



RESEARCH REPORT

Effectiveness of Graphing Calculators in K–12 Mathematics Achievement: A Systematic Review

Madhab Khoju
Empirical Education Inc.

Andrew Jaciw
Stanford University

Gloria I. Miller
Empirical Education Inc.

November 21, 2005

About Empirical Education Inc.

Empirical Education Inc. is an independent research company based in Palo Alto, California. It was founded in 2003 with the mission of helping school districts make cost-effective decisions in adopting new instructional and professional development programs. Empirical Education Inc. uses rigorous quantitative research methods combined with qualitative data collected from teachers and classrooms. For a program being piloted in a school district, these methods quantify its potential impact in terms of the improvement in student achievement that a school can expect. Empirical Education Inc. provides research services to K–12 school districts, as well as to commercial and nonprofit developers and publishers, to determine the effectiveness of programs. The company draws on the expertise of world-class researchers and methodologists, assuring that the research is objective and takes advantage of current best practices in experimental design and analysis.

Empirical Education Inc.
425 Sherman Avenue, Suite 220
Palo Alto, CA 94306
(650) 328-1734
www.empiricaeducation.com

Executive Summary

In this report we systematically review research that examines the effect of calculator use, including the graphing calculator, on K–12 students' mathematics achievement. Our goal was to determine whether there is scientific evidence of effectiveness of graphing calculator use on students' mathematics learning. A thorough review of the research literature and a careful examination of the methods used narrowed our selection of reports to those that used acceptable methods and adequately reported quantitative findings. We summarize a total of 13 studies. For four of these studies, which address the impact of graphing calculators specifically on algebra achievement, we conducted a meta-analysis, yielding evidence of a strong effect of the technology.

Selection of Qualified Research

To support the emphasis of the No Child Left Behind Act of 2001 (NCLB) on teaching methods with evidence of effectiveness, the U.S. Department of Education established the What Works Clearinghouse (WWC) in 2002. The clearinghouse has established the WWC Study Review Standards, which research studies must pass to be included in their reviews. Our work on this review makes use of a study-screening and classification procedure that closely parallels the one used by the WWC. These criteria were the following:

- The research should assess the effect of calculator (scientific and graphic) use on mathematics achievement.
- The research should be experimental (randomized control or quasi-experimental). The research should be analyzed quantitatively and provide information for calculating effect sizes.
- The research should be conducted in elementary to secondary schools (K–12) levels.
- The research should be published within the past 20 years, i.e., since 1985.
- The research paper should be accessible.

The search led to six published research papers and seven unpublished dissertations. The following list provides the author, publication date, sample student grade levels and mathematics topics covered by the studies.

1. Ruthven, K. (1990) Upper secondary students in England. Symbolization and interpretation. .
2. Graham, A.T., and Thomas, M. O. J. (2000) Year 9 and 10 students in New Zealand. Algebra
3. Thompson, D. R., and Senk, S.L. (2001) Grades 10 and 11 in Chicago. Second-year algebra
4. Hollar, J. C., and Norwood, K. (1999) University freshmen in U.S. Intermediate algebra
5. Autin, N. P. (2001) Grade 12 students in U.S. Trigonometry
6. Drottar, John F. (1998) Grades 10, 11, and 12 U.S. Algebra II
7. Rodgers, K. V. (1995) Algebra II class students U.S. Quadratic equations
8. Wilkins, C. W. (1995) Grade 8 students in U.S. Factoring quadratic equations
9. Szetela, W., and Super, D. (1987) Grade 7 students in Canada. Translation process and complex problems
10. Loyd, B. H. (1991) Grades 8, 9, and 10 in U.S. Subsets of 4 different item types
11. Liu, S. (1993) Grade 5 students in Taiwan. Mathematics computation problem-solving ability

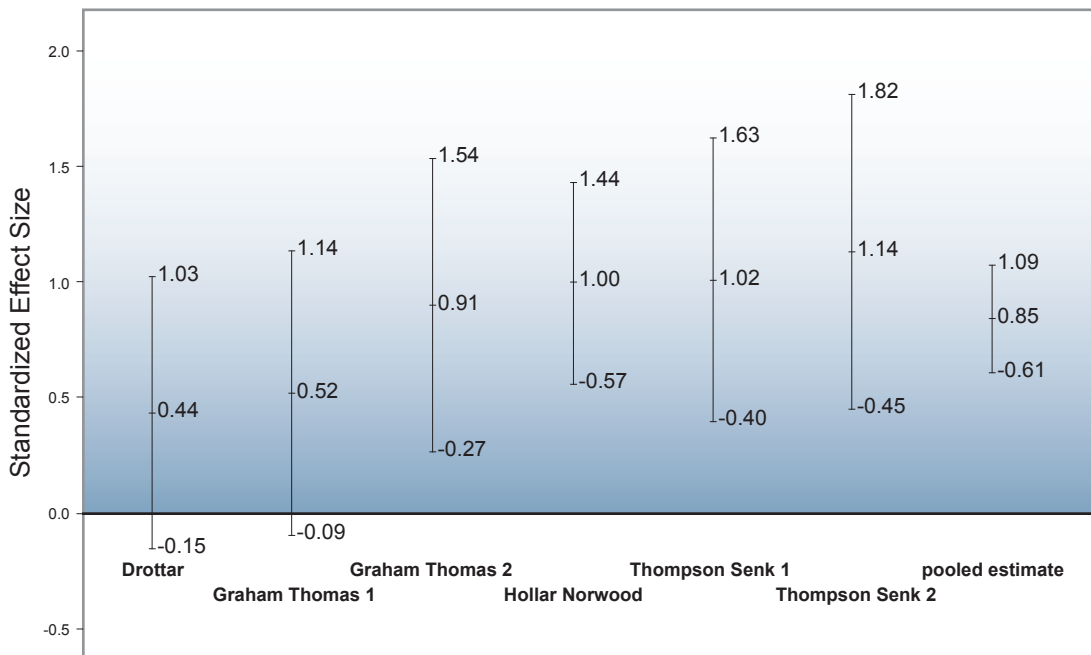
- 12. Ellerman, T. B. (1998) Grades 7 and 8 students in U.S. Mathematics concepts and applications
- 13. Glover, M. A. (1991) Grades 5, 6, 7, and 8 students with Learning Disabilities, U.S. Computation and problem solving

Meta-analysis of Graphing Calculator Impact on Algebra Achievement

A meta-analysis gives us a way of combining the impact of multiple studies to arrive at a single estimate of the impact. Impact is expressed as an effect size, which uses the metric of the standard deviation.

A meta-analysis requires that the studies being combined be studies of the same or closely related educational problems or interventions. First, studies are selected that address similar problems based on researcher judgment. Second, a statistical test of homogeneity is used to verify that the studies have reasonably similar effect sizes. Since our initial focus of the review was on graphing calculators, we restricted the meta-analysis to these studies. There are four published research papers and four unpublished dissertations that investigated the effect of graphing calculators. Among these studies, the researchers measured the impact on a variety of skills and abilities, most commonly on algebra. We judged that four of the studies that met the inclusion criteria measured the effect of using graphing calculators on algebra skills. Our meta-analysis addresses these studies only. Two of the studies report two separate effect sizes. We treated these as separate outcomes, so we worked with six outcomes in the meta-analysis.

We computed standard errors for the effect sizes. We then carried out a statistical test of homogeneity to determine that the studies can reasonably be described as sharing a common effect size. The point estimates for the effect sizes for the six results are displayed in the figure below.



Each point estimate is centered on its 95% confidence interval. The rightmost confidence interval represents the result for the pooled estimate, which has an effect size of .85 and a 95% confidence interval that does not contain zero. This result gives us strong evidence that the use of graphing calculators is associated with better performance in algebra.

