The Impact of Electronic Calculators on Educational Performance

Dennis M. Roberts
The Pennsylvania State University

Thirty-four empirical studies at the elementary, secondary, and college level were reviewed concerning the impacts of electronic calculators on mathematics-related achievement and attitudes. The dominant research design was pretest-posttest with the principal analytic procedure being ANOVA. In the typical study, an experimental group (usually one or more classes) was allowed to use calculators during their mathematics instruction whereas students in the control group were not. Results showed support for the computational benefits due to calculator use, especially when the calculators were also allowed on criterion tests. However, support for conceptual benefits was minimal. Hypothesized changes in general attitudes toward mathematics for those using calculators was unsupported, although effects on immediate and task-specific, affective measures were found. Interpretation of the results of many of the studies was hampered by defective research designs including assignment of students to conditions, contamination of treatment with control groups, control of the teacher variable, and the lack of use of calculators on the posttests.

Educators have long been skeptical about the influence of innovations which either find their way into schools or catch students’ eyes outside school. Programmed instruction, with all its teaching machines, is a case in point. Educators have argued that teaching machines would merely reinforce rote memory associations rather than encourage students to solve problems creatively. What about television? Has there been an educator yet who has not cursed television, at one time or another, as a menace to school learning? After all, time spent watching television probably means time lost from an activity like homework. Most recently, small electronic calculators have come under a similar barrage of skepticism. Rather than planning careful investigations to seek answers to important questions relevant to calculators, many educators have dismissed the new technological innovation as a toy. Obviously, toys do not belong in the mathematics curriculum. Such attitudes suggest that a careful review of the empirical literature on the impacts of calculators would be helpful. The present review was completed with the above concerns in mind.

One of the major spin-offs of the space program has been the miniaturization of electronic components to a level which has allowed the small hand-held calculator industry to mushroom. In fact, current availability has reached the near-saturation point in terms of the number of units in the marketplace. While calculator use is widespread, the debate continues as to what performance and psychological benefits may accrue due to calculator utilization. Bell (1977) has indicated a number of important areas where research and development efforts are needed to better under-
stand the impacts of calculators. Most concerns about the effects of calculator usage have surfaced from educational institutions, especially parents, teachers, and principals of elementary school-age children. Opponents argue that calculators may have detrimental impacts on the development of children’s mathematical abilities. Conversely, proponents assert that calculators may actually facilitate mathematical learning. Obviously, the resolution of the debate depends on evidence from carefully planned and executed empirical investigations.

The present review examines the impacts of electronic calculators on achievement and psychological (i.e., attitudes) parameters in learning settings related to mathematics. While the literature on effects of mechanical calculating devices generally predates the research on electronic calculators (Cech, 1972; Durrance, 1965; Johnson, 1971; Ladd, 1974; Mastbaum, 1969; Shuch, 1976), the focus of the present review deals exclusively with electronic devices. However, there have been some investigations reported, but not reviewed here, where at least partial focus was placed on the electronic calculator (Majumdar, 1977; Muzeroll, 1976), but where the major purpose was not to make comparisons between groups allowed to use calculators and those not allowed to use calculators. Thus, the primary focus here is with research where the major intent was to draw conclusions about the role of electronic calculators on achievement and attitudes. Investigations reviewed span research reported from 1973 to mid-1978 and include readily available published articles (10) and completed doctoral dissertations (24). Except where noted, both the original articles and the actual dissertations were examined. The majority of the research appeared in 1976 (17 entries).

Basic Research Design

In the majority of the studies (71 percent), the basic research design utilized was the pretest-posttest arrangement. An experimental group (E) received an intermediate calculator treatment (typically embedded in the mathematics learning setting) while the control group (C) received no such calculator treatment (typically only the normal mathematics curriculum). See Table 1 for a schematic representation. It should be pointed out that some of the investigations had more than a single E group and a single C group; for example, there might have been two E groups emphasizing different amounts of calculator utilization. As a fairly typical example of the basic research design, Fischman (1976) examined high school students' attitudes and concept learning in business arithmetic where some classes used electronic calculators to complete their work, and others did not. All students were tested on the New York Computation Test and the Aiken Revised Math Attitude Scale at the beginning and end of the school year (two semesters). In the three E classes (N = 48), students were allowed to use calculators in their daily class work, whereas students in the three C

---

1 Since the review was completed, the following additional references were located in the 1978 issue of Dissertation Abstracts International: Cooper, Dyce, Eckmier, Gallery, Gooden, Hopkins, Laursen, and Wilson. One study missed in the current review is Aldridge (1977). These are not reviewed here but are listed in the references. It is important to emphasize that the present review does not take into account a variety of other empirical work such as surveys on calculator use and other less formal investigative activities. Interested persons can contact the Calculator Information Center, 1200 Chambers Road, Ohio State University, Columbus, Ohio, for bulletins describing these other works.
classes (N = 52) were not. No treatment effect was found on the attitude measure, but there was an overall positive change for both E and C groups from the beginning to the end of the school year. In terms of achievement, the E group (which was posttested twice, once when using calculators on one form of the posttest and a second time not using calculators on an alternate form of the posttest) was better than the C group when calculators were allowed on the test. When the E groups’ results were based on noncalculator use, no differences were found in comparison to C. As with the attitudinal data, there were also improvements in the performance in both E and C groups from the beginning to the end of the year.

Review by Grade Level

Studies were examined for possible calculator impacts depending on the general grade level at which they were conducted. As rough categories, elementary was considered to be through grade 6, secondary from grades 7 to 12, and finally college level. Using this scheme, 11, 13, and 10 studies were respectively classified although there was some overlap in several cases.

