Alternative Designs for Evaluating Computer-Based Instruction

F J King
Professor, Florida State University
and
M. D. Roblyer
Associate Professor, Florida A & M University

Abstract. As more computer-based materials and methods appear in the educational marketplace, there is an increasing need for studies to validate the effectiveness of these products of technology and to provide more substantial guidelines for future development. However, efforts to study the usefulness of computer-based methods and materials with students are rare in comparison with the opportunities for such studies. Although many reasons have been hypothesized for why more research on the effects of various computer-based methods and materials has not been accomplished, two which seem most likely are: (1) practical difficulties in arranging studies with traditional randomized two-group designs in non-laboratory settings, and (2) lack of available expertise to develop and employ alternative designs and analyze resulting data. This paper:

(1) Presents five designs (2 two-group and 3 one-group) which can be used effectively when a randomized two-group design is not practical
(2) Gives examples of evaluation and research studies which effectively employ each of the designs
(3) Provides references to procedures to analyze data resulting from these non-traditional methods; and
(4) Provides a Decision Flowchart for selecting the most appropriate design for the evaluation question and constraints of the situation.

The widespread use of microcomputers in schools, which began in the first part of this decade, signaled a new era for educational computing. Before this movement, most studies of computer-based instruction were limited to large-scale centralized projects such as the Stanford math and reading (Fletcher & Atkinson, 1971; Suppes & Morningstar, 1971), Computer Curriculum Corporation (Lysiak, Wallance, & Evans, 1976; Crandall, 1977; Holland, 1980), Control Data PLATO (Alessi, Siegel, Silver & Baines, 1982-83; Poore, Qualls, & Brown, 1981), and TICCIT (Jones, 1978) programs. These studies were usually performed by grant-funded projects with a research and evaluation component built into them. With the acceptance of low-cost computing in schools and the explosion of new computer-based materials, the opportunities for evaluating the effectiveness of instruction in this area have greatly expanded and changed in nature. Small school and classroom-based programs have been started throughout the country. Not only do studies of these programs have the potential for making a significant contribution to the field, they are, in fact, essential if we are to make any progress in establishing the usefulness of new computer-based methods and materials.

Since the focus of school-based educational computing projects is often implementation rather than research, at least two factors are likely to inhibit systematic study of their use. First, there are practical difficulties in arranging studies with the traditional, randomized two-group designs in school settings. Second, project personnel often lack the expertise to employ alternative designs and to analyze the resulting data. If we are to take advantage of the valuable information available from classroom computer projects, alternative designs to study and evaluate the effectiveness of computerbased activities must be employed. The authors:

- Present five designs (three one-group and two two-group) which can be effectively employed to study computer-based methods when a randomized two-group experimental design is not practical.
- give examples of evaluation studies which effectively employ each design,
- Provide references to procedures which can be used by persons with limited training in statistics and evaluation methods,
- Give a decision flowchart and describe how to select the most appropriate alternative design for a given project.

Design #1: Sequential Analysis

Sequential analysis originated almost forty years ago (Wald, 1947), but, with few exceptions, it has been used for quality control of manufactured products rather than to evaluate the quality of educational materials. It can be used to construct an effective design for evaluating instructional materials. This technique could be especially useful to school administrators responsible for developing or purchasing computer-based programs for district-wide distribution and use. Before they make the considerable expenditure of time, money, and resources required to implement computer-based programs, administrators should have some data to indicate their instructional effectiveness. They may find it logistically difficult to arrange for a control group and to have an appropriate number of students, copies of materials, and computers for such a study.

An administrator would find that sequential analysis has many advantages for evaluation of computer-based materials besides not requiring a control group. Its principal advantage is efficiency. Decisions can be made with as much as 50% fewer observations than would be required for procedures in which the sample size must be specified in advance. This makes it an ideal choice when the number of people who work with the materials is limited by factors such as the number of computers available, by not being able to obtain a large student group at one time, or by the expense of implementing the program.