Contents

Executive Summary	i
Selection of Qualified Research	i
Meta-analysis of Graphing Calculator Impact on Algebra Achievement	ii
Effectiveness of Graphing Calculators in K-12 Mathematics Achievement: A Systematic Review	1
Selection of Qualified Research	1
Table 1. Sample Student Grade Levels and Mathematics Topics	2
Table 2. Sample Sizes and Interventions	3
Summaries of Research on Graphing Calculators	4
(1) Ruthven (1990)	4
(2) Graham and Thomas (2000)	5
(3) Thompson and Senk (2001)	6
(4) Hollar and Norwood (1999)	7
(5) Autin (2001)	8
(6) Drottar (1998)	9
(7) Rodgers (1995)	10
(8) Wilkins (1995)	11
Summaries of Research on Non-Graphing Calculators	12
(9) Szetela and Super (1987)	12
(10) Loyd (1991)	13
(11) Liu (1993)	14
(12) Glover (1991)	16
(13) Ellerman (1998)	17
Causal Validity	18
Table 3. Causal Validity and Other Study Characteristics: Published Research Papers	18
Table 4. Causal Validity and Other Study Characteristics: Unpublished Dissertations	18
Meta-analysis of Graphing Calculator Impact on Algebra Achievement	19
Table 5. Effect Sizes in Published Research Papers	19
Table 6. Effect Sizes in Unpublished Dissertations	20
Figure 1: For studies of algebra: Estimates of the size of the difference between treatment and control groups indicating the 95% confidence interval	21
References	21

Effectiveness of Graphing Calculators in K-12 Mathematics Achievement: A Systematic Review

The objective of this report is to systematically review the research that examines the effect of calculator use, including the graphing calculator, on K–12 students’ mathematics achievement. Our goal was to determine whether there is scientific evidence of effectiveness of graphing calculator use on students’ mathematics learning. A thorough review of the research literature and a careful examination of the methods used narrowed our selection of reports to those that used acceptable methods and adequately reported quantitative findings. We summarize a total of 13 studies. For four of these studies, which address the impact of graphing calculators specifically on algebra achievement, we conducted a meta-analysis, yielding evidence of a strong effect of the technology.

Selection of Qualified Research

Policymakers in education have been duly concerned about the undersupply of mathematicians and scientists who are critical for global economic leadership and innovation. The No Child Left Behind Act of 2001 (NCLB) was a major effort to improve proficiency of K–12 students through strong accountability for results and an emphasis on teaching methods that have been shown to work through scientifically based research. To support NCLB’s emphasis on teaching methods with evidence of effectiveness, the U.S. Department of Education’s Institute of Education Sciences established the What Works Clearinghouse (WWC) in 2002. The objective of WWC is to facilitate informed decision-making in education. It does this by providing a central source for referral by policymakers, educators, researchers, and the public on educational interventions (programs, products, practices, and policies) that have been shown to improve student outcomes. Although it does not endorse particular interventions, the clearinghouse has established the WWC Study Review Standards, which research studies must pass to be included in their reviews.

Our work on this review makes use of a study-screening and classification procedure that closely parallels the one used by the WWC. The WWC reviews a study in three stages:

- Stage 1: Screening for relevance.
- Stage 2: Determination of whether a study provides strong evidence of causal validity, weaker evidence of causal validity, or insufficient evidence of causal validity.
- Stage 3: Review of other important study characteristics.

The studies for review in this report were selected following the WWC Study Review Standards, including the following:

1. The research should assess the effect of calculator (scientific and graphing) use on mathematics achievement.
2. The research should use randomized control or quasi-experimental methods.
3. The research should be analyzed quantitatively and provide information for calculating effect sizes.
4. The research should be conducted in elementary to secondary schools (K–12)
5. The research should be published within the past 20 years, i.e., since 1985.
6. The research paper should be accessible.

The search for appropriate research reports was done at the library at the University of Illinois at Urbana-Champaign. Priority was given to published journal articles. The following electronic databases were used for the search:

- Educational Resources Information Center (ERIC)
- PsycInfo
- WorldCat
- EBSCO

The references and bibliographies in the research papers that met the above WWC criteria were also used as sources for locating other potential research studies. This search led to six published research papers and seven unpublished dissertations. The objective of most of these studies was to evaluate the benefits of graphing calculators on students' understanding of a particular topic in algebra. Sample student grade levels and mathematics topics covered by the studies are summarized in Table 1. The sample sizes and the interventions of these studies are summarized in Table 2.

Table 1. Sample Student Grade Levels and Mathematics Topics

Study	Student Grades	Math Topics
Ruthven, K. (1990)	Upper secondary students in England	Symbolization and interpretation
Thompson, D. R., and Senk, S.L. (2001)	Grades 10 and 11 in Chicago	Second-year algebra
Hollar, J. C., and Norwood, K. (1999)	University freshmen in U.S.	Intermediate algebra
Graham, A.T., and Thomas, M. O. J. (2000)	Year 9 and 10 students in New Zealand	Algebra
Szetela, W., and Super, D. (1987)	Grade 7 students in Canada	Translation process and complex problems
Loyd, B. H. (1991)	Grades 8, 9, and 10 in U.S.	Subsets of 4 different item types
Autin, N. P. (2001)	Grade 12 students I in U.S.	Trigonometry
Drottar, J. F. (1998)	Grades 10, 11, and 12 in U.S.	Chapter 6 and 7 in Algebra II
Wilkins, C. W. (1995)	Grade 8 students in U.S.	Factoring quadratic equations
Rodgers, K.y V. (1995)	Algebra II class students in U.S.	Quadratic equations
Glover, M.I A. (1991)	Grades 5, 6, 7, and 8 students with Learning Disabilities, U.,S.	Computation and problem solving
Ellerman, T.e B. (1998)	Grades 7 and 8 students in U.S.	Mathematics concepts and applications
Liu, S.. (1993)	Grade 5 students in Taiwan	Mathematics computation problem-solving ability

Table 2. Sample Sizes and Interventions

Study	Sample Size	Intervention
Ruthven, K. (1990)	47 in treatment group; 40 in comparison group	Different teachers in treatment and comparison groups but same curriculum. Treatment group with regular access to calculators.
Thompson, D. R., and Senk, S. L. (2001)	22 and 16 in treatment classes vs. 24 and 23 in comparison classes	UCSMP and regular algebra curriculum. UCSMP group with access to graphing calculators. Different teachers.
Hollar, J. C., and Norwood, K. (1999)	46 in treatment group; 44 in comparison group	Textbook with graphing calculator activities and access to graphing calculator for treatment group vs. regular textbook without calculator in control group. Different teachers.
Graham, A.T., and Thomas, M. O. J. (2000)	21 in treatment and 21 in comparison in each of two sets of classes	“Tapping into Algebra” module with graphing calculator in treatment group vs. normal teaching in control group. Different teachers.
Szetela, W., and Super, D. (1987)	290 students in 14 classes in CP group; 195 in 10 classes in P group; 338 in C group	Problem-solving strategies with calculators (CP), problem-solving strategies without calculators (P), and no problem-solving strategies and no calculator group (C).
Loyd, B.a H. (1991)	4 groups of 40 examinees, 70 with calculator, 90 without	Four subsets of items, some favoring calculator use and others problematic with calculator use
Autin, N. P. (2001)	29 in treatment and 29 in comparison groups. All male students.	Researcher and classroom teacher team-taught both classes. Same syllabus and textbook except graphing calculator use for treatment group.
Drottar, J. F. (1998)	22 in treatment and 23 in comparison group for first part, 19 and 21 in second	Both treatment and comparison groups were taught by the researcher and used the same UCSMP textbook. Graphing calculator to treatment group.
Wilkins, C. W. (1995)	75 in treatment group; 24 in comparison group	Researcher taught the treatment group; second teacher taught control groups. Same textbook but treatment group had graphing calculators.
Rodgers, K. V. (1995)	17 in treatment class; 21 in comparison class	Both classes taught by the same teacher using same textbook, content and activities. Calculator group used graphing calculators.
Glover, M.I A. (1991)	35 in treatment group; 33 in comparison group. Learning-disabled students.	Experimental students trained in Math Explorer calculator prior to calculator instruction in regular class. Control students with no Math Explorer training.
Ellerman, T. B. (1998)	579 in treatment group; 491 in control group	Teachers required to provide calculators to treatment group on the day of the test.
Liu, S. (1993)	43 in T group; 50 in C group; 53 in P group; 47 in C plus P group	Four classes randomly selected as Traditional (T) group, Calculator group (C), Problem-solving group (P), Calculator plus Problem-solving group (CplusP)

Summaries of Research on Graphing Calculators

There were only four published research studies and four unpublished dissertations examining the effect of graphing calculators on mathematics achievement. Each of these research studies is summarized below.

(1) Ruthven (1990)

K. Ruthven compared the performance of students of upper secondary school mathematics classes with graphing calculators to other students who were matched based on similar background and curriculum but without graphing calculators used to improve their understanding of algebraic functions. Such matched classes were identified in four English secondary schools. Of the two classes in each school, students in one class had regular access to graphing calculators (treatment), while students in another class did not have access to graphing calculators (comparison). Students were tested on two sets of problems—one set consisting of symbolization items (requiring students to write the equation for a given graph) and another of interpretation items (requiring students to extract information from a given graph).