Elementary

All 11 of the investigations at this level were of the general pretest-posttest design. In the first study, Hohlfeld (1974) examined the effect of a calculator programmed to provide immediate feedback on working simple multiplication problems. Students were in the fifth grade. Within each of seven classes, four students (total N = 84) were assigned to one of three groups: E, used calculators as a feedback device; C1 used paper and pencil to work by hand the same problems as did the experimental group, and C2 was the group that followed the normal classroom routine without any particular attention being spent on multiplication drill. The Mathematics Computation section of the Metropolitan Achievement Tests and a researcher-developed 100-item multiplication test were given as pretests (and used as covariates), whereas alternate forms of the multiplication test were readministered as a posttest (after 1 month of treatment), as a short-term retention test (after 1 more month), and as a long-term retention test (after an additional 3 months). Using ANOVA to analyze the data on the posttest and both retention tests, it was found that E was better than
both the \( C_1 \) and \( C_2 \) groups on the posttest and the first retention test, but all groups were the same on the long-term retention test. It should be noted, however, that on the average, the E group worked nearly twice the number of problems as did the \( C_1 \) group. This additional practice may have accounted for the improved performance.

Spencer (1975) used fifth and sixth graders to observe the impact of calculators on computational skills and arithmetic reasoning abilities. The 84 students consisted of 42 males and 42 females. The arithmetic section of the Iowa Tests of Basic Skills (subtests on arithmetic concepts and problem-solving) was given both before and after the 8-week treatment. Students in the E group \( (N = 42) \) were allowed to use calculators on all in-class work plus the actual posttests. Students in the C group \( (N = 42) \) had no access to calculators. ANOVA was used to compare the gain scores between the groups; separate analyses were made for each grade. For the fifth grade, \( E > C \) on the problem-solving test whereas in the sixth grade, \( E > C \) on the arithmetic concepts (computation).

Miller (1977) examined whether calculators would be effective instructional aids in developing the concept and skill of long division with fifth-grade students. Two classes were used; one was randomly assigned to each of the E \( (N = 24) \) and C \( (N = 23) \) conditions. Pretests (for covariates) included an arithmetic readiness test, an investigator-developed division test, and the mathematics section of the Comprehensive Test of Basic Skills. The investigator-developed test (two difficulty levels) was used as a posttest measure. All students received instruction emphasizing the subtractive approach to long division with E students allowed to use calculators on the posttest. Using ANCOVA, and doing separate analysis for low- and high-ability groups, results showed \( E \) (low) > \( C \) (low) but \( E \) (high) = \( C \) (high).

In two studies utilizing the same sixth-grade students, Jones (1976) and Allen (1976) investigated the effects of using hand-held calculators on mathematics achievement, attitudes, and self-concept. Six intact classes were randomly assigned: four to the E condition \( (\text{total } N = 113) \), and two to the C condition \( (\text{total } N = 62) \). In one study, \( N \) for \( C = 58 \), while in the other study, \( N \) for \( C = 62 \). Pretests included the SRA Assessment Survey (mathematics), the Criterion Referenced Test in Metrics Measurement, and a researcher-developed test on decimals. Treatment consisted of allowing E students to use calculators during their mathematics classes to solve problems, to check work, etc. Students in the C condition had no access to calculators during class sessions. After 1 month, the Criterion Referenced Test in Metrics Measurement and the decimal test were readministered. After an additional month, the SRA was readministered along with Dutton’s Attitude Toward Arithmetic Scale and the Piers-Harris Children’s Self Concept Scale. Students in the E condition were not allowed to use calculators on the posttest. In Jones’ work, SRA gain scores along with posttest attitude and self-concept data were analyzed, whereas in Allen’s work, metric measurement and decimal test gains were examined. ANCOVA was the statistical procedure with SRA pretest scores used as the covariate. On the SRA, \( E > C \) on the posttest; however, \( E = C \) on attitudes and self-concept. On the individual metric measurement and decimal tests, \( E = C \); however, with a linear combination of both measures, \( C > E \). One problem encountered in these data was the admission by some of the C students \( (6 \text{ percent}) \) of having used calculators outside of the classrooms during the study.

With fourth- through seventh-grade summer school students, Nelson (1976) investigated impacts of calculators on computational skills and attitudes. Sixteen classes
were randomly assigned to one of four conditions: E₁ (N = 45) was a commercial program that had calculator work built in; E₂ (N = 47) was a locally developed, remedial program with calculator work built in; E₃ (N = 55) was the regular program with calculators available but not part of the regular instructional emphasis; and C (N = 49) was the regular program with no calculators available. Students were pretested and posttested on the Shaw-Hiehle Computation Test and the SMSG attitude inventory, PX 010 Scale Incentive Code, “Arithmetic Fun vs. Dull.” Students were not allowed to use calculators on the posttest. The treatment lasted 4 weeks with daily 50-minute sessions. For both computation and attitudes, E₂ was the best, followed by E₁ = E₃, and all E groups were superior to the C group. It should be noted that no means for any of the groups or any of the tests were presented in the dissertation itself.

Borden (1977), using sixth graders, examined how calculators might aid in teaching decimal concepts. Six classes were used: three were assigned to the E (N = 44) condition and three to the C (N = 41) condition. During the 4 weeks of instruction, only E students were allowed to use calculators during the unit of study. Pretest, posttest, and retention test forms of a decimal examination were administered along with 20 items taken from the NLSMA Attitude Inventory (developed by SMSG) and used as a posttest. While both E and C groups gained significantly on the decimal test, there were no differences between the two groups’ performance on either the posttest or the retention test, nor were there significant differences on the attitude inventory.

Sutherlin (1977) used fifth- and sixth-grade students to investigate the effect of calculators on decimal estimation skills. Both E (N = 84) and C (N = 88) classes received decimal estimation instruction, E students used machines to make precise calculations following an initial estimate, whereas C students used paper-and-pencil algorithms to do precise calculations. Students were pretested, posttested (3 weeks later), and given retention tests (after another 4 weeks). The test used was a decimal estimation examination developed by the investigator. Using ANOVA, significant gains in decimal estimation skills were found in both E and C; however, E = C on both the posttest and retention test.

Schun and Lang (1976) utilized 48 summer elementary school students to see if hand-held calculators would improve their computational skills. The treatment lasted 1 month and consisted of work on basic arithmetic operations. E students (N = 26) used calculators to check and work problems while C students (N = 22) used paper-and-pencil techniques. All students were pretested and posttested with alternate forms of the Individualized Computational Skills Program Computational Test 3-4. Students in the E group were not allowed to use calculators on the posttest. Using ANOVA with the gain scores, it was found that E > C on computational performance.