In such studies, the following strategy would be employed: (a) an observation would be made (e.g., one or more students are tested on some skills), (b) the results are recorded on a chart or graph developed for this purpose, and (c) by reading the chart, a decision is made to reject the instruction, to accept the instruction, or to make another observation.

There are some limitations on when sequential analysis can be effectively used. Since there is no control in this design for prior learning, it is better used when there is little likelihood that students would be able to perform the skills involved *before* exposure to the materials. It is also desirable that the skill units be relatively short, to reduce effects due to maturation and history (Campbell & Stanley, 1963). If the materials under study are lengthy, they can be divided into short units, and a sequential analysis done on each one. This technique is also most useful when teachers want to make absolute judgments about materials ("Are these computer-based units effective?") rather than a comparative one ("Are these lessons better than other non-computer lessons?").

Procedures to Implement Sequential Analysis

The following procedures are required. For detailed explanations of the procedures, see Epstein (1975), Epstein and Knerr (1978), Wald (1947), and Burr (1980). Lathrop (1983) also gives a detailed description of statistical procedures required for sequential analysis, although he used it for evaluating student performance rather than instructional materials.

STEP 1: Set acceptable/unacceptable quality level. Instructional effectiveness is defined in terms of the preparation of students who successfully complete the instruction. Since absolute accuracy in determining this proportion cannot be achieved with a sampling technique, it must be translated into a range of proportions which define acceptable and unacceptable instructional quality.

STEP 2: Set risk parameters. The probability of rejecting high quality instruction (Alpha) and the probability of accepting lowquality instruction (Beta) must be set.

STEP 3: Construct Operating Characteristics Curve (OC), Average Sample Size Curve (ASN), and truncation number. An OC curve is constructed which will show the probability of rejecting or accepting the program for any true value of the success proportion under consideration (given the specifications of the test). Then the ASN, or the average number of student scores needed to reach the decision for any true value of the success proportion, is calculated. Finally, the truncation number is calculated, which is the largest number of students that could be needed to make a decision without altering the risks the experimenter is willing to accept.

STEP 4: Construct graph for recording observations. The graph for recording observations contains three zones corresponding to the three possible decisions which are considered each time an observation is made (to accept, reject, or take more samples). The actual decision choice is determined by the pattern of recorded data. Students are tested, one by one, until the decision is made to accept the instruction, reject the instruction, or until the truncation number is reached. If the truncation number is reached, the instruction would be accepted if the last point was nearer the line indicating acceptable instruction, or it would be rejected if the opposite prevailed.

Example of Sequential Analysis Use

A study using the sequential analysis method is currently being undertaken by the U.S. Army to validate the effectiveness of microcomputer/videodisc training materials (King, n.d.). The modules were designed to teach specialist students to perform a lengthy and expensive procedure for testing the purity of jet fuel. No other training is currently being given, so no control condition is available. A sampling procedure such as sequential analysis is an appropriate design for this evaluation situation since the length of the instruction and the limited number of videodisc units available would make it difficult to test a large group of students at once. Effectiveness is to be defined in terms of the proportion of students who successfully complete the simulated task in no more than two trials. In this case, the proportion is set at .80.

Design #2: Value-Added Analysis

Many computer-based instructional materials are large-scale programs which deal with such areas as reading and writing skills and motor skills development. A great deal has also been written about the impact of LOGO on problem-solving skills. Although the expense of implementing these programs can be considerable, and, in the case of LOGO-related activities, may call for revolutionizing classroom methods, few studies have been done to measure the impact of these materials. School personnel wishing to gather evidence of effectiveness before investing in such innovation will find that study of these materials must often extend over a long period of time in order for measurable effects to occur. The value-added analysis is appropriate for evaluation of these long-term activities, since it controls for natural maturation processes.

The basic idea underlying value-added analysis is that the effect of instruction can be estimated by comparing the average observed growth between pretest and posttest with the estimated growth that is expected to occur in the absence of instruction (Bryk & Weisberg, 1976). An unbiased estimate of growth rate can be provided by way of ordinary least square regression of pretest on the student's age, if age at the time of pretest is not related to systematic growth. When growth (estimated from a cross-section of the population) is not linear, transformations must be made before value-added analysis can be completed.