The Graphic Calculators in Mathematics project in England had enabled each teacher in six small groups of classroom teachers to work with at least one class of students with calculators for a two-year advanced-level mathematics course. The participating teachers did not have any previous experience with graphing calculators. These teachers were not required to follow any prescribed program of calculator activities and planned their own classroom work, but met periodically to exchange ideas and review progress. Four schools in the project identified classes (comparison group) that were parallel to a project class (treatment group), similar in previous attainment and following the same mathematics course, but differing only in their access to graphing calculators. In addition to some background information, including their mathematics grade in GCSE (an external examination taken before attending the current course), a 40-minute test containing 12 graphing items was administered. The resulting sample consisted of 87 students; 47 were in the treatment group and 40 were in the comparison group. However, 7 students in the comparison group who had their own graphing calculators were dropped from the group. Based on background information, the two groups were comparable (similar) in their abilities. Scores on symbolization and interpretation items on the test administered near the end of the first year of the course constituted outcome measures.

Several considerations were taken into account in designing the test. First, the test covered materials drawn from two topic areas central to any advanced-level course, where the use of graphs is normal practice. Second, the test items were designed to test competencies for which there is no automatic graphing calculator procedure.

At the end of the first year of the two-year advanced-level mathematics course, the students were administered a 40-minute test. Of the 12 items in the test, the first 6 were symbolization items and the second 6 were interpretation items.

The covariance analysis of students' test scores indicated significant treatment effect on symbolization items but not on interpretation items. The treatment group outperformed the control group in symbolization items, with the effect size of 1.81. Moreover, there was also a significant treatment gender interaction for symbolization items. The female students outperformed male students in the treatment group but were outperformed by male students in the comparison group.

(2) Graham and Thomas (2000)

A.T. Graham and M. O. J. Thomas were motivated by the research findings of Tall and Thomas, 1991, which demonstrated improvements in students' algebra performance using computer activities. Since a graphing calculator is portable and an affordable alternative to computers for many schools, this study sought to analyze whether students' performance in algebra can be significantly improved by using graphing calculator activities. The researchers used the "Tapping into Algebra" module—a classroom-based research program that uses an experimental design to compare the teaching of the concept of 'variables' in algebra with and without the use of a graphing calculator. The students in the treatment and comparison groups were similar in ability and background. The study compared the pretest and posttest performances of treatment and comparison groups of students in two schools in New Zealand. The tests were designed to measure understanding of the use of letters as specific unknowns, generalized numbers, and variables in elementary algebra. The treatment groups significantly outperformed the comparison groups on the posttest, even though there were no differences on the pretest.

Although teachers from six New Zealand schools volunteered to take part in this research project, comparison groups similar to the treatment groups in ability and background were found only in two schools. Of the 147 treatment students in six classes and 42 students in two comparison classes, 118 were from year 9 (age 13 years) and 71 from year 10 (age 14 years), and covered different ability groups. Since comparison classes similar to treatment classes were found in only two schools, the results reported here are based on those four classes—two treatment and two comparison classes. Each of these classes had 21 students. The students in these classes did not differ much in their abilities based on pretest results. The "Tapping into Algebra" module was taught during terms one and two of 1996 by the classroom teachers, and a graphing calculator was provided to each student in the treatment class. The comparison classes received algebra work similar to the treatment group but were taught by different teachers using their normal teaching program. The researchers were not present in any of the classrooms, and the teachers were encouraged to use their normal teaching approach.

Both the treatment and comparison groups were administered a pretest and posttest based on Kuchemann's (1981) study comprising 68 questions. Students were not given their papers or any answers to the questions until after the posttest. Student scores on the posttest constituted the outcome measures in this study. The maximum possible score was 68. The outcome measures were compared between the treatment and comparison groups separately for each of the two schools with control groups.

The research design for this study can be considered quasi-experimental. The sample students in the treatment group were the students in classes of six teachers who volunteered to take part in this research. Since comparison groups similar to treatment groups in ability and background were found only in two schools, t-tests were used to compare the posttest performance between the treatment and comparison groups separately for each of these two schools only. In each school, the treatment group significantly outperformed the control group ($p < 0.05$). The posttest scores of the remaining treatment classes in four other schools, used as a triangulation group, showed similar gains. The information about the means and standard deviations in the pretest and posttest were used to calculate the effect sizes following Chen (1994, p.91). The effect size was 0.249 for school A and 0.485 for school B¹. The study did not report detailed gender information about students.

¹The effect sizes reported here are computed using a method that adjusts for discrepancies in performance between the treatment and comparison groups prior to intervention. This yields a more conservative estimate than the commonly used measure of effect size, which is based on the posttest only. For purposes of meta-analysis, the more commonly employed estimates are used. For this study they are .52 and .91, respectively.

(3) Thompson and Senk (2001)

D. R. Thompson and S. L. Senk compared student achievement in second-year algebra between the University of Chicago School Mathematics Project (UCSMP) classes and comparison classes.

Participants in the study were recruited through advertisements in UCSMP and NCTM publications. A school needed at least four sections of second-year algebra, two UCSMP classes, and two comparison classes, and the staff had to promise to keep classes intact for a full year. UCSMP and comparison classes were expected to have “similar students who have had the same previous work.” The evaluators used a matched-pair design for the study. A pretest measuring entering algebra and geometry knowledge was given over two days to assess the proficiency of the students. This pretest developed by UCSMP is composed of 46 multiple-choice items. This test was used to match UCSMP and comparison classes in the same school. Two well-matched pairs were formed in each school. Even though UCSMP and comparison classes were not assigned randomly, the teachers of the two groups of students had comparable academic backgrounds. The difference in the pretest score means of the two classes (within each pair) was not significant even at $p=0.25$. Students using UCSMP materials were expected to have continual access to graphing technology (calculators or computers). The research design for this study can be considered quasi-experimental.

Four schools that participated in the study represented a broad range of educational and socioeconomic conditions in the United States. These four schools were one each from Georgia, Illinois, Mississippi, and Pennsylvania. In each school, two classes used advanced algebra materials produced by the UCSMP, and two other classes used regular textbooks. The texts used in the comparison classes and in UCSMP advanced algebra overlap considerably. To eliminate potential teacher selection bias, in each school each teacher had to agree to teach either curriculum before assignment. In each school, two teachers were assigned to two sections using UCSMP advanced algebra, and the other two teachers were assigned to two comparison classes which used the textbook currently in place at the school.

UCSMP advanced algebra is compatible with a variety of instructional styles. Instead of depending primarily on lecture to introduce content, teachers are also asked to pose problems, engage students in class discussion, and encourage students to learn to read their textbooks. UCSMP advanced algebra and the comparison texts treat technology very differently. The UCSMP developers assume that graphing calculators are available for student use at all times. The comparison texts' authors do not assume that any calculators will be used, although optional activities are included for use with scientific calculators.

A total of 150 students were in the UCSMP classes, and 156 students were in the comparison classes. The performance of students is measured in eight pairs of second-year algebra classes that had been matched on the basis of pretest scores at the start of the school year. Since only the comparison students in the school in Chicago did not own calculators, only the results from this school are considered. In this school, one treatment class had 22 students compared to 24 students in its matching comparison class. Similarly, another treatment class had 16 students compared to 23 in its matching comparison class.

About two weeks before the end of the school year, teachers administered several instruments, including a multiple-choice posttest to assess students' knowledge of the content of second-year algebra. The posttest contained 36 items. However, both UCSMP and comparison teachers at the Chicago school reported that their students had the opportunity to learn the needed content only for 26 items, and so a test containing these 26 items was called a fair test. The reliability of the fair test was 0.635. Similarly,

there were 15 items for which all the teachers in the study indicated that their students had opportunities to learn the needed content, and so a test containing these 15 items was called a conservative test. The reliability of the conservative test was 0.635.

The results of a matched-pairs t-test indicated significant ($p < 0.05$) differences between two curricula. The UCSMP students outperformed comparison students in the fair and conservative test in the Chicago school. The USCMP group outperformed the control group in the fair test, and the effect size was 1.02 in one matched pair of classes and 1.14 in the second matched pair. Similarly, the USCMP group outperformed the control group in the conservative test, The effect size was 0.80 in one matched pair of classes and 0.82 in the second matched pair.

(4) Hollar and Norwood (1999)

J.C. Hollar and K. Norwood extended O'Callaghan's study by comparing students using a graphing approach to the curriculum with the aide of TI-82 graphing calculators with students using a traditional approach. The function concept in mathematics is one of the most central concepts. O'Callaghan studied the effects of the Computer-Intensive Algebra (CIA) curriculum on college algebra students' understanding of the function concept by comparing students using CIA with students using a traditional curriculum. He developed a test to assess students' understanding of functions. Each question on the test was designed to assess one of the following aspects of conceptual knowledge: (1) modeling a real-world situation using a function; (2) interpreting a function in terms of a realistic situation; (3) translating among different representations of functions; and (4) reification (transitioning from the operational to the structural phase of using functions). O'Callaghan (1998) found that CIA students were better than traditional students in understanding modeling, interpreting, and translating concepts but no different in reification. The objective was to examine the effects of using a graphing approach to the curriculum on each of the four aspects of conceptual knowledge of functions.