Kasnic (1978) studied the effect of calculator usage on mathematical problem-solving in relation to three levels of ability of the sixth-grade students tested. Four schools were used and each was randomly assigned to one of the four treatments: E₁ (N = 30) used calculators to practice problems but did not use the machines on the posttest; E₂ (N = 30) used calculators for both practicing problems and on the

---

₂ The article states that N = 60, but a personal communication showing all data points adds to N = 48.
posttest; C₁ (N = 30) used paper-and-pencil methods to practice the problems and were not allowed to use calculators on the posttest; and C₂ (N = 30) had no particular treatment. The treatment lasted nine days with 50-minute sessions each day. Posttesting involved a problem-solving measure. Using a 2-factor ANOVA with pretest ability as a blocking variable, there were no differences between E and C groups, nor were any differences found for the different ability levels between E and C.

In the final study, Whitaker (1978) explored calculator impacts with first-grade children. Classes in one school were randomly assigned to E and C conditions. E used calculators to check work and had machines available during their free study time. C students had no access to calculators at any time. The treatment lasted for 30 instructional days. Three instruments were used at both pretesting and posttesting; these were the mathematics section of the California Achievement Tests, a computational arithmetic problem test designed by the researcher, and the Mathematics Attitude Scale. Using alpha as .1, two mathematics gain scores favored the E group while one gain favored the C group. 3 No attitudinal differences were found between E and C.

In brief summary, the majority of the studies completed at the elementary level showed computational advantages (6 of the 11) from the introduction of calculator usage into the mathematics instruction, even though the use of calculators was not allowed on the posttest. However, in only one study of the five investigating concepts were there conceptual benefits due to calculator usage and in only one study of the four investigating attitudes were there attitudinal benefits.

Secondary

Of the 13 studies at the secondary level 12 were of the pretest-posttest arrangement. In the first, Quinn (1976) used honors eighth-and regular ninth-grade students to observe whether the use of a programmable calculator would facilitate algebra achievement and attitude towards mathematics. Classes in one school (which had the calculator) served in the E condition (N = 105), while the students from the other served in the C condition (N = 79). The Cooperative Mathematics Test: Algebra I and the Mathematics Attitude Inventory were given during pretest and posttest sessions. Select data from the Comprehensive Tests of Basic Skills and the Short Form Test of Academic Aptitude were used as covariates. In E classes, the treatment consisted of incorporating the use of a programmable calculator into routine instruction throughout the year; however, the calculator was used only after the students had proven that they could work the problems by hand. E students were not allowed to use the calculator on the posttest. Using ANOVA, no achievement differences were found between E and C, but E > C on attitudes.

Zep (1976) examined whether there was an interaction between the use of a calculator and ability level in ninth-grade and college students' solutions to proportion problems. Based on a pretest, students were assigned into high, medium, and low levels depending on how well they solved proportion problems. Half of each level was then assigned to the E (N = 184) condition and the other half to the C (N

---

3 The primary source was not available; therefore, it was impossible to see if results were significant at the more traditional .05/.01 alpha levels. Also, E and C sample sizes are not known.
students used calculators throughout the 2-week programmed instructional sequence on proportions. Students in E were allowed to use calculators on the posttest which was again a proportion problems test. Using a 2-factor ANOVA (separate analyses for ninth grade and college), no differences were found between E and C, although there were differences at the ninth grade due to ability level.

Gaslin (1975) compared the achievement and attitudes of ninth graders using either conventional or calculator-based algorithms for operations on positive rational numbers. Six classes were used, two from each of three schools. There were three treatments: a conventional algorithm set (CAS; N = 38) where operations were performed by the usual textbook approach; an alternative algorithm set (AAS; N = 32) where fractions were converted to decimals on a calculator first and then the various operations were performed with the decimals using the calculator; and the C condition, CAS (N = 31) with no calculator use. CAS with students being allowed to use calculators was E1. The AAS with the use of calculators was E2. Treatments lasted 10 weeks followed by a retention test after 2 weeks. Students in E1 and E2 were allowed to use calculators on posttests and retention tests. Criterion measures included an operating on fractions test, a transfer test, a fractions retention test and semantic differential attitude scale about mathematics. Both ANOVA and ANCOVA (achievement and intelligence test scores as covariates) were used. Significant treatment effects were found on both posttest achievement measures with E2 > E1 > C, and on the retention test with E2 > E1 = C; however, no differences on attitudes were found.

Eight studies at the secondary level were reported in 1976–1977, including the Fischman (1976) investigation already discussed in the Basic Research Design section. In work done with ninth-grade general mathematics students, Wajeeh (1976) examined effects of a newly developed program of meaningful and relevant mathematics on achievement and attitudes. For E1, students (N = 75) used the developed program plus calculators. E2 (N = 75) used the program but without the benefit of calculators. C1 (N = 75) was not exposed to the new program but was taught by the same teachers who taught E1 and E2. C2 (N = 75) was the same as C1; except that different teachers were utilized. The treatment lasted 15 weeks and was preceded and followed by mathematics subtests of the California Achievement Test and Dutton's Attitude Toward Arithmetic Scale. It was not clear whether the students in E1 were allowed to use calculators on the posttests. Using both ANOVA and ANCOVA, results showed superiority of both E1 and E2 over the C groups on both achievement and attitudes but little difference was found between E1 and E2.

Hutton (1977) looked at effects of mini-calculator use on achievement and attitudes of ninth-grade algebra students. Pretests and posttests were the SMSG Mathematics Inventory Form 122A (achievement) and SMSG PY-408 Pro-Math Composite Scale and PY-408 Math-Fun vs. Dull Scale (attitudes). For treatments, both E1 (N = 53) and E2 (N = 45) received calculators to use with the teachers in E1 incorporating them into the mathematics instruction, whereas the teachers in E2 did not incorporate them. In the C classes, students (N = 72) were not allowed to use calculators. Treatments lasted for 4 weeks, and the unit of study was a chapter on real number powers, roots, and radicals. Students in E1 and E2 were not allowed to use the calculators on the posttest. Using t tests, no differences were found between any of the E and C groups on any of the achievement or attitudinal variables.
Jamski (1977) investigated the impact of hand-held calculators on seventh graders' learning of decimal/percent conversion algorithms. Classes were randomly assigned to E and C conditions. For both groups, the treatment period lasted 4 weeks with E students (N = 66) being allowed to use calculators during their mathematics instruction, whereas C students (N = 70) did not use calculators. The pretest measure used to compare E and C (for equivalency) was Form 7S-3, Test D from the SMSG series, a test emphasizing computations. The criterion test was developed by the researcher and was used for both posttesting and as a retention test 5 weeks later. E students were allowed to use calculators on the posttest but not the retention test. Using ANOVA, results showed E > C on achievement on the posttest but no differences were noted on the retention test.