Certain conditions must be present in order to use value-added analysis. An independent variable which can be measured without error (i.e., chronological age) must be available, and it must be correlated with pretest scores. A large enough number of students must be available to be able to compute a product-moment correlation. In addition, it must be assumed that longitudinal growth can be estimated through study of a cross-sectional sample.

Procedures to Implement Value-Added Analysis

Value-added analysis involves the following steps. For formulas and detailed descriptions of statistical procedures, see Bryk & Weisberg (1976) and Bryk, Strenio, and Weisberg (1980).

STEP 1: Estimate of growth rate. First, an unbiased estimate of growth is obtained. This is done by regressing pretest scores on age and using the resulting regression coefficient as the estimate of growth rate. This figure is then entered into a formula to obtain an unbiased estimate of growth rate.

STEP 2: Calculate value added. The value added by the instruction is calculated using a formula to determine the effect of the instruction which takes into account the natural growth of students over time. To increase the precision of this model, Bryk et al. (1980) also describe how it can be extended to incorporate background variables that may be related to individual growth rates. This is done by regressing the pretest on age and the first order interactions of age and the background variable.

Example of Value-added Analysis Use

This procedure was used in a study of a microcomputer-based reading and writing program in grades K-1 (Garretson, Vertuno, King, & Roblyer, 1983). All students in these grades (approximately 100) received the computer-based instruction, which was to take place over the school year. Several measures were used to assess skill levels, including a Nonsense Words Test devised by the school resource teacher and several subtests of the California Test of Basic Skills (CTBS). The value-added analysis was used with the Nonsense Words Test. (Analysis methods for the other test will be discussed later under another design.) The results indicate that the instruction resulted in significantly more learning than that expected from normal maturation.

Design #3: Non-Equivalent Dependent Variables (NEDV)

This design, like sequential analysis, is a time-saving evaluation procedure. Very often, schools want to evaluate the effectiveness of not one but several computer-based programs by a vendor. The NEDV design would allow evaluation of several programs at once. For example, suppose school personnel wanted to evaluate a computer-based module on fractions as well as one on geometry. Students could be given pretests on both units, instruction on fractions only, and then posttests on both units. If statistically and educationally significant changes occur on the fractions unit but not only the geometry test, the fractions unit is judged to be effective. Then instruction on the geometry would be given, followed by a test on geometry. If the change from the first to the second geometry tests is not significant, but change from the second to the third is significant, then the geometry unit is judged effective.

Of course, the evaluators would have to be satisfied that the tests were equally reliable, and that students were not already scoring at the ceiling of either test at the time they were pretested. In addition, the investigators would have to be satisfied that the change in performance in either case was not due to some learning experience other than the modules, such as some parent tutoring on this topic during the instructional period.

Such an approach has two advantages. First, no non-instruction control group is needed, since performance on one variable acts as a control for the other. Second, older students can be used, since, unlike the value-added analysis, no independent variable such as age is required. This design may be diagrammed as:

$$\begin{array}{lllll}
0_{1A} & X_A & 0_{2A} & & \\
\text{- - - - - - -} & \text{- - - - - - -} & \text{- - - - - - -} & \text{- - - - - - -} & \text{etc.} \\
0_{1B} & & 0_{2B} & X_B & 0_{3B}
\end{array}$$

where 0_{1A} and 0_{2A} are pre- and posttests for the variable that is hypothesized to be affected by the instruction on fractions and 0_{1B} and 0_{2B} are pre-and posttests for the variable hypothesized *not* to be affected during the first treatment period.

In order to be sure this design will be credible for the situation, the evaluator must (a) specify in advance which variables are expected to change and which are not; (b) demonstrate that differential change is not due to differential reliability of measures or to ceiling/basement effects; and (c) use variables which are conceptually so similar that both would be equally affected by the same threats to internal validity, namely maturation and testing.