The participants in this study were students enrolled in Intermediate Algebra at a large state university. These students scored the lowest on the university's mathematics placement examination. Four sections of a semester-long intermediate algebra course taught by two instructors were used in this study. Of the two instructors, each taught one treatment and one control class. A sample of 90 students participated in this study—46 in the treatment group and 44 in the control group.

One of the two simultaneous morning sections and one of the two simultaneous afternoon sections were selected to use the experimental curriculum. To determine any initial differences among the four classes, researchers used ANOVA procedures to compare the classes in terms of the following outcomes: results of the O'Callaghan Function Test pretest, math background (number of previous algebra courses); mathematics ability (math SAT scores); and predicted grade-point average in mathematics calculated by departmental formula. The analysis indicated that the four classes were similar. Similarly, pretest scores indicated no significant differences among the four classes on prior knowledge of functions.

The instructors followed the same plan of study, adhering to the course syllabus. From interviews and random observations of the classes, the researchers concluded that the instructors were not biased to (in favor of) any approach.

In the treatment group, the college text *Intermediate Algebra: A Graphing Approach* (Hubbard & Robinson, 1995) included calculator activities and was used in conjunction with the TI-82 graphing calculator. The text consists of both the graphing calculator activities and traditional algebra work. The

students had access to calculators and were able to explore, estimate, and discover graphically and to approach problems from a multi-representational perspective. However, the students did not have access to calculators for the O'Callaghan Function Test or the traditional final examination.

In the comparison class, the text *Intermediate Algebra: Concepts and Applications*, fourth edition (Bittinger, Keedy, & Ellenbogen, 1994), was used, and the text covered the same topics as the experimental text. The focus of the text was on simplifying and transforming expressions and solving equations. The comparison group had no known access to graphing calculators.

The O'Callaghan Function Test was administered without access to calculators, first as the pretest at the beginning of the semester and later as a posttest at the end of the semester. Each question on the test was designed to assess one of the following aspects of conceptual knowledge: (1) modeling a real-world situation; (2) interpreting a function in terms of a realistic situation; (3) translating among different representations of functions; and (4) reifying functions. To evaluate students' traditional algebra skills, a departmental final examination consisting of a 50-question test of conventional algebra skills was used. The traditional final examination was administered to all four classes during the final week of the semester.

MANOVA was used to analyze students' understanding of the function concept on the four component scores and the total score on the O'Callaghan Final Posttest. MANOVA results indicated that the treatment classes outperformed the comparison classes in O'Callaghan's Function Test and also in each of the four components of the test. The effect size for the total test was 1.00. The effect sizes for the four components are 0.60 for modeling a real-world situation, 0.70 for interpreting a function in terms of a realistic situation, 0.64 for translating among different representations of functions, and 5.03 for reifying functions.

(5) *Autin (2001)*

Nancy P. Autin investigated the impact of the use of graphing calculators on both students' understanding of inverse trigonometric functions and on their problem-solving approaches. It is an effort to investigate topics for which integrating graphing technology in mathematics teaching is well-suited. Students in two 12th-grade trigonometry classes at a large, metropolitan, all-male private high school in Louisiana constituted the sample in this study. Each of these students had completed full-year state-approved courses in algebra I, algebra II, and geometry. One of the two classes involved in this study was randomly chosen as the treatment class, and the other as the comparison class. Each of the two classes contained 29 students for a total of 58 students: 55 white, 5 black, 2 Vietnamese, and 1 Hispanic. The researcher and the classroom teacher team-taught both classes for two weeks, following the same syllabus and using the same textbook, except that the treatment class was allowed to use a graphing calculator.

A pretest was administered to measure students' understanding of the general nature and behavior of functions. An F-test indicated no significant difference in pretest scores between the two classes. Students' algebra II grades and ACT math scores were used to further investigate whether students in the two classes had a similar understanding of functions at the beginning of the study. An independent samples t-test indicated no significant differences between the classes. A posttest consisting of two parts was administered on the final day of instruction. Part 1 consisted of 20 short-answer questions; Part 2 had six free-response questions. The six free-response items required students to justify their responses in a variety of ways, including through the use of graphs, and algebraic arguments. Scores on the posttest were the sum of raw scores in Part 1 and Part 2 of the test. The maximum possible score on the pretest is

60, on the posttest Part 1 it is 72, on the posttest Part 2 it is 30, and for the total posttest it is 102.

Analysis of covariance was used to test for a difference in understanding of inverse trigonometric functions at posttest between the treatment and comparison classes. The pretest scores were used as the covariate in the study in order to account for preexisting differences that may have existed between the intact groups. ANCOVA was chosen since it is considered to be an appropriate procedure for adjusting for preexisting differences between two intact groups. Further, ANCOVA, which combines regression and analysis of variance, controls for the effects of extraneous variables, and increases the precision of the research by reducing error variance (Hinkle, Wirsman, and Jurs, 1998, p. 518).

F-tests indicated significant differences in the total posttest scores between the treatment and control classes. The treatment class significantly outperformed the comparison group in both total posttest scores and scores in Part 2 of the posttest. However, there was no significant difference between the two classes in Part 1 of the posttest. The effect sizes were 0.64 for Part 1 of the posttest, 1.02 for Part 2 of the posttest, and 0.91 for the total posttest.

(6) Drottar (1998)

John F. Drottar compared the impact of graphing calculator on both the overall math performance and four particular aspects of student understanding as defined by the University of Chicago School Mathematics Project (UCSMP): Skills, Properties, Representations, and Uses. Both the treatment and comparison groups were taught by the same teacher following the same curriculum, except that the students in the treatment class were allowed to use graphing calculators.

Students from two intact algebra II A-level (with average to above average ability) classes at a four-year suburban high school in eastern Massachusetts participated in this study. Using the flip of a coin, one of the two classes was chosen as the treatment group and the other, as the comparison group. Both groups used the UCSMP advanced algebra textbook and were taught by the same teacher (the researcher of this study). The content and pacing as well as instructional strategies were the same for both classes. The treatment group differed from the control group only in its access to graphing calculators (TI-83). Chapters 6 and 7 were covered in the study. To measure performance, for both Chapters 6 and 7, Form A was used as a pretest and Form B as a posttest. These chapter tests have specific questions relating to each of the four components: skills, properties, uses, and representations. The study compared the treatment group with the control group on overall performance and on each of the four components.

The treatment group for the first part of the study included 22 students (10 males and 12 females), of whom 9 were in grade 10, 10 in grade 11, and 3 in grade 12. Similarly, the comparison group included 23 students (16 males and 7 females), of whom 13 were in grade 10, 7 in grade 11, and 3 in grade 12. Based on t-test results on Chapter 6 pretest scores, the treatment group was not significantly different from the comparison group. The issue of ability equivalency between the groups was further explored by comparing students' previous year's math grades. A t-test indicated no significant difference between the two groups in the students' previous year's math grades. Some students dropped out of the school in the second part of the study when the treatment and control groups were switched for Chapter 7 tests. As a result, in the second part of the study, the treatment group included 19 students and the control group, 21 students. One male Caucasian student in the control group and four students (1 female Caucasian, 2 male Caucasian, and 1 Hispanic male) in the treatment group dropped out. A t-test on Chapter 7 pretest data indicated no significant difference between the two groups.

In the first part of the study, students' performance on the Chapter 6 posttest constituted the outcome measure. The test also identifies the questions related to each of the four components of understanding: skills, properties, uses, and representations. Similarly, the Chapter 7 posttest performance constituted the outcome measure in the second part of the study. The test also identifies the questions related to each of the four components of understanding. In each of these chapter posttests, 10, 4, 10, and 4 questions were related to skills, properties, uses, and representations, respectively, for a total of 28 questions.

In the first part of the study based on the Chapter 6 posttest, the treatment group outperformed the control group, and the effect size was 0.440. However, the calculated t-statistic of 1.50 for the difference was not statistically significant. Of the four components of understanding, the treatment group significantly outperformed the control group only in the area of the representations component.

Similarly, in the second part of the study based on the Chapter 7 posttest, the treatment group also outperformed the control group, and the effect size was 0.303. However, the calculated t-statistic of 1.05 for the difference was not statistically significant. Of the four components of understanding, the treatment group significantly outperformed the control group only in the area of the skills category.