Using ninth-grade general mathematics classes, Vaughn (1977) examined effects of calculator usage when they were specifically integrated into the curriculum. Of the eight classes, four were randomly assigned as E and four as C. The instructional topics were decimals and percents with E students (N = 55) allowed to use calculators during the 8 weeks of instruction, while C students (N = 46) were not allowed to use calculators. The measures used in pretesting and posttesting were an investigator-prepared test on decimals/percent and the Aiken Revised Math Attitude Scale. In a retention testing that occurred 2 weeks after the posttesting session, the decimals/percent achievement test was readministered. It is not clear whether E students were allowed to use calculators on the posttest and/or retention tests. Using a stepwise regression procedure, results indicated the E > C on achievement at the posttest, but E = C on attitudes at the posttest and on the decimals/percent examination given at the retention testing.

In a study executed somewhat differently from the other pretest-posttest arrangements, Lenhard (1977) split each of a variety of mathematics classes (grades 7 through 12, total N = 125) into two groups (Group 1 and Group 2) and then let them alternate between the E and C conditions throughout the experiment. Each class took between 8 and 10 class tests during the semester. Of these tests, each student would have taken approximately half under E (use with calculator) and half under the C (no-calculator-use) conditions. Besides administering both pretests and posttests in mathematics concepts and computation, Lenhard also gave the Aiken Revised Math Attitude Scale after each of the in-class tests. In a variety of analyses using t and ANOVA procedures, the higher ability students made fewer concept and computation errors than did the lower ability students, and also had more positive attitudes. No differences were found between scores generated in the E and C conditions. However, since Group 1 and Group 2 alternated between the E and C conditions, contamination of treatment across all students appears to have been a deterrent to finding differential effects due to calculator usage.

In a study completed with seventh-grade students, Andersen (1977) was interested in the effects of restricted versus unrestricted use of calculators on mathematics achievement and attitudes. Three classes were selected at random (from seventh-grade mathematics sections) from each of four schools, one of the three being assigned to two experimental conditions and one to a control condition. In E1 (N = 106), students were only allowed to use calculators to check hand computations (restrictive use); in E2 (N = 105), students were allowed to use calculators at any time (unrestricted use); and in C (N = 114), no students had access to calculators. The
study lasted for 20 weeks and students were both pretested and posttested on achievement (computations and problem-solving) and attitudes. During posttesting, E₁ and E₂ students were allowed to use calculators on the computational tests but not the problem-solving tests. Using ANCOVA as the principal analysis procedure, E₂ = E₁ > C on attitudes.

Rudnick and Krulik (1976) investigated whether the availability (not integrated use in curriculum) of calculators would affect seventh-grade mathematics achievement and attitudes. Half of the seventh-grade classes in both of two schools was randomly assigned to the E and C conditions. After all students received instruction on the use of calculators, the E students (N = 258) were then free to use the machines any time at their own initiative. No special changes in the mathematics program were made to accommodate the calculators. Students in the C condition (N = 209) were not allowed to use calculators. Students were pretested with the Cooperative Mathematics Test and an attitude measure at the beginning of the school year, retested with the achievement test in January, and then retested again with both the achievement and attitude measures at the end of the year. Participants in the E conditions were not allowed to use calculators on the first retesting. However, two forms of the achievement test were administered at the second retesting, at which time students in the E condition were allowed to use the machines on one of the tests. Data reported only cover the pretesting and the first retesting. Using ANCOVA, there were no achievement differences found between E and C on the retest. It should be noted that there were significant differences favoring the C group on the pretest of achievement.

In a study involving a nonmathematics content area, Bolesky (1977) investigated the influence of calculators on achievement in chemistry. Eighty 11th-grade college preparatory students were assigned (40 each) to E and C classes. The E students were allowed to use calculators to review homework, to assist in computations for laboratory exercises, to do computations for chemistry problems, and to assist them on in-class quizzes. Students in the C condition were not allowed to use calculators. The treatment period lasted for one semester. At the conclusion of the study, the Anderson-Fisk Chemistry Test and a teacher-constructed chemistry test were given in the following 2 × 2 factorial design: E condition, calculators on the posttest; E condition, no calculators on the posttest; C condition, calculators on the posttest; and C condition, no calculators on the posttest. Results showed no significant main effects or interaction.

And finally, Boling (1977) investigated whether the use of calculators would differentially improve achievement and/or attitudes for students in consumer mathematics classes. Students in grade 12 were used. Pretests and posttests were comprised of the Mathematics Problems section of the California Achievement Tests and a modified version of The A-V Scale of Attitude Toward Mathematics. In the E condition, students (N = 51) were allowed to use calculators to complete their daily class calculations, while students (N = 43) in the C condition were required to do all their calculations by paper-and-pencil methods. The treatment lasted 19 weeks. Using both ANOVA and ANCOVA, no differences between E and C were found for either achievement or attitudes.

Again, in brief summary, a majority of the secondary-level studies (6 of the 11 computation studies) found computational benefits due to calculator use. However,
as was the case in the elementary studies, very little support was found for the hypothesis that calculator benefits transfer to the more conceptual (1 of the 8 concept studies) and affective areas (2 of the 9 attitude studies).

