Student Achievement in the Acquisition of Problem Solving Skills
Equivalent Instruction Except for Calculator Use

	Composite				Productivity				Selectivity			
Study #	Maintenance		Extension		Maintenance		Extension		Maintenance		Extension	
3			-.0773	.0329							.3904	.2046
			.1026	.3637							.1565	-.0911
5	.3877	.7275	.1453	.9732								
10			-.6017	.2290								
12											.6527	.0780
14							.3554	.2729				
							.1939					
15			.1430								-.2876	
18	-.8724	-.7880	-.9622	-.0130								
	-.3225	.0962	.1059	.3527								
23	.2701	.2886										
	.4898											
26	.3158								.2902			
28	-.2425	-.0481	.6482	.7293					-.4554	-.2261	.5877	-.0796
	.1940	.1999	.8363	.9138					-.1369	.1872	-.1546	-.2290
			.9433	1.1795								
31			.2846	.4241								
33			-.8548	-.5136								
			-.4814									
34			.3583	.6481								
35			-.0440								.2838	
36	.1582	.2433										

Table C-2 Continued

	Composite		Productivity		Selectivity	
Study #	Maintenance	Extension	Maintenance	Extension	Maintenance	Extension
41	-.3551 -.2286					
	.0142 .1374					
42		1.8421				
46					.0705 .6522	
47	.0123					
48						.4581 .2952
53		.6769				

Table C-3: Effect Sizes by Study Number
Dependent Variables Related to Student Achievement in the Retention and Transfer of Skills
Equivalent Instruction Except for Calculator Use

	Retention of Operational Skills		Retention of Problem Solving Skills		Transfer of Skills – Extension	
Study #	Maintenance	Extension	Maintenance	Extension	Operational	Problem Solving
8	-.2297 -.1446 -.0973 -.1376					
10		1.3153		-1.2295 .6803		
12		-.4035			.0165 .3712	
15		.5245		.3195 .2460		-.1905
22		.2959				
42		.5378				

Table C-4: Effect Sizes by Study Number
Dependent Variables Related to Student Attitudes
Equivalent Instruction Except for Calculator Use

Study #	Attitude Toward Mathematics		Attitude Toward Use of Calculator in Math		Anxiety Toward Mathematics	Self-Concept in Mathematics	
2	.9727						
4	-.0965						
14	-.0256		.0388			.0440	
23	.4119						
24	-.3657	-.1945				-.3582	-.1546
32	-.1834						
33	.2403	.5028					
	.5074						
37	.4882	.5698					
40	.4023						
41	-.0447	.0296	-.7924	-.4385			
	.2874	.3539	-.4030	-.3718			
42	-.4216				.3539	.3070	
44	.6497					1.7462	
47	-.7940						
50	.3216						
51	-.3114						
52	.0732			.7207	-.1742	.2214	

Table C-4 Continued

Study #	Motivation to Learn Mathematics	Attitude Toward Mathematics Teachers	Perception of the Value of Mathematics in Society
14	.0329		
24			-.1157 .3470
52	.1126	-.1859	-.0871

Table C-5: Effect Sizes by Study Number
Dependent Variables Related to Student Achievement and Attitude
Special Instruction with Calculators versus Traditional Instruction without Calculators

	Composite Operational		Computation Operational		Problem Solving	
Study #	Maintenance	Extension	Maintenance	Extension	Maintenance	Extension
6		1.0065				
16						1.0820 1.0916
						1.5208
30				-.0279		
33			.9864 -.0468			.0250 .1795
			.1594			.2561
48				.1474		-.3483

Study #	Attitude Toward Mathematics		
33	.4246	.0396	.3635
48	.3538		

Appendix D

Results of Data Collection – Independent Variables

Table D: Independent Variable Data by Study Number

Study #	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	D	S	G	9-12	Mix		Algebra I	L
2	D	S	B	8	Mix	Low	Problem Solving	L
3	J	S	B	10-12	Mix		Problem Solving	L
4	D	S	B	3	Mix	Middle	Number Concepts	L
5	J	S	B	7, 8	Mix			L
6	M	X	G	8	High		Geometry	M
7	J	S	G, S, B	11	Mix			L
8	D	X	B	6	Mix/High	Middle	Fraction and Decimal	L
9	D	R	G	11, 12	Mix		Transformations	L
10	D	S	B	7, 8	Low	Middle	Business Mathematics	L
11	M	X	G	11, 12	Mix		Functions	L
12	D	X	G	10-12	Mix		Algebra II	M
13	D	S	B	1	Mix		Addition and Subtraction	L
14	D	S	B	7, 8	Mix			L
15	D	R	B	7	Mix		Fractions	M
16	D	S	B	6	Mix			L
17	D	X	G	9-12	High		Functions	L
18	D	X	B	5-8	Low		Fractions	L
19	U	X	G	6, 7	Mix/High		Variable	L
20	U	X	G	7, 8	Mix		Variable	L