(7) Rodgers (1995)

Kathy V. Rodgers analyzed the impact of supplementing the traditional algebra II curriculum with graphing calculator activities on achievement scores, retention scores, and students' attitudes towards mathematics for average ability students. The students in two intact standard (average ability) algebra II classes at a four-year high school in rural western Kentucky are the study participants. Students in these classes were of average ability (based on their past performance in math) and were randomly assigned to one of the two classes by the school's computer-scheduling program before the beginning of the classes. The same teacher taught both classes, and one of the classes was randomly assigned (by a flip of a coin) to be the treatment class and the other to be the comparison class. Both the treatment and control classes were taught by the same teacher; the content, examples, assignments, and activities were identical for both classes except the treatment class was allowed to use graphing calculators (TI-82). The research was focused on the study of quadratic equations.

The treatment class consisted of 17 students; the control class, 21 students. The differences in the achievements of these students in the pretest and posttest constituted the dependent variable. A maximum score of 100 was possible for both the pretest and posttest. All the problems in the tests could be solved without the use of a graphing calculator. Students were required to solve the first three items in the tests using the traditional method and display paper-and-pencil calculations, while other items could be solved with or without graphing calculators. Treatment and comparison classes were also compared separately on their achievement in paper-and-pencil items and other problem-solving items. KIRIS (Kentucky Instructional Information System) scores (based on a combination of performance-based questions and traditional multiple-choice questions) of these students constituted the covariate in the analysis of covariance (ANCOVA). The maximum possible score for the paper-and-pencil items as well as for problem-solving items in the test was 18. Students' semester averages from the fall semester were also used separately as the covariate in the ANCOVA. The treatment and comparison classes were equivalent in terms of their KIRIS scores and also their previous fall semester averages.

This study utilized ANCOVA to test for a difference between pretest and posttest achievement on items related to quadratic equations. Students' KIRIS scores and previous fall semester averages were separately used as the covariates. ANCOVA results with KIRIS scores as a covariate indicate that supplementing the traditional algebra II curriculum with graphing calculator activities improved overall

achievement. The treatment class therefore outperformed the comparison group in overall achievement. The effect size of 0.75 indicated that the treatment group outperformed the control group by 0.75 of a standard deviation. Similarly, ANCOVA results with students' previous fall semester averages as a covariate indicate that supplementing the traditional algebra II curriculum with graphing calculator activities improved overall achievement. The treatment class outperformed the control group in overall achievement.

ANCOVA results for the difference scores in paper-and-pencil items between the pretest and posttest achievements with KIRIS scores as a covariate indicated that supplementing the traditional algebra II curriculum with graphing calculator activities worsened paper-and-pencil achievement. The comparison class outperformed the treatment class. The effect size of -1.11 indicated that the control group outperformed the treatment group by 1.11 standard deviations. On the other hand, ANCOVA results for difference scores on problem-solving items with KIRIS scores as a covariate indicated that supplementing the traditional algebra II curriculum with graphing calculator activities improved achievement on problem-solving items. The effect size of 6.79 indicated that the treatment group outperformed the comparison group by 6.79 standard deviations.

(8) Wilkins (1995)

Cynthia W. Wilkins examined the effect of integrating graphing calculator use into the study of factoring in an eighth-grade algebra I program of study. The objectives of the study included investigating two research questions: (1) whether students who are taught to factor by using a graphing calculator perform significantly better than students taught traditionally without a graphing calculator, and (2) whether the effect of graphing calculator use is different between male and female students. Since the National Council of Teachers of Mathematics (NCTM) has recommended the use of graphing technology beginning in eighth grade at the pre-algebra level of math instruction, this study examined whether the graphing calculator was helpful to students at that level.

The sample included eighth-grade students enrolled in two schools in Mississippi. Seventy-five students in three classes in a public school constituted the treatment group; 24 students in a class in a parochial school constituted the control group. Of the 75 students in the treatment group, 40 were female and 35 male. Similarly, of the 24 in the control group, 14 were female and 10 male. The researcher taught all three classes in the treatment group, while another teacher taught the comparison group. The researcher selected the comparison group teacher based on that teacher's attitude, teaching style, teaching philosophies, and collaborative work experience. The researcher and comparison group teacher had different approaches to presenting the unit in factoring. Both teachers used the same textbook, but the researcher developed a unit consisting of 10 lessons that integrated the graphing calculator (TI-81) into her instruction; the textbook was used only as a reference tool. The comparison group teacher followed the lesson order and format in the textbook. The comparison group teacher also supplemented the text with some additional materials. The researcher trained the comparison-group teacher in factoring methods that were used in the treatment group. The comparison group also had access to graphing calculators; however, the comparison group teacher as well as all the teachers in his/her school were not trained in how to incorporate graphing calculators into the factoring unit, so the risk of experimental diffusion was low.

Both the treatment and comparison groups took the same pretest, the Stanford Achievement Test, and the same posttest. A panel of experts and an outside evaluator established the content validities of the pretest and posttest. No reliability estimates for the pretest and posttest were given. Both groups were

given the pretest immediately prior to the five week period devoted to this unit of study. Of the 25 multiple-choice problems in the pretest, 12 problems in Section A were designed to measure basic factoring skills; 3 word problems in Section B were designed to measure basic applications of factoring skills; and 10 problems in Section C were designed to measure concepts and understanding beyond the basic level.

The independent sample t-test indicated a significant difference in prior ability in Section A of the pretest (basic factoring) between the treatment and comparison groups but not in Sections B (basic applications of factoring skills) and C (concepts and understanding beyond the basic level). The groups were also significantly different in prior ability in basic math skills as measured by Stanford Achievement Test scores. These scores were used as covariates in the analysis of covariance. The last day of the five-week study period was used to administer the posttest. The posttest was an alternate form of the pretest.

ANCOVA was used to test for a difference in scores in sections A, B and C of the posttest, with pretest scores and Stanford Achievement Test scores used as covariates to account for preexisting differences between the intact groups. The results indicated that the treatment and comparison groups differed significantly in basic applications of factoring skills (Section B), and concepts and understanding beyond the basic level (Section C) but not in basic factoring skills (Section A of the posttest). The treatment group outperformed the comparison group in Sections B and C but not in Section A. Since the adjusted means were not reported, the effect sizes were based on posttest means and standard deviations. The effect sizes were -0.25 in Section A, 0.41 in Section B, and 2.42 in Section C. T-tests also indicated no significant differences between male and female students in either the pretest or posttest scores.

Summaries of Research on Non-Graphing Calculators

The National Council of Teachers of Mathematics has recommended the use of graphing technology beginning in eighth grade at the pre-algebra level of math instruction. Since the council as early as 1980 had recommended the use of calculators at all grade levels, this review also included a few studies examining the effect of other calculator use on elementary and middle school children's mathematics achievement. One study specifically investigated the effect of calculator use on mathematics achievement of students with learning disabilities. Each of these is summarized as follows.

(9) Szetela and Super (1987)

W. Szetela and D. Super compared performance in mathematics for three groups of seventh-grade students in British Columbia, Canada. Teachers adopted problem-solving strategies with calculators (CP group) with the first group, problem-solving strategies without calculators (P group) with the second group, and no problem-solving strategies and no calculators (C group) with the third group. The following instruments were used in the study:

- Operations with Whole Numbers Test (PREOP) and Operations with Rational Numbers Test (RAT). Each of these tests was a 40-item multiple-choice test used in British Columbia. The reliability indices were 0.88 for PREOP and 0.91 for RAT.
- Translation Problems Tests (TRAN1 and TRAN2). Each of these tests, which consist of 20 translation problems, was constructed and pilot-tested by the authors and was aimed at measuring the performance on elementary school math problems. TRAN1 was administered at midyear and TRAN2 at the end of the year. Reliability indices were 0.75 for TRAN1 and 0.72 for TRAN2.

- Process Problem Tests (PROP1 and PROP2). Each of these tests consists of 20 process problems and was constructed and pilot-tested by the authors. Strategies taught in the two problem-solving groups—CP and P—were needed to solve these problems. Reliability indices were 0.78 for PROP1 and 0.77 for PROP2.
- Complex Problems Test (COMP). A four-item test of complex problems was constructed and pilot-tested by the authors to determine whether teaching problem-solving strategies resulted in superior performance in the complex problems than in the translation and process problems.

PREOP was administered at the beginning of the year, TRAN1 and PROP1 were administered midyear, and the three tests (TRAN2, PROP2 and COMP) were administered in one sitting at the end of the year. The performance data were analyzed by using a partially nested analysis of covariance with treatment and sex nested within class. The pretest scores on PREOP were used as the covariate. This method of analysis effectively treats the class as the unit of analysis. The CP group scored significantly higher than the C group on TRAN1 and TRAN2 tests.