Berquist and Graham (1980) expanded this design to allow the evaluation of many objectives over many waves of measurement. They point out that this approach is useful to evaluate instruction with single students as well as groups. For details about when this approach is most appropriate, see Berquist and Graham (1980).

If more than one unit of instruction is to be evaluated and the group is sufficiently large (e.g, 60 students), this design can be converted to a true experimental design which does not require a nonCAI control group. Where instruction can be individually administered to each student (as with CAI), the following design could be used:

$$\begin{array}{cccccc}
R & 0_{1A} & 0_{1B} & X_A & 0_{2A} & 0_{2B} \\
R & 0_{1A} & 0_{1B} & X_B & 0_{2A} & 0_{2B}
\end{array}$$

Here, students must be randomly assigned (R) to either fractions or geometry (X_A or X_B) and pre- and posttested with both unit tests (0_{2A} and 0_{2B}). Then two separate analyses must be done: one in which the group that received the geometry instruction serves as the control group for the fractions instruction and the other in which the fractions group serves as a control for the geometry instruction.

This design is really better in some ways than one having a nonCAI control group because it is a better control for Hawthorn effects. The novelty of the CAI experience is the same for both groups so that the computer aspect alone cannot be expected to produce significant achievement differences between groups.

Procedures for Non-Equivalent Dependent Variables Design

See Cook and Campbell (1979) for a discussion of the analysis of this design. Briefly, steps to implement the NEDV design are as follows:

STEP 1: Randomly assign and pretest students. Students are randomly assigned to two instructional groups, and both are pretested on both modules.

STEP 2: Give instruction. One group receives instruction in fractions, and the other geometry instruction.

STEP 3: Compare posttest means. Students are posttested and the means of the two groups are compared to evaluate the effectiveness of the fractions instruction. The same analysis is done to evaluate geometry instruction.

Example of True Experimental Design Involving Non-Equivalent Dependent Measures

One study using this design as part of its total evaluation program was the ETS/LAUSD study of computer-assisted instruction and compensatory education (Ragosta, Holland & Jamison, 1982). Schools in the Los Angeles Area School District used math and language CAI available from the Computer Curriculum Corporation. In Grade 4 students were randomly assigned to receive two sessions of CAI daily. They received either: (a) two sessions of mathematics (MM), (b) one session of reading and one of language arts (RL), or (c) one session of mathematics with one session of reading or language arts on alternate days (MRL). All students were pretested in the fall with the Iowa Test of Basic Skills (ITBS) and a curriculum-specific test (CST) in mathematics, language, and reading. They were posttested the next spring with the California Test of Basic Skills (CTBS) and the CST. Thus, the RL students were controls for evaluating the effectiveness of two levels of mathematics instruction (MM and MRL), while the MM students were controls for the evaluation of two levels of reading (RL and MRL). A regression analysis was used to analyze the data.

Although this was a complex study with many kinds of groups and taking place over several years, the results were generally positive for both math and reading groups. Math results were usually higher than for reading.

Design #4: Regression-Discontinuity

Although this is the one design which actually requires a noinstruction control group, the regression-discontinuity design is especially useful when compensatory or enrichment programs are to be evaluated. This is because of its requirement that students be pretested and all students below (or above) a certain score level be placed in an instructional group. For example, suppose a school district purchased a computer-based remedial math program with Chapter III compensatory education funds. Any students not in the compensatory Chapter III program could serve in the control group. Before any instruction is given, all students are pretested. Those whose scores fall below the cut-off level are placed in the instruction group. Those above this level are placed in a non-instruction group. At the end of the instruction time, the posttest is administered to all those who received the pretest. If the difference between the adjusted means of the groups is statistically significant, the instruction is effective.

The principal advantage of the regression-discontinuity design over other quasi-experimental designs is that it is not biased by errors of measurement in the pretest. That is, in the absence of an instructional effect, the expected posttest adjusted mean difference is zero. That is not true in designs in which unknown selection variables operate in the formation of instruction and control groups.