**College**

The 10 investigations completed at the college level reflect a wide range of potential calculator impacts, and thus are represented by more diverse research designs. In fact, only one investigation used the pretest-posttest arrangement, the design almost exclusively used in the elementary and secondary studies. Sosbee and Walsh (1975) were interested in whether students who had used calculators on in-class chemistry examinations would do better than those not using calculators. During the final examination, students (N = 169) were asked to self-report whether or not they had used calculators on the other examinations. The mean examination score for those reporting that they had used calculators was compared to the mean score for those who reported that they had not. N's for the separate groups were not given. Using t tests, results showed that those reporting calculator use scored better on the examination. To counter the possibility that ability level could explain the results (smarter students more often bought calculators), a comparison of scores on the American College Testing Program admission tests was made between calculator users and noncalculator users. No differences in ACT scores were found.

In the only college level pretest-posttest study, Nichols (1976) examined the influence of calculator use on achievement and attitudes of students in a basic mathematics course. Four classes were assigned to either E or C conditions. E (N = 48) students were allowed to use calculators for daily work and on in-class examinations; C (N = 50) students had no access to calculators. Scores from the ACT tests were available along with a locally constructed mathematics survey instrument as pretests, as was the Semantic Differential Mathematics; the mathematics survey test and the attitude survey scale were again given during posttesting. It was not clear whether students in the E classes were allowed to use calculators on the posttest. Using ANCOVA, no differences were found between E and C on either achievement or attitudes.

Ayers (1977) was interested in the effects of situational problem-solving and calculators on statistics performance. Participants were 172 statistics students randomly (and equally) assigned to one of four conditions: S-C, situational-realistic teaching heuristics coupled with classroom access to calculators; S-NC, situational-realistic teaching heuristics coupled with no classroom access to calculators; NS-C, nonsituational-realistic teaching heuristics coupled with classroom access to calculators; and finally, NS-NC, nonsituational-realistic teaching heuristics coupled with no classroom access to calculators. The three posttest criterion measures used were a statistics content assessment device (75 items), the McCallon and Brown Attitude Toward Mathematics Scale, and a view of statistics inventory (30 items). Again, it was not clear whether students in the S-C or NS-C conditions were allowed to use calculators on the posttest. Using several analytical procedures (ANOVA, Mann-Whitney U test, etc.), results showed better achievement in the calculator groups and more positive attitudes in the situational-realistic teaching heuristics groups. No attitudinal differences were found between the groups using or not using calculators.
O’Loughlin (1976) investigated the effects of using a programmable electronic calculator on the achievement of students in two sections (total N ≈ 30) of a one-semester calculus course. (N’s for separate E and C groups were not given.) In the E section, students used the programmable machine as an aid to instruction and to work laboratory exercises, whereas students in the C class did not have access to the calculators. The criterion measure was a calculus test designed to assess competency in six calculus areas: limits of functions, definite integrals, etc. Students in the E group were not allowed to use the calculators on the posttest. Using a 2-factor ANOVA (blocking over three ability levels based on previous mathematics achievement), differences on three of the six calculus subtests were found in favor of the E group; no differences favored the C group.

Smith (1977) examined achievement and attitudinal impacts of calculator usage in teaching the simplex method to business and economics majors. The summer school students were randomly assigned to two sections of a beginning business and economics course. The instructional period lasted for 10 days in which the simplex method was taught; E students (N = 34) were allowed to use calculators during their class periods, whereas C students (N = 35) were not. A simplex test of 12 items and the Aiken Revised Math Attitude Scale were given as criterion measures. E students were not allowed to use calculators on the posttest. Using previously available ACT mathematics scores as a blocking variable, no achievement or attitudinal differences were found between E or C groups. (One problem in this study was that all students were informed ahead of time that their performance would have no bearing on their course grades.)

In the only series of investigations reported, Roberts and colleagues (Roberts & Fabrey, 1978; Roberts & Glynn, 1978a, 1978b; Roberts & Lerner, 1978; and Roberts, Seaman, & Lerner, 1978) focused directly on computational benefits of using calculators. For all five studies, introductory statistics students worked numerous statistical problems (means, standard deviations, correlation coefficients, etc.) under a variety of conditions. Three criteria that were common to all studies included the number of correct answers, the time to work the problems, and efficiency (number correct per unit of time). In four of the five studies, a posttest attitudinal measure was also administered. Each study is briefly discussed below.

In a pilot investigation, Roberts and Glynn (1978a) employed three types of calculation methods: hand, 4-function calculator, and 4-function calculator plus memory/square root. Each student (total N = 48) was assigned at random (and equally) to one of the three calculation modes for solving a set of five routine statistical problems. Students worked in small groups; however, there was no confounding of the experimental conditions. Participants recorded their beginning and ending times. Criteria were number of problems correct, time to work the problems, and efficiency. Using a priori contrasts, the results showed a combined machine condition to be superior to the hand condition on all dependent measures. In no case, however, were the two machines different from each other.

In a follow-up investigation, Roberts and Glynn (1978b) used the same calculation modes but added task difficulty as a second factor. Separate sets of easy and hard problems were constructed: easy problems primarily involved single-digit numbers with no decimals, while hard problems involved three-digit numbers with decimals. Students (total N = 60) were randomly (and equally) assigned to one of the three
calculation modes but all students worked both an easy and a hard set of problems. Order of problem presentation was counterbalanced. In addition to the three major dependent measures, a posttest Likert-type attitudinal measure was administered to assess students’ feelings about themselves and the problems they had just completed. The scale had five clusters of three items each measuring attitudinal subcomponents called activity, affect, capability, facility, and value. Using a 2-factor ANOVA, the results showed superior machine performance (combined) on number correct, time, and efficiency. For task difficulty, no effects were found on number correct, but easy problems were worked faster and more efficiently than hard problems. Also, for number correct and time, there was an interaction between calculation modes and task difficulty with larger differences in performance being found between hard and easy problems when worked by hand. For the posttest attitudinal measure, the combined machine groups gave more positive ratings on each of the five clusters than did the hand group. In a later study (total N = 102), which was basically a replication of the Roberts and Glynn (1978b) study, Roberts and Lerner (1978) found similar results for both the performance and attitudinal dependent variables. However, greater distinctions were made between machine conditions: the simpler machine had four functions plus memory/square root, and the advanced machine had scientific and automatic statistical functions. In this case, differences were found not only between each machine condition and the hand condition, but also between the two machine conditions themselves.