Legend for Independent Variables:

X_1: Publication Form
X_2: Evaluation Instrument Type
X_3: Type of Calculator
X_4: Grade Level
X_5: Student Ability Level
X_6: Socio-Economic Status
X_7: Subject Matter
X_8: Experimenter Bias

Table D Continued

Study #	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
21	D	X	G	11, 12	Mix		Trigonometry	L
22	J	R	G	12	Mix		Functions	L
23	D	S	B	8	Mix			M
24	D	S	B	5	Mix-High		Advanced Topics	M
25	D	S	B	K	Mix			L
26	D	R	B	6	Mix		Problem Solving	L
27	D	R	G, S, B	10-12	Mix			L
28	D	R	B	4	High		Word Problems	M
29	E	S	G, S, B	11	Mix			L
30	M	X	G	10-12	Mix-High	Middle	Functions	L
31	E	S	B	4, 8	Mix		Fractions and Decimals	L
32	D	S, X	B	9, 11	Low-Mix		Probability and Statistics	M
33	D	S, R	B	5	Mix	Middle		L
34	J	S	G, S, B	8, 10	Mix			L
35	J	S	B	8-11	Mix			L
36	D	X	B	10-12	Mix		Consumer Mathematics	L
37	D	X	G	10-12	Mix	Middle	Functions	L
38	E	S	G & S	11, 12	Mix			L
39	D	X	G	9-12	Mix		Algebra I	M
40	D	R	G	11, 12	High		Functions	L

Legend for Independent Variables:

X_1: Publication Form
X_2: Evaluation Instrument Type
X_3: Type of Calculator
X_4: Grade Level
X_5: Student Ability Level
X_6: Socio-Economic Status
X_7: Subject Matter
X_8: Experimenter Bias

Table D Continued

Study #	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
41	D	R	B	3-6	Mix			L
42	D	R	G	9-12	Mix		Functions	M
43	J	X	G	11, 12	Mix		Functions	L
44	D	R	G	10-12	Mix		Conic Sections	L
45	J	R	S	9-12	Mix		Percents	M
46	E	X	B	3, 4	Mix			L
47	M	X	B	6	Mix	Low	Fractions and Decimals	M
48	J	R	B	7	Mix		Problem Solving	L
49	D	R	G	9-12	High		Calculus I	L
50	D	R	G	8	Mix		Functions	L
51	D	R	G	9, 10	Mix	Middle	Inequalities	L
52	D	S	S	8-11	Mix	Middle	Algebra I	L
53	D	X	G	8	High	Middle	Factoring	M

Legend for Independent Variables:

X_1: Publication Form
X_2: Evaluation Instrument Type
X_3: Type of Calculator
X_4: Grade Level
X_5: Student Ability Level
X_6: Socio-Economic Status
X_7: Subject Matter
X_8: Experimenter Bias

Table D Continued

Study #	X_9	X_{10}	X_{11}	X_{13}	X_{14}	X_{15}	X_{19}	X_{20}
1	T	P	1-1		1996	185		1.9
2	T	P	1-1	F	1990	45		2.3
3	T	F	1-1		1989	0		2.2
4	T	F	1-1		1986	80		1.9
5	T	P	1-1		1992	185		2.2
6	S	P	1-1		1999	185		1.9
7	T	F	1-1		1995	0		2.2
8	T	P	1-1		1985	25	4	2.1
9	T	P	1-1		1992	10		2.3
10	T	F	1-1		1985	120	12	2.4
11	T	P	1-1		1992	90		2.0
12	T	P	1-1		1999	24		2.0
13	T	F	1-1		1983	30		2.2
14	T	F	1-1	M	1998	30		2.1
15	T	F	1-1		1988	4	2	2.3
16	S	P	1-1		1988	45		2.5
17	T	P	2-1		1990	22		1.9
18	T	F	1-1		1991	11		1.6
19	T	P	1-1		2000	15		1.6
20	T	P	1-1		2000	15		1.6