A primary disadvantage of the design when compared to a randomized instruction-control group design is that it has low power to reject the hypothesis of no differences between groups. If a randomized design used a sample of 100 students, the regression-discontinuity design would require 275 students to have equal power (Reichardt, 1979). In addition, if one of the groups is relatively small, the estimate of the regression of posttest on pretest for that group may be unstable.

Two other precautions should be kept in mind when using this design. First, the process of assigning students to groups should be done without error. If the pretest is to be used as the covariate in the analysis, selection should be based on it *only*. If other sources of information, such as teacher or parent judgments, are used in addition to or instead of the pretest, the analysis may be seriously biased. For a further discussion of these problems and some ways of remedying them, see Huitema (1980) and Reichardt (1979).

The second precaution in using this design concerns the fact that non-linear relationships between pre- and posttest, which result from selection-motivation factors, may cause spurious treatment effects to occur. Reichardt (1979) suggested that the first step in analyzing data from this design should be to plot the raw data to see what relationships are suggested. If non-linear or interactive effects are suggested, the analytic model given previously can be expanded in an attempt to account for them.

Procedures for implementing this design are detailed in Huitema (1980). They are outlined as follows:

STEP 1: Administer the pretest and group students.

STEP 2: Give instruction and posttest students.

STEP 3: Assign treatment variables. A "dummy" treatment variable is constructed by assigning a score of "1" to each member of the treatment group and a "0" to each member of the non-treatment group.

STEP 4: Perform regression analysis. An analysis is performed using the following model:

$$Y = a + b_1 T + b_2 \text{ Pre}$$

where Y is the adjusted posttest score, T and Pre are the treatment and pretest variables, a is the regression intercept, b_1 is the difference between the adjusted means of the treatment groups, and b_2 is the regression coefficient for the pretest. If b_1 is statistically significant, the instruction is considered to be effective.

Example of Regression-Discontinuity Use

No example of this design in evaluation of instructional materials could be found in the literature. However, Cook and Campbell (1979) cite a study by Seaver and Quarton (1973) which examined how students' grades in one quarter were affected by making the Dean's list based on grades from the previous quarter. The investigators obtained the grades of a group of students during both quarters, and placed students in groups according to who did and did not make the Dean's list. Students who made the list were expected to do better than they would have done if they had not had the recognition of being on the Dean's list. Cook and Campbell (1979) interpreted the results of the study as not supporting the hypothesis because of the curvilinear trend in the data.

Design #5: Cohort Design

Cohorts are groups of persons who follow each other through an institution. For example, last year's first graders are cohorts of this year's first graders. The cohort design compares the performance of students in one instruction group with their cohorts who did not receive the instruction. This design is especially useful in evaluating the effectiveness of newly-installed, year-long computerbased programs where it is not feasible to withhold the program from some students in order to have a control group. Suppose a school district wished to study a popular and highly publicized new computerbased reading program for a given grade level before they decide whether or not to continue it in subsequent years. Attempting to select one or more classes which will not get to use the package the first year may be a political problem when parents and teachers find out what has happened. With a cohort design, all students in the grade level can use the package, if desired.

The design makes the assumption that student cohorts are comparable in most respects, and that differences between them on a dependent variable can be attributed to instruction administered to one of the groups.

In its simplest form, the design is as follows:

$$\begin{array}{ccc} 0_1 & & \\ & X & \\ & & 0_2 \end{array}$$

0_1- represents, for example, a final achievement measure of students in a certain grade in a given year, and 0_2- $_p$ represents the same final achievement measure of students in the same grade the year after that, after having received some experimental instruction. If a t-test of the differences between means of 0_1- and 0_2- yields a statistically significant result, and that difference is large enough to be considered educationally significant, the instruction is said to be successful.