Roberts and Fabyan (1978) extended the previous work by adding two new independent variables and clarifying the calculation mode factor. The new factors were amount of prepractice (two trials vs. five trials) and instructional work set (instructions to work fast or work accurately). The modification to the calculation mode was that a larger gap between the capabilities of the two machines was introduced: currently, the simple calculation condition had a machine with four functions plus memory/square root features, whereas the advanced machine had a variety of scientific functions plus numerous automatic statistical features (as was in Roberts & Lerner, 1978). There was no hand group in the present study. The design was a $2 \times 2 \times 2$ factorial ANOVA with two levels of prepractice, two calculation modes, and two instructional work sets. The 60 students were randomly (and equally) assigned to one of the four cells made by crossing the first and second factors. All participants worked a separate set of problems under each of the instructional work sets in a counterbalanced order. Results showed that amount of prepractice had no effect on the number correct, time, or efficiency dependent measures. For the calculation mode, the advanced machine was consistently superior to the basic machine. For the instructional work set, the work-accurately condition produced more correct answers, longer working times, and less efficient problem solution. For the attitudinal data, the advanced machine condition produced more positive ratings on four of the five attitudinal clusters.

And finally, Roberts, Seaman, and Lerner (1978) were interested in estimating the discrepancy in performance and attitudes if one pitted the least sophisticated calculation condition (hand work) against the best calculation condition (advanced calculator). In addition to calculation mode, task difficulty and instructional work set were other factors manipulated. The 64 students were randomly (and equally) split into hand and machine groups, and then all participants worked two sets of
problems under one of the following four combinations: easy first under work fast set, followed by hard second and work accurately; easy first under work accurately followed by hard second and work fast; hard first under work fast followed by easy second and work accurately; or hard first under work accurately followed by easy second and work fast. Using ANOVA, results showed large differences between calculation modes in favor of the advanced calculator (double number correct took one-sixth the time, and 15 times more efficient). As to task difficulty, significant effects were also found with easy problems being worked more accurately, in shorter times and with greater efficiency. Contrary to the earlier work of Roberts and Fabrey (1978), no effects for instructional work set were found, although the trends were the same. There also were interactions between calculation mode and task difficulty for the three performance dependent measures with the largest differences occurring between easy and hard problems done by hand. The attitudinal data showed large and significant differences between calculation modes with more positive task and self-perceptions being expressed by students using the advanced calculator.

In summary, a majority of the studies (seven of the eight that studied computations) provided support for the computational benefits of the calculator. In addition, affective (attitudinal) effects were found (four of the seven studied attitudes) favoring groups which had used calculators. This was especially true when the dependent measure tapped task-specific, affective responses. There was some support (two of the three that studied concepts) for the conceptual impact of calculator use.

Review by Effects

Computational

Computational benefits resulted when students who had used calculators during a treatment could perform routine computations (not solutions to word problems, etc.) more accurately and/or rapidly than those not having access to calculators during the treatment. Such benefits might occur when either the students were allowed to use a calculator on the actual task (posttest) or not, or both. Although it was the present author’s judgment as to what criterion tests were more computational than conceptual, such a classification still appeared to be instructive. Of the 30 studies examining computational skills, 19 (63 percent) reported positive findings in favor of the E groups. In no study was there an overall performance difference where the C group was better than the E group. Looking more closely at the 19 studies, the results show that 11 of the 19 positive findings (58 percent) were in cases where students were allowed to use calculators on the criterion test, whereas only 6 of the 19 (32 percent) had positive findings when calculators were not allowed during posttesting. In two of the cases it was not clear whether students in the E group could or could not use calculators on the posttest. Thus, it appears that the data support the hypothesis that using calculators during instruction benefits routine calculations and that the benefit is most pronounced when students continue to use the machines while actually performing the test computations. While there were many design problems in the studies, we would expect such problems to work towards washing out E and C comparisons. The fact that a majority of studies still found E > C would again suggest real computational benefits due to calculator usage. The most compelling evidence supporting this hypothesis comes from the series of investigations by Roberts and colleagues in which three criterion performance measures were used:
number correct, time to work problems, and efficiency. Systematically increasing the sophistication level of the calculation mode produced large increments in performance. Differences in performance due to increasing levels of calculation mode were most dramatically seen in the efficiency dependent measure. In some cases, the advanced calculator group performed 15 times more efficiently than the handwork calculation mode.

**Conceptual**

One could argue that for certain types of mathematics problems, a calculator might facilitate concept-formation abilities. Thus, if some mathematical principles require numerous, laborious calculations in order to be well understood, then concepts should be acquired faster if calculators can aid the student to leap through the computations. In addition, and perhaps more importantly, if calculators can reduce frustrating computational errors, then the quality of the concept attainment may be improved. Although the proposition that calculator usage can have an impact on mathematical concept formation seems reasonable, it is not supported thus far by the empirical data available. In fact, a strong case can be made that this hypothesis has not been adequately tested since few studies made any real attempt to carefully integrate calculator use into the curriculum that would illustrate how calculators can facilitate concept learning. Only 4 of the 16 studies (25 percent examining concepts) showed superiority of the E group over the C group on tests that could be considered to emphasize concepts rather than sheer computations. Obviously, positive results related to conceptual benefits of calculator usage would not be expected to occur as often as simple computational benefits because conceptual acquisition is a more complex task. Thus, attempts to demonstrate conceptual advantages require more intense efforts to fully integrate calculator usage into mathematics instruction. For example, to simply show students how to operate the calculator when used in mathematics problem-solving situations would appear not to be sufficient. If the desire is to have the calculator facilitate concept acquisition, then the precise ways in which the calculator can be utilized to more efficiently and/or effectively solve the problem must be demonstrated. To illustrate this difficulty, Rudnick and Krulik (1976) stated, “After a three-day learning period with the calculator, these students were ‘on their own’ as to how and when to use the machine. No modifications were made in the mathematics curriculum” (p. 655, italics theirs). With a research strategy like this, it is no wonder that very few conceptual benefits have been found.4