The study involved a total of 42 classes. Of these, 14 classes with 290 students were in the CP group, 10 classes with 195 students in the P group, and 18 classes with 338 students in the C group. Although test results were available for 42 classes for the midyear tests, the results for only 36 classes were available for the end-of-year tests. Three teachers in the C group, one teacher in the P group, and two teachers in the CP group dropped out of the study. Based on the results of a pretest, the three groups were not significantly different in their knowledge of whole-number operations.

This study used analysis of covariance with treatment by sex nested within class to analyze test score differences between groups. The outcome measures that were collected at the end of the year consisted of scores on two tests—TRAN2 and PROP2—which tested translation and process problems, respectively. Each test consisted of 20 items. PREOP scores were used as the covariate in the analysis of covariance of mathematics achievement data.

The ANCOVA results indicated significant treatment effects for TRAN2 and PROP2. The information about the means and standard deviations in the report were used to calculate the effect sizes. Following Glass, McGaw, & Smith (1981), the standard deviation of the comparison group was used to calculate the effect size. The effect size for TRAN2 between CP and P groups was 0.17 and between CP and C groups was 0.374. Similarly, the effect size for PROP between CP and P groups was 0.152 and between CP and C groups was 0.434.

The calculator effect was also compared between gender groups. There were no significant differences in TRAN2 and PROP2 scores between boys and girls in each group.

(10) Loyd (1991)

Brenda H. Loyd examined four item types on which performance was expected to vary differentially depending on conditions of calculator use. The identification of item subtypes as they relate to calculator use could be used to increase predictability of test score results with and without calculator use in a standardized testing situation. The study was motivated by previous research that had provided conflicting findings about whether using calculators changes the difficulty of mathematics tests or the time needed to complete them.

One hundred and sixty students attending a summer enrichment program at a state university during the summer of 1988 participated in this study. Twenty-seven students were 13 years old, 64 were 14

years old, 50 were 15 years old, 18 were 16 years old, and 1 was 17 years old. In the group 45% were in the eighth grade, 36% were ninth grade, 18% were 10th grade, and 1% was in 11th grade. Of the 160 students, 69 were boys and 91 female. Ten percent were black, 83% were white, and 7% were of other races. Ninety percent of the students owned their own calculators.

The math test administered to students was a composite of four subsets of items. The first subset of eight items was developed to favor examinees who were allowed to use calculators. This set included items that involved a more difficult level of computation as well as items requiring estimation, for which calculators could be used to approximate results. The second subset of eight items was developed as items that could be answered using a calculator, but could also be answered without using a calculator. These items were designed so that use of a calculator did not provide an advantage over the non-calculator group. The third subset of eight items required examinees to select the correct strategy or setup rather than a numerical answer. For this set of items, the use of a calculator would not be applicable. The fourth subset of eight items was more difficult or problematic for those using the calculators.

Four groups of 40 examinees were administered the 32-item test. Eighteen identical TI-1706 II solar-powered calculators were available for the study. Within each group, half of the students were allowed to use a calculator. Among the students seated for the test, half were randomly selected and assigned calculators. The students with calculators were permitted to use them, but there was no requirement that the calculator be used.

To examine whether there was a difference in the performance on the four subsets of items between students who were allowed use of the calculator and those who were not, a two-group discriminant analysis was used with the group variable consisting of an indicator of calculator use or nonuse. The four predictor variables were the scores on the four subsets. A significant discriminant function was followed up with t-tests for each subset.

Of the 160 students, 70 were allowed to use a calculator and 90 were not allowed to use a calculator. The results of the discriminant analysis indicated that the two groups could be distinguished in terms of their performance on the four subsets. The t-tests indicated a significant difference between the two groups on the first set of items but not in the other three subsets. The findings of the study support the contention that high school students' performance on math tests is affected by calculator use. The effect of calculator use also differs by item types.

(11) Liu (1993)

Shiang-tung Liu examined the effects of teaching calculator use and problem-solving strategies on attitudes towards mathematics, mathematics computation ability, and problem-solving ability of fifth-grade male and female students in Taiwan. Certain professional organizations, like the National Advisory Committee on Mathematics Education (NACOME) and National Council of Teachers of Mathematics (NCTM), recommend the use of calculators for instruction, while other researchers like Elliott (1981), Higgins (1990), and Suydam (1979) argue against calculator use. This study was an effort to investigate whether there were advantages to calculator use in elementary school classrooms.

The subjects in the study were students in four fifth-grade classes from four schools in Taiwan. Each of the four classes was randomly selected and assigned to one of the four treatment groups: traditional, calculator use, problem solving, and calculator plus problem solving. Of the four treatment groups, the traditional group had 43 students (24 males and 19 females); the calculator group had 50 students (23

males and 27 females); the problem-solving group had 53 students (32 males and 21 females); and the calculator plus problem-solving group had 47 students (24 males and 23 females).

Each teacher of the four classes received specific teaching instructions from the researcher. The teachers were asked to maintain the same teaching pace and to give the same amount of practice to students. The researcher occasionally visited the classroom of each teacher to observe the progress of instruction. The Arlin-Hills Attitude Survey (AHAS), the Test of Prior Computation Skills (TPCS), and the Mathematics Problem Solving Ability Scale (MPSAS) were used to examine differences between groups in attitude and ability prior to the intervention.

Students in the calculator use group and calculator plus problem-solving group had access to calculators. The teacher in the traditional group was asked to follow a traditional teaching style. The teacher in the calculator use group was instructed on how to teach students to use calculators and to encourage calculator use in solving problems. The teacher in the problem-solving group was taught Polya's four steps to problem solving and was instructed to have students write down their problem solving processes. The instructions given to the teacher of the calculator use group and the teacher of the problem-solving group were given to the teacher of the calculator use plus problem-solving group. At the end of the nine-week intervention, the students were administered the TCA and posttests of MPSAS and AHAS. The students' scores on these posttest were compared across the four treatment groups.

Students' performance on the three posttests—the Arlin-Hills Attitude Survey (AHAS), the Test of Computation Ability (TCA), and the Mathematics Problem-Solving Ability Scale (MPSAS)—constituted the outcome measures. AHAS measures attitudes towards mathematics, TCA measures computation ability; and MPSAS measures problem-solving ability.

AHAS was developed by Arlin and Hills (1976) to assess fourth-grade to sixth-grade students' attitudes toward mathematics. The AHAS, consisting of 15 questions, was first translated into Chinese, and an English teacher was asked to translate this version back into English. Another English teacher was asked if the translation was appropriate to make sure the two versions were equivalent. The scores for AHAS range from 0 to 15, and the reliability of the pretest Chinese AHAS based on student scores from the four groups was 0.88 and that for the posttest was 0.91.

The TPCS consisted of 28 paper-and-pencil items that were used to measure students' computation skills before the intervention. These items were adapted from textbooks, and the scores ranged from 0 to 28. The reliability for this test was 0.93.

The TCA was designed to measure students' computational ability at the end of the study. The TCA also consisted of 28 items that were adapted by the researcher from students' textbooks. The scores ranged from 0 to 28 and the reliability for the TCA was 0.93.

Similarly, the MPSAS was developed by Liu (1989) to assess the mathematics problem-solving abilities of fifth-grade to eighth-grade-level Taiwanese students. There were two forms of this test: A and B. Form A had 16 items (64 sub-questions) and Form B had 15 items (64 questions). The scores in each form ranged from 0 to 64, and the reliability coefficients for Form B were 0.77 (based on the pretest) and 0.87 (based on the posttest).

The pretest scores on the Arlin-Hills Attitude Survey, the Test of Prior Computation Skills, and the Mathematics Problem-Solving Ability Scale constituted baseline data. These scores were used to examine differences in ability among the groups prior to the intervention.

This study utilized a three-factor analysis of covariance to test for differences on the posttest across the four treatment groups. This was done separately for each posttest. The three factors consisted of treatment status, achievement level, and gender. The researcher ranked the sum of two semesters-worth of mathematics scores for each group from highest to lowest and divided them into three achievement levels—high, middle and low. If the ANCOVA results indicated significant differences across the four treatment groups, then Dunnett’s one-tailed follow-up test was performed to find out which of the groups were different from one another.

Based on the F-ratios from the ANCOVA summary table, the mathematics computation scores for the groups without calculators (traditional and problem-solving) were not significantly higher than those of the calculator use groups (calculator use and calculator use plus problem-solving). This finding indicates that calculator use did not hurt students’ computation ability. However, findings indicate that the calculator use plus problem-solving instructional approach is likely to be the best of the four teaching methods. In addition to comparing posttest scores across the four treatment groups, separate comparisons were also made between males and females. The posttest scores were not significantly different between genders.