A disadvantage to this design is that its use depends on the yearly availability of appropriate equivalent measures. Also, Cook and Campbell (1979) indicated that the design in this form is weak because differences between the two groups (other than the instruction) could cause a difference in achievement (selection). Events other than the instruction could depress achievement for the control cohort or enhance it for the instructional cohort (history). Variations in testing conditions or procedures could also be responsible for differences in performance. The design can be strengthened by adding a pretest if it is available. This augmented design is shown as follows:

$$\begin{array}{ccccc} 0_1 & 0_2 & & & \\ & & 0_3 & X & 0_4 \end{array}$$

The pretest could be an end-of-year test or one that is routinely administered in the fall. The data would be analyzed using regression analysis, the same method used in the regression-discontinuity design. The cohort design is considered to be superior to some designs in which the control group is not equivalent to the experimental group (e.g., the groups are obtained from different schools). The populations from which the cohort were drawn were probably more similar than those using students from different schools. Also, the school environment (principal, teachers, facilities, etc.) for the cohorts would probably be more similar than would the environments in different schools.

Procedures in Using the Cohort Design

STEP 1: Calculate means and standard deviations. The means and standard deviations of both groups on both measures are determined.

STEP 2: Perform regression analysis. This analysis actually involves several steps. For a detailed discussion, see Cook and Campbell (1979) and Huitema (1980).

Example of Cohort Design Use

Garretson, Vertuno, Roblyer and King (1983) used this design in a study of a computer-based reading and writing program, the "Writing to Read" program from IBM, Inc. There were approximately fifty students in each grade (K-1), and it was not considered feasible to withhold the instruction from any of them. A readiness test was available for both the instruction and control kindergarten cohort, so it was used as a pretest. No pretest measure was available for the first grade students. The dependent variable was the reading score (prereading for kindergarten) from the California Test of Basic Skills (CTBS) administered at the end of both kindergarten and first grade.

The results for neither grade showed a significant advantage over instruction from the year before. These findings are not in agreement with those in the value-added analysis performed in the same study in which the instruction effect for the Nonsense Words Test was significant.

Discussion and Summary

Depending upon the needs of the evaluator and the characteristics of the situation, each of the five designs discussed here can serve as a practical alternative to a design requiring a no-treatment control group. Several caveats are, however, in order. Selecting the best design for the need and using appropriate procedures during the study are of critical importance.

Choosing the Appropriate Design

Each of the designs described here can be appropriate for studying various aspects of computer-aided instruction. But like all evaluation designs, each has assumptions which must be met and requirements which must be considered if the research is to be perceived as valid. The flowchart shown in Figure 1 can be used to choose the most appropriate design for the situation.

After selecting the most useful of these designs, the researcher must recognize that, like all designs, each has its strengths and weaknesses. Using two of these designs together can work to make the total study stronger. For example, in Garretson et al. (1983), both the cohort design and the value-added analysis were used. Ideally, the results of these two designs would be the same and thus support the same conclusions.

Specifying Power

The sequential analysis method is the only one of the five which forces the evaluator to set Alpha, Beta and effect size prior to conducting an evaluation. However, this practice is desirable for all designs (including true experiments) in order to determine that the study has sufficient power to reject the null hypothesis when it is false or to allow acceptance when it is true (Cohen, 1969).

Required Expertise

It should be apparent that using these designs requires some expertise in both evaluation methods and statistics. The minimum training would probably include introductory courses in research methods, descriptive statistics, and inferential statistics. Some experience in implementing evaluation studies is also helpful. To those with such expertise available to them, the designs described here can and should be useful in evaluating the effectiveness of computerbased instruction and should help direct the course of computer use throughout the field.

References

Alessi, S. M., Siegel, M., Silver, D., & Baines, H. (1982-83). Effectiveness of a computer-based reading comprehension program for adults. *Journal of Educational Technology Systems, 11* (1), 43-57.

Berquist, C. C., & Graham, D. L. (1980). Developing models for special education. Using a multiple time-series, multiple baseline design in the ESEA Title IV project. *Evaluation Review, 4* (3), 307321.