On the other hand, it might alternatively be suggested that a reason conceptual benefits were not noted more often was that the learning settings in which these studies were conducted did not generally emphasize concept-formation skills, thereby providing a partial explanation. Of the four positive findings, two were obtained when students used calculators on the posttest and two were obtained when calculators were not allowed. Not surprisingly, if the calculator were successful in

---

4 One additional difficulty that was constantly emphasized at a recent calculator research conference sponsored by the National Institute of Education (January 22-24, 1979) was the failure of most researchers to carefully build the E treatment in such a way that it took advantage of the unique capabilities of calculators. In most studies, there was little attention given to the integration of calculators into the treatment curriculum.
facilitating learning the concept and the concept were really at the understanding level, then utilization of the calculator to apply the principle should have had relatively little bearing on obtaining the correct solution on a posttest. However, it would have been interesting if time data had been collected on the concept-oriented tests because, while number correct might not have been affected, the time to complete the test might have been decreased. Until investigators use calculators as a strategy for solving problems, the concept-formation benefits of calculators will still be an unresolved issue.

Attitudinal

The results on attitudinal criteria are also disappointing even though the data do seem to be more interpretable. Of the 20 studies (35 percent examining attitudes) reported results where E groups gave more positive ratings than did C groups. Four of these 7 were in the Roberts series that emphasized students’ immediate reactions concerning their feelings about themselves and the problems they had just completed. There the subcomponents on the attitude scale were more congruent with the context of the experimental situation, thus rendering it more sensitive to differences between E and C students. For the vast majority of the studies, three factors may have worked against finding attitudinal impacts. First, the measures used were too oriented toward general traits rather than characteristics which were more task specific and sensitive to calculator influences. Second, the time frame for most studies was too short to expect any real change in attitudes. Third, since many of the investigations did not allow E students to use the calculators on the posttest, many of the students might have been disgruntled which, in turn, might have influenced the rating on the attitude scales. In summary, there seems to be evidence that calculators influence immediate and specific attitudinal perceptions, but evidence supporting more general and lasting changes is not available.

Summary of Pertinent Data

Table II–IV present summaries of pertinent objective and subjective data related to the 34 investigations reviewed. A brief explanation is given here as to the descriptors used in each table. The studies are listed from elementary school (Table II) and secondary school (Table III) through college (Table IV). Content refers to either traditional mathematics (M) settings or others (O) (statistics, chemistry, etc.). The design listing includes the pretest-posttest (PP) arrangement (most popular) and other (O) arrangements (posttest only, etc.). Treatment length is self-explanatory. The retention test columns refer to whether a second (or third) posttesting occurred (Yes, Y, or No, N) after the treatment was completed. The time between the first posttest and retention testing varied considerably. Calculator use describes whether students in the E group were allowed to use calculators (Yes, Y, or No, N) on the posttests. In cases where there are check marks for both Y and N, either the same students took two tests, one using calculators and another not using calculators, or different groups of students completed one test under each condition. (In four cases it was not clear from the original manuscripts whether a Y or N was the appropriate classification.) The student assignment classification is somewhat complicated: S means professed (by investigator) random assignment of students; C means by classroom; O means other. In many of the investigations it was difficult to ascertain
how students were placed in the E and C conditions. Results are classified into computational (Cm), conceptual (Cn), and Attitudinal (A) benefits. It must be emphasized that the line between Cm and Cn was somewhat arbitrary and, in many cases, fuzzy at best. However, the essential difference was that if the posttest involved routine calculations as the primary task, it was considered to be Cm; if it emphasized interpretive exercises like word problems, it was considered to be Cn. The dot indicates that the variable was measured in a given study. The analysis breakdown concerns whether the investigator used simple t tests, analysis of variance (An), analysis of covariance (Ac), or other procedures (regression, etc.). The internal validity columns refer to four factors considered: S for assignment of students to conditions (this automatically received a minus (−) if it were not random), I represented the instrumentation used (primarily tests), C symbolized whether contamination between the E and C groups could have occurred easily, and T refers to control of the teacher variable. The I decision was basically one of whether the tests used could have been (on initial examination by this reviewer) sensitive to possible E and C differences. For the C decision, if E and C students were in the same school, especially if taught by the same teacher, it was given a minus (−) because the E and C students could converse about the experiment. The T decision had to do with whether the teacher variable was reasonably controlled. If multiple teachers were assigned (essentially at random) to the different E and C conditions then it was rated as a plus (+) but if only one teacher taught both E and C, or if one teacher taught E and another taught C, then it received a minus (−) rating. Finally, the comments columns simply represent various images which stood out in reviewing the studies. The absence of comments from certain investigations does not necessarily mean that they were free of problems.

Research Difficulties

While there are numerous internal and external validity problems associated with many of the studies reviewed, four will be isolated for discussion because of their particular relevance: assignment of students to groups, contamination of treatment with control group, control of the teacher variable, and use of calculators on tests.

Assignment of Students to Groups

The most serious design strategy problem associated with calculator research has to do with how the students were assigned to the E or C conditions. As can be seen from Table II-IV, only 12 of the 34 studies (35 percent) reported having assigned students at random. The typical situation was to assign classes at random. Hence, since the class was the unit of assignment, statistical analyses should have been based on using the number of classes as N, not the number of students. In no case was it reported that the analysis was done that way. To counter the possible pretest differences between E and C groups, ANCOVA was then used to control for those possible differences in the posttest analysis. Since ANCOVA also assumes random assignment and does not overcome that flaw, E and C groups could have been different in many ways which were not included as the covariates and such potential differences could have either produced effects or prohibited demonstration of effects. (See Rabinowitz & Roberts [1977] for additional discussion of the foregoing point.) The continued incorrect use of ANCOVA to rectify design inadequacies seems an
anachronism today in light of knowledge about statistical analysis. If the assumptions for both ANOVA and regression are met, then ANCOVA may (but not necessarily) reduce the error term if there is a reasonable correlation between the covariate and the criterion measure. However, ANCOVA will not make nonrandom groups random! In a mild defense of random assignment, it should be pointed out that 67 percent of the studies using random assignment found some positive results favoring the E groups, whereas only 59 percent found some positive results in the nonrandom assignment studies.