(12) Glover (1991)

Michael A. Glover examined the effects of handheld calculator usage on the computation and problem-solving achievement of children with learning disabilities in grades five, six, seven, and eight. Students with learning disabilities tend to lack computational skills that are foundational at the upper elementary and beginning secondary school levels (McLeod and Armstrong, 1982). Therefore, these skills were targeted in the intervention.

All students in this study had been identified by their school district as having a learning disability and were attending regular mathematics classes. The treatment group received mathematics instruction with calculators. Students in this group used the calculator for all homework, quizzes, and tests in the regular math class. They also received instruction in the use of the calculator. The comparison group students with learning disabilities attended regular math classes but didn’t have access to calculators.

Students with learning disabilities in a small (2500 students) rural school district in western New York participated in this study. They were attending regular mathematics classes. The number of students in the treatment group was 8, 9, 8, and 10 in grades five, six, seven, and eight, respectively. Similarly, there were 7, 11, 9, and 6 controls in grades five, six, seven, and eight, respectively. Both the treatment and comparison group students received assistance from their special education teachers, who accompanied them to the regular math classes. The treatment group students were trained in the use of the TI Math Explorer calculator prior to the implementation of calculator instruction in the regular class. Throughout the project, the special education teacher provided the students with calculator instruction as it pertained to the regular mathematics curriculum. The treatment students used the calculator each day during classroom math instruction, while the control group students continued to use paper-and-pencil algorithms to complete assignments. Both the treatment and control group students received assistance from their special education teachers, who accompanied them to the regular math classes.

A 23-item computation test and a 7-item problem-solving test were administered to all students. The items tested addition, subtraction, multiplication, and division of integers and fractions. The treatment and comparison groups were administered the same test both before and after the intervention. Students completed one form of the test using paper and pencil methods and another form using the calculator.

The performance of the students in the treatment group was compared to those of the comparison group. Mean scores of students on pre- and posttests were compared to measure the effect of intervention.

Both the treatment and comparison groups scored higher on both the computation and problem-solving tests when using the calculator than when using pencil and paper methods. Posttest comparisons indicated that the treatment group had significantly higher computation scores when using the calculator. The treatment groups exhibited greater amounts of growth than the control groups. At each grade level, the treatment group outperformed the control group when a calculator was used during posttesting. In three of the four treatment groups, the pencil-and-paper posttest scores were higher than the pencil-and-paper pretest scores. This supports Roberts' (1980) contention that calculator instruction does not harm pencil-and-paper performance, and therefore, the calculator must be introduced early in a child's education.

(13) Ellerman (1998)

Tracie B. Ellerman examined the effects of calculator usage on the mathematics achievement of seventh- and eighth-grade students and also students' and teachers' attitudes towards mathematics. Students from two North Central Louisiana School systems constituted the sample for this study. Students' mathematics achievement was measured by administering California Achievement Tests, Fifth Edition, Form A, Level 17 and 18, Mathematics Concepts and Applications section. Level 17 was designed for seventh graders and Level 18 for eighth graders. The reliability of the 50-item Level 17 test was reported by the test publisher to be 0.77 and that of the Level 18 test was 0.75.

Data for this study were collected during the first semester of the 1997-98 school year. TI-108 calculators were used. The researcher and the school principal randomly assigned the intact classes into treatment or control groups on the day of the test by flipping a coin. Teachers were required to allow the use of calculators in the tests for the treatment group, regardless of how well-integrated calculator use was in the class. Of 1,070 students, 491 were in the control group and 579 in the treatment group; 446 were in seventh grade compared to 624 in eighth grade; 525 were black, 534 white, and 11 others Asian or Hispanic. Of the 33 teachers involved, 28 were females and 5 were males.

The mean scores of the treatment and comparison groups were examined for differences in the number of correct responses in the mathematics concepts and applications section of the CAT. A T-test indicated that the treatment group outperformed the controls in the number of questions answered correctly. This result was statistically significant. The effect size was 0.13. Further, the mean score for male students was significantly higher than for females, with an effect size of 0.05. Results of this study indicate that calculator usage during assessment has a positive influence on student mathematics achievement. Student and teacher survey responses supported calculator usage for both instructional and assessment purposes.

Causal Validity

The causal validity and other characteristics of the studies reviewed in this report are summarized in Table 3 for published research papers and in Table 4 for unpublished dissertations.

Table 3. Causal Validity and Other Study Characteristics: Published Research Papers

Study	Causal Validity	Intervention Fidelity	Outcome Measured	People, Settings & Timing	Testing within SG	Analysis	Statistical Reporting
Ruthven, K. (1990)	Y	○	○	○	●	○	●
Thompson, D. R., and Senk, S. L. (2001)	Y	○	●	○	∅	○	●
Hollar, J. C., and Norwood, K. (1999)	Y	○	○	○	∅	○	●
Graham, A.T., and Thomas, M.O. J. (2000)	Y	○	○	○	∅	○	●
Szetela, W., and Super, D. (1987)	Y	○	●	○	○	○	●
Loyd, B. H. (1991)	N (Not an acceptable design.)						
Note: Y = Meets WWC evidence standards with reservations; N = Does not meet WWC evidence standards ● = Fully meets criteria; ○ = Meets minimum criteria; ∅ = Does not meet criteria							

Table 4. Causal Validity and Other Study Characteristics: Unpublished Dissertations

Study	Causal Validity	Intervention Fidelity	Outcome Measured	People, Settings & Timing	Testing within SG	Analysis	Statistical Reporting
Autin, N. P. (2001)	Y	○	●	○	∅	○	●
Drottar, J. F. (1998)	Y	○	○	○	∅	○	●
Wilkins, C. W. (1995)	Y	○	○	○	○	○	∅
Rodgers, K. V. (1995)	Y	○	●	○	∅	○	●
Glover, M.I A. (1991)	N (Not an acceptable design.)						
Ellerman, T. B. (1998)	N (Not an acceptable design.)						
Liu, S. (1993)	Y	○	●	○	●	○	●
Note: Y = Meets WWC evidence standards with reservations; N = Does not meet WWC evidence standards ● = Fully meets criteria; ○ = Meets minimum criteria; ∅ = Does not meet criteria							

Meta-analysis of Graphing Calculator Impact on Algebra Achievement

A meta-analysis gives us a way of combining the impact of multiple studies to arrive at a single estimate of the impact. Impact is expressed as an effect size, which is in standard deviation units. Specifically, we calculate this value by taking the mean of the treatment group minus the mean of the control group and dividing this difference by the pooled standard deviation.

However, a meta-analysis requires that the studies being combined be studies of the same or closely related educational problems or interventions. First, studies are selected that address similar problems based on researcher judgment. Second, a statistical test of homogeneity is used to verify that the studies have reasonably similar effect sizes.

To begin, the effect sizes for our 13 studies are summarized in Table 5 for published research papers and in Table 6 for unpublished dissertations.

Table 5. Effect Sizes in Published Research Papers

Study	Group	Sample Size	Mean	SD	Effect Size	
Ruthven, K. (1990)	T	47	57	17	1.81	
	C	33	28	16		
Thompson, D. R., and Senk, S.L. (2001)	Class 1	T	22	66.8	12.8	1.02
		C	24	53.5	12.9	
	Class 2	T	16	68.3	15.6	1.14
		C	23	51.2	14.1	
Hollar, J. C., and Norwood, K. (1999)	T	46	21.02	5.87	1.0	
	C	44	15.62	4.70		
Graham, A.T., and Thomas, M. O. J. (2000)	School A	T	21	0.476 ^a		0.52
		C	21	0.227 ^b		
	School B	T	21	0.924 ^a		0.91
		C	21	0.439 ^b		
Szetela, W., and Super, D. (1987)	TRAN2	T ^c	12	11.59		0.37
		C ^d	15	10.00		
	PROP2	T ^c	12	9.81		0.43
		C ^d	15	8.02		
Loyd, B. H. (1991)		X				
Note: T = Treatment Group, C = Comparison Group, X = Does not meet WWC evidence standards, ^a = Posttest Effect Size, ^b = Pretest Effect Size, T ^c = Problem solving strategies with calculators, C ^d = No problem-solving strategies and no calculators						

Table 6. Effect Sizes in Unpublished Dissertations

Study	Group	Sample Size	Mean	SD	Effect Size	
Autin, N. P. (2001)	T	29	81.31	11.46	0.91	
	C	29	70.79	15.12		
Drottar, J. F. (1998)	T	22	11.82	6.35	0.44	
	C	23	9.09	5.80		
Wilkins, C. W. (1995)	T	75	Adjusted Means Not Reported			
	C	24	Adjusted Means Not Reported			
Rodgers, K. V. (1995)	Paper-and-Pencil	T	17	12.29	3.70	-1.11
		C	21	15.95	3.28	
	Problem-Solving	T	17	7.58	4.08	6.79
		C	21	0.45	1.05	
Glover, M. A. (1991)	X					
Ellerman, T. B. (1998)	X					
Liu, S. (1993)	T ^e	47	28.47	7.39	0.02	
	C ^f	42	28.28	7.19		
Note: T = Treatment Group, C = Comparison Group, X = Does not meet WWC evidence standards, T ^e = Calculators plus problem solving, C ^f = Traditional and no calculators						

Since our initial focus of the review was on graphing calculators, we restricted the meta-analysis to these studies. There are four published research papers and four unpublished dissertations that investigated the effect of graphing calculators. Among these studies, the researchers measured the impact on a variety of skills and abilities, most commonly on algebra. We judged that four of the studies that met the inclusion criteria measured the effect of using graphing calculators on algebra skills. Our meta-analysis addresses these studies only. Two of the studies report two separate effect sizes which were considered independent since they involve separate classes or schools. Thus, we worked with six outcomes in the meta-analysis.