Bryk, A. S., Strenio, J. F., & Weisberg, H. I. (1980). A method for estimating treatment effects when individuals are growing. *Journal of Educational Statistics, 5* (1), 5-34.

Bryk, A. S., & Weisberg, H. I. (1976). Value-added analysis: A dynamic approach to the estimation of treatment effects. *Journal of Educational Statistics, 1,* (2), 127-155.

Burr, I. W. (1976). *Statistical quality control methods.* New York: Marcel Dekker, Inc.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasiexperimental designs for research.* Chicago: Rand McNally College Publishing Company.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation designs and analysis issues for field settings.* Chicago: Rand McNally College Publishing Company.

Crandall, N. (1977). *An analysis of the impact of CAI on a program designed to ameliorate the ef-*

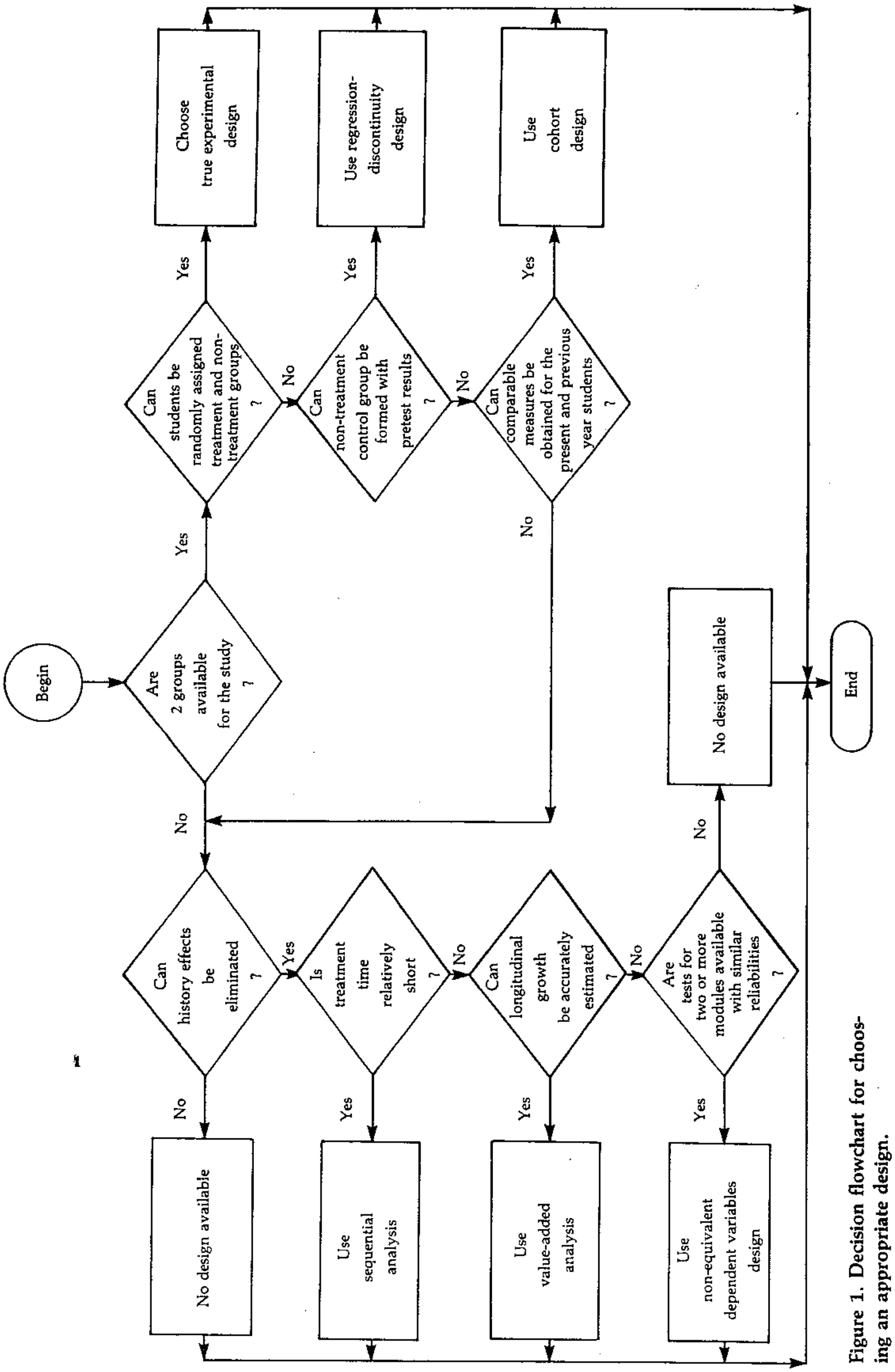

Figure 1. Decision flowchart for choosing an appropriate design.

fects of racial isolation in the Los Nietos School District. Washington, DC: Office of Education. (ERIC Document Reproduction Service No. ED 147 395)

Epstein, K. I. (1975, April). *Sequential plans and formative evaluations*. Paper presented at the Annual Meeting of the American Educational Research Association, Washington, DC.

Epstein, K. I., & Knerr, C. S. (1978). Applications of sequential testing procedures for performance testing. In D. J. Weiss (Ed.), *Proceedings of the 1977 computerized adaptive testing conferences*. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program.

Fletcher, J. D., & Atkinson, R. C. (1972). Evaluation of the Stanford CAI program in initial reading. *Journal of Educational Psychology, 63* (6), 597-602.

Garretson, P., Vertuno, E., Roblyer, M., & King, F J. (1983). A computer-based reading and writing program for K-1. Presentation at the Florida Instructional Computing Conference, Tampa.

Holland, P. W. (1980, April). Computer-assisted instruction: A longitudinal study. Panel presentation at the Annual Conference of the American Educational Research Association, Boston.