Contamination of Treatment with Control Groups

Theoretically, during the treatment period, only the students in the E groups would have access to calculators. While this may have been true during the actual mathematics class periods, it may not have been true in either the students' other classes or outside the school setting. The lack of control of calculators used by C students may have been influenced by two factors: (1) the general availability of calculators in the home or at other locations; and (2) the extent to which E and C students communicated about the experiment in progress. While the first factor is essentially uncontrollable, the second factor is controllable. Randomly assigning E and C conditions to multiple schools and/or multiple grade levels in the same school would have been a reasonable way to exert some control over possible contamination. Unfortunately, most of the studies reported generally had both E and C conditions in the same schools and at the same grade level. Thus, the contamination of the treatment by the control groups was very likely to have occurred. The suggestion that this could have happened often may help to explain the pattern of results found. If one examined E and C differences on computationally oriented tests, especially if calculators were allowed on the tests, then contamination would probably have had little effect. Since the calculator aids one while working the problems, the fact that C students may have used a calculator illegally during the treatment period would not change the fact that they would not have access to it on the posttest. However, on the conceptually oriented tests and the attitudinal measures, intervening illegal calculator use might have improved either their understanding of the mathematical principles and/or their attitudes about mathematics. Thus, on the criterion tests, even when calculators were not available to C students, the residue of impact might have confounded the results. While the second type of contamination could have been controlled in most of the previous studies, it is doubtful if it would be very controllable in the future because of widespread availability of calculators in the population. Unfortunately, for aspiring researchers, the growth of calculators has not occurred slowly but rather has occurred in one large onslaught. While it still may be possible to show computational benefits in highly controlled studies, such investigations are not likely to have much theoretical appeal since the data already support positive effects in this area. Therefore, the areas in which there still remains doubt (conceptual and long-term attitudinal) are going to be more difficult to investigate with sufficient control over the possibility of "contamination of the second kind." It will take creative research efforts and designs to overcome this problem.

Control of the Teacher Variable

Another difficulty with many of the studies had to do with who administered the
treatments. In some cases, the researcher gave both E and C conditions; in others, one teacher took E and a second teacher took C. In either case, inconsistencies in how the E and C routines were implemented throughout the treatment period could have either heightened or masked calculator impacts. For example, if one teacher taught both E and C and it happened that the teacher were not very enthusiastic about calculator use in the classroom, then this could have possibly entered into the daily instruction and lessened the true effects. On the other hand, the opposite might have occurred if the teacher administering the E treatment were very devoted to calculator use, thus giving extra but inappropriate help to the E students. Seventeen of the studies (50 percent) appeared to have adequate control over the teacher variable. Of the studies having satisfactory control, 71 percent found positive results favoring the E groups while only 53 percent of the studies found positive results where the teacher variable was suspect.

Use of Calculators on Criterion Tests

One of the most peculiar aspects of the calculator research (at least to this reviewer) is the extent to which investigations (47 percent) did not allow students to use calculators on the posttests. The assumption is that this reflects the philosophy that the real question is whether the use of calculators will harm students' performance when they are faced with doing calculations by paper-and-pencil methods. While such an approach is certainly a legitimate way to test theory, it appears to be a negative orientation rather than a positive one. Instead of examining potential positive impacts, the focus is on demonstrating the lack of negative effects. However, it seems more realistic to assume that calculators may have more positive benefits than negative. If that is the working assumption, then allowing one to use the calculator on the actual test seems the more useful research strategy. As was noted earlier, computational benefits were more often found in cases where students were allowed to use the calculators on the tests. It would have been instructive to test for more conceptual benefits by allowing calculator use on those types of criterion measures. The current reviewer's opinion is that if one wants to maximize performance, then calculator use on a criterion measure is a more crucial test of the research hypothesis. If one uses calculators on the posttests and there still is no E and C difference, assuming, of course, that the treatment properly integrated calculators, then one is in a very strong position to argue that the calculator has little effect. The reverse procedure, however, does not necessarily provide that position of strength.

Conclusions and Implications

There seems to be little doubt about the computational value associated with calculator use. Sufficient pretraining enables one to work problems more accurately, rapidly, and efficiently. Another definite advantage is that calculators allow one to complete more problems per unit of time, thus in effect affording greater amounts of practice. However, when the discussion shifts to higher level conceptual and attitudinal impacts due to calculator use, there is less consensus as to what facts can be gleaned from the research literature. Also, at a philosophical level, there is less agreement as to what should be done regarding the question of whether to incorporate calculators into instruction. For those who believe that calculators should be used in the schools, the question may still remain as to when such aids would be introduced.
into the mathematics program. Unfortunately, the research literature offers no guidance on this point. No one has described either a reasonable cross-sectional or longitudinal investigation as to the critical timings for entering calculators into the mainstream of quantitative instruction, or at what different times, depending on the mathematical principles being emphasized. This is one area where future research could have an important influence on instructional strategies.

Unfortunately, the overall future in basic calculator research may be without much promise because of the tremendous boom in calculator availability. Most thematic research areas proceed through stages of rough and troubled waters before implementing tight and appropriate strategies. Usually, such phases extend over many years with only small increments in progress from year to year. However, calculator research happens to be an area where the ability to control the “treatment” of calculator use only spans 3 or 4 years. The earliest study appeared in 1973 with 1976 as the banner year. Because most of the research in this area has been through the medium of doctoral dissertations (only a few of which have been published), the dissemination of findings has been very slow with a parallel difficulty in providing evaluation and feedback by the profession regarding the methodological weaknesses. It is doubtful that this problem can be resolved easily in the future. Unfortunately, this may mean that the content areas in which work still needs to be done (concepts and long-term attitudes) will be neglected because adequate design controls are no longer available. However, there may be some promise if a shift is made from examining the number of correct answers toward the use of measures such as time or efficiency. The work of Roberts and colleagues implies that such measures may be more sensitive to the impacts of calculators on educational goals. However, even more crucial will be the necessity to make a real effort to develop treatments that utilize unique capabilities inherent to calculators. So far, most studies have not adequately integrated calculator use into the instructional process. Future investigators will have their hands full in order to keep calculator research alive—and healthy.

References