Legend for Independent Variables:

X_9: Curriculum
X_{10}: Method of Calculator Use
X_{11}: Calculator Availability
X_{13}: Gender-related Differences
X_{14}: Publication Year
X_{15}: Length of Treatment
X_{19}: Length of Retention Period
X_{20}: Research Design Rating

Table D Continued

Study #	X_9	X_{10}	X_{11}	X_{13}	X_{14}	X_{15}	X_{19}	X_{20}
21	T	P	1-1		1992	15		2.4
22	T	P	1-1		1998	185		1.8
23	T	F	2-1		1987	70		2.2
24	T	P	1-1		1983	185		2.1
25	T	F	2-1		1986	100		2.1
26	T	P	1-1		1984	12		2.1
27	T	F	1-1		1998	370		2.7
28	T	F	1-1		1983	5		2.1
29	T	F	1-1		1994	0		1.8
30	S	P	1-1		1996	15		1.7
31	T	F	1-1		1983	0		1.8
32	T	F	1-1		1991	85		2.2
33	T, S	P	1-1	M	1993	45		2.4
34	T	F	1-1		1989	0		2.3
35	T	F	1-1		1991	0		2.0
36	T	F	1-1		1985	45		1.9
37	T	P	1-1		1999	6		2.2
38	T	F	1-1		1992	0		1.8
39	T	P	1-1		1993	120		1.7
40	T	P	1-1		1990	185		1.9

Legend for Independent Variables:

X_9: Curriculum
X_{10}: Method of Calculator Use
X_{11}: Calculator Availability
X_{13}: Gender-related Differences
X_{14}: Publication Year
X_{15}: Length of Treatment
X_{19}: Length of Retention Period
X_{20}: Research Design Rating

Table D Continued

Study #	X_9	X_{10}	X_{11}	X_{13}	X_{14}	X_{15}	X_{19}	X_{20}
41	T	P	1-1		1992	120		2.3
42	T	P	1-1		1995	8		2.1
43	T	P	1-1		1990	185		1.6
44	T	P	1-1		1994	30		2.4
45	T	F	1-1		1994	5		2.2
46	T	P	1-1		1994	650		1.6
47	T	P	1-1		1989	40		2.1
48	S	P	2-1		1987	185		2.1
49	T	P	1-1		1993	10		2.3
50	T	P	1-1	M	1990	8		1.9
51	T	P	1-1		1998	6		2.1
52	T	P	1-1		1989	140		2.3
53	T	P	1-1	F	1995	25		2.2

Legend for Independent Variables:

X_9: Curriculum
X_{10}: Method of Calculator Use
X_{11}: Calculator Availability
X_{13}: Gender-related Differences
X_{14}: Publication Year
X_{15}: Length of Treatment
X_{19}: Length of Retention Period
X_{20}: Research Design Rating

VITA

Aimee J. Ellington was born in Fredericksburg, Virginia on June 27, 1968. She developed an interest in a career in education at an early age. The decision to pursue her dream to become a teacher was solidified through the encouragement and inspiration of her high school mathematics teacher. After graduating from Shenandoah Valley Academy in 1986, she enrolled at Columbia Union College. She received a Bachelor of Science degree with a major in mathematics and a minor in computer science in 1991. The last year of her undergraduate education was spent at IBM in Austin, Texas. Through a cooperative education opportunity, she was able to enhance her computer science skills. She wrote computer programs that generated statistical information from designs of computer chips. This memorable experience fostered her interest in incorporating technology in pedagogical activities and educational research endeavors.

In 1991, she was given the opportunity to pursue a graduate-level mathematics degree at the University of North Texas. Her first teaching engagements were the result of her participation in the mathematics department as a teaching fellow. She acquired a Master of Science degree in mathematics in 1993. In order to gain experience in the classroom, she spent two years as a full time mathematics instructor in Chattanooga, Tennessee. In 1995, while continuing to teach full time in Chattanooga, she enrolled in the University of Tennessee at Knoxville. Graduate courses gave her another opportunity to integrate her interests in mathematics, technology, and education. She was awarded a Doctor of Philosophy degree in education in December of 2000.