Huitema, B. F. (1980). *The analysis of covariance and alternatives*. New York: John Wiley and Sons.

Jones, M. (1978, March). *TICCIT applications in higher education: Evaluation results*. Paper presented to the Association for Development of Computer-based Instructional Systems, Dallas. (ERIC Document Reproduction Service No. ED 107 039)

King, F J. (n.d.) *A plan to evaluate a microcomputer/videodisc system to teach the jet fuel thermal oxidation test (JFTOT)*. Final report submitted to the Battelle Columbus Laboratories.

Lysiak, F., Wallace, S., & Evans, C. (1976). *Computer-assisted instruction 1976-76 evaluation report*. Fort Worth: Fort Worth Independent School district, 1976. (ERIC Document Reproduction Service No. ED 140 495)

Mosteller, F., & Tukey, S. W. (1977). *Data analysis and regression*. Reading, Massachusetts: Addison-Wesley.

Poore, J. H., Qualls, J. E., & Brown, B. L. (1981). The educational effectiveness and economics of delivery of the PLATO basic skills mathematics lessons: A field study. *AEDS Journal, 15* (1), 31–51.

Ragosta, M., Holland, P.W., & Jamison, D. P. (1982). Computer assisted instruction and compensatory education: The ETS/LAUSD study. Final report, -19. Princeton, NJ: Educational Testing Service.

Reichardt, C. S. (1979). The statistical analyses of data from non-equivalent group designs. In T. D. Cook & D. T. Campbell (Eds.), *Quasi-experimentation design and analysis: Issues for field settings*. Chicago: Rand McNally College Publishing Company.

Seaver, W. B., & Quarton, R. J. (1973). *Social reinforcement of excellence: Dean's list and academic achievement*. Paper presented at the 44th Annual Meeting of the Eastern Psychlogical Association, Washington, DC.

Suppes, P., & Morningstar, M. (1972). *Computer-assisted instruction at Stanford, 1966-68: Data, models, and evaluation of the arithmetic programs*. New York: Academic Press.

Wald, A. (1947). *Sequential analysis*. New York: John Wiley and Sons.