The procedures are as follows. We computed standard errors for the effect sizes. We then carried out a statistical test of homogeneity to determine whether the studies can reasonably be described as sharing a common effect size (Hedges & Olkin, 1985). Under the null hypothesis that the effect sizes are equal,

the test statistic, $Q = \sum_{i=1}^k \frac{(d_i - d_+)^2}{\hat{\sigma}(d_i)}$, (where d_+ is the estimated pooled effect size and d_i are estimated

study-specific effect sizes,) has an asymptotic chi-square distribution with $k-1=5$ degrees of freedom. In the current meta-analysis, Q has a value of 4.37. A value of Q as large as that obtained would occur between 25 and 75% of the time if the effect sizes are equal. Hence, we do not reject the hypothesis of homogeneity of effect size, and we consider pooling the data to obtain an estimate of the common effect size.

The point estimates for the effect sizes for the six results are displayed in Figure 1. Each point estimate is centered on its 95% confidence interval. The rightmost confidence interval represents the result for the pooled estimate. The 95% confidence interval does not contain zero, therefore, we reject the hypothesis that the common effect size is zero at the $\alpha=.05$ level of significance. The point estimate is .85 with a confidence interval (0.61, 1.09), which gives strong evidence that the use of graphing calculators is associated with better performance in algebra. A fixed effects model is assumed in the computation of the standard error of the pooled estimate. (Note that outcomes for quasi-experiments may be biased, and this caution should be kept in mind when interpreting results.)

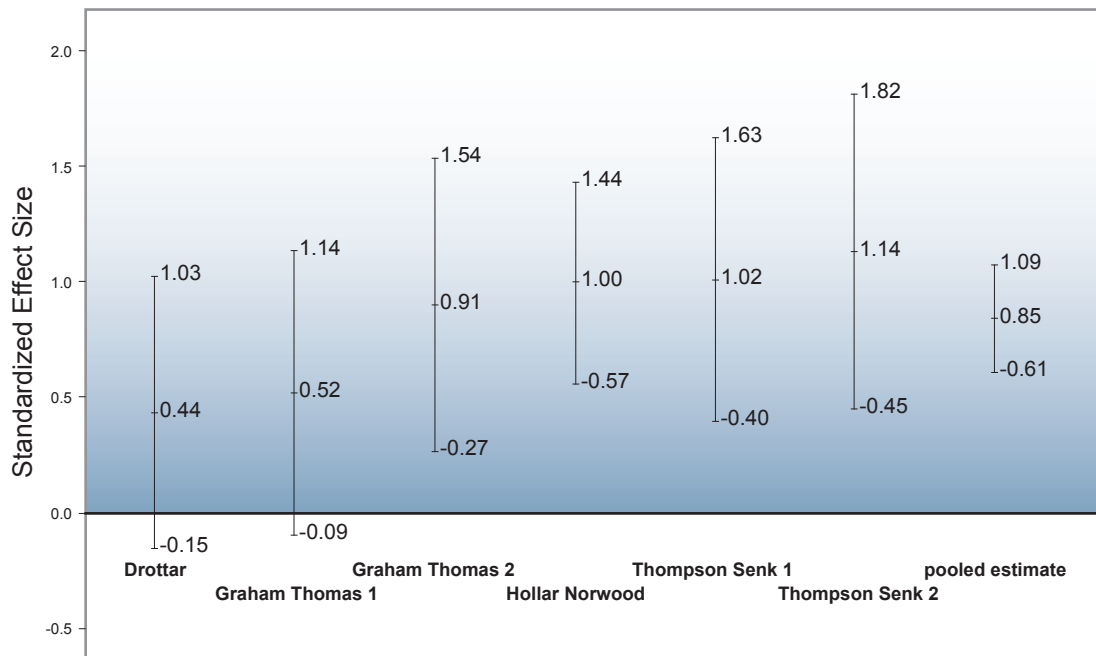


Figure 1: For studies of algebra: Estimates of the size of the difference between treatment and control groups indicating the 95% confidence interval

References

- Arlin, M. & Hills, D. (1976). Manual for Arlin-Hills attitude survey. Jacksonville, IL: Psychologists & Educators, Inc.
- Autin, N. P. (2001). The effects of graphing calculators on secondary students' understanding of the inverse trigonometric functions (Doctoral dissertation, University of New Orleans, 2001). *Dissertation Abstracts International*, 62, 0890A. (UMI No. AAT3009261..)
- Bittinger, M. L., Keedy, M. L., Ellenbogen, D. J. (1994) *Intermediate Algebra: Concepts and Applications*. Addison Wesley Longman
- Chen, T. (1994). A meta-analysis of effectiveness of computer-based instruction in mathematics (Doctoral dissertation, University of Oklahoma, 1994). *Dissertation Abstracts International*, A56/01, p.125.
- Drottat, J. F. (1998). An analysis of the effect of the graphing calculator on student performance in Algebra II (Doctoral dissertation, Boston College, 1998). *Dissertation Abstracts International*, 60, 56A.
- Ellerman, T.B. (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.
- Elliott, J. W. (1981). The effects of the use of calculators on verbal problem-solving ability of sixth grade students. *Dissertation Abstracts International*, 41(8), 364A.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- Glover, M. A. (1991). The effect of the handheld calculators 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.

- Graham, A. T., & Thomas, M. O. J. (2000). Building a versatile understanding of algebraic variables with a graphic calculator. *Educational Studies in Mathematics*, 41(3): 265-282.
- Hedges, L., & Olkin, I., (1985). *Statistical methods for meta-analysis*. New York: Academic Press Inc.
- Higgins, J. L. (1990). Calculus and common sense. *Arithmetic Teacher*, 37(7), 4-5.
- Hollar, J.C., & Norwood, K. (1999). The effects of a graphing-approach intermediate algebra curriculum on students' understanding of function. *Journal for Research in Mathematics Education*, 30(2), 220-226.
- Liu, C. M. (1989). The evaluation report of mathematical problem-solving behavior. Taipei, Republic of China: National Science Council (NSC77-0111-S026-01A).
- 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.
- Loyd, Brenda H. (1991). Mathematics test performance: The effects of item types and calculator use. *Applied Measurement in Education*, 4(1): 11-22.
- McLeod, T. M., & Armstrong, S. W. (1982). Learning disabilities in mathematics—skill deficits and remedial approaches at the intermediate and secondary level. *Learning Disabilities Quarterly*, 5, 305-311.
- Roberts, D. M. (1980). The impact of electronic calculators on educational performance. *Review of Educational Research*, 50(1), 71-98.
- Rodgers, K. V. (1995). The effect 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.
- Ruthven, K. (1990). The influence of graphic calculator use on translation from graphic to symbolic forms. *Educational Studies in Mathematics*, 21(5):431-450.
- Suydam, M.N. (1979). The use of calculators in precollege education: A state-of-the-art review. Columbus, OH: Calculator Information Center. (ERIC document Reproduction Service No. 171 573)
- Szetela, W., and Super, D. (1987). Calculators and Instruction in problem solving in grade 7. *Journal for Research in Mathematics Education*, 18(3):215-229.
- Thompson, D.R., & Senk, S.L. (2001). The effects of curriculum on achievement in second-year algebra: The example of the University of Chicago School Mathematics Project. *Journal for Research in Mathematics Education*, 32(1):58-84.
- What Works Clearinghouse (2005) *WWC Study Review Standards*. Available on the WWC site: http://whatworks.ed.gov/reviewprocess/study_standards_final.pdf
- Wilkins, C. W. (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.