



Curriculum Research: *Uses of Evaluation*¹

Ted Ward

Educational evaluation demands both *measurement* and *valuing*. Measurement requires the use of a procedure to quantify and hold up against an arbitrary standard—to provide an objective description. Valuing requires a procedure for assessing the objective description in terms of its worth—to allow the objective data to be related to some basis of value and significance.

Most educational evaluation is concerned with comparisons. Often the basic concern is for assessment of the consequence of a particular educative procedure. Sometimes the comparison is between two (or more) educative procedures or experiences. But in any case, the evaluation is incomplete until *what is* is related to *what is desired or intended*.

In the field of curriculum research, evaluation is basic. Scientific methods of analysis, description, and prediction make their greatest contribution to curriculum development through the procedures of evaluation. Curriculum decisions are of two basic sorts: 1) The *what*, *when*, and *how* and, 2) The *why*. Scientific methods of evaluation can provide considerable help with the former set of questions, but the help really useful only within a context of values arising from the continued analysis of *why*—the reasons, the intentions, the goals, and the beliefs about what is important.

Thus the application of scientific methods of data-gathering and analysis are relatively easy to relate to the *what*, *when*, and *how* questions. But curriculum research workers must learn not to stop here; there are ways to apply scientific investigation procedures to the data underlying the *why* decisions as well. The major questions in this sector—the valuing sector—deal with *who* has the responsibility for setting goals, *how well* are the goals clarified and communicated, *how responsive* the system is to the goals.

The following series of questions and problems indicate something of the range of issues to which any thorough educational evaluation must be related. A sequence is implied in the list, but the particular conditions in a particular evaluation project will usually demand alteration of this sequence. The major purpose of this list is to make the evaluator more aware of the many factors that impinge on an educational program, thus he or she can be more alert to the unchallenged assumptions and the inquiry gaps that afflict any particular evaluation task that is defined as less than thorough.

First, consider the assumptions upon which the list is built:

Assessment of educational programs and services must have at least three aspects: 1) information (data gathering), 2) evaluation, and 3) planning.

Data Gathering—determining what data are needed and getting them—is an essential step. Logical though this statement is, it is amazing how often an educational assessment fails to gather data before pushing ahead with evaluation! Many evaluations begin from a framework of given services (institutions and programs). This approach is always destined to give a less-than-complete picture of learners and their needs.

Evaluation—determining what is important and weighing the status quo against identified values and aspirations—is the focal problem. Quantitative data are but one sort of input; even more crucial are data about values—the inputs that provide a means for establishing the relationship between empiricism and aspiring humanism.

¹ Preparatory document for Curriculum Research Seminar, November, 1967. Michigan State University

Planning is less classically a part of the assessment process in education, but concern for improvement in education demands that assessment programs be designed so as to make direct linkages between evaluation and planning. Here is the ultimately important “so what?” dimension of assessment. Here is the essential justification for inquiry. Assessment that concludes with the shelving of reports has little justification in a pragmatic era.

I. Information Problems:

- A. The target population as learners
 - 1. Who are they?
 - 2. Where are they?
 - 3. Which of them are being reached by educational services?
 - 4. Which of them have needs that are not being met?
- B. The service programs
 - 1. What are the species of service programs for this target population?
 - 2. What agencies currently have responsibility for each of the services?
 - 3. What articulation exists among the programs and institutions?
 - 4. What changes (enlargements, reassignments, etc.) have taken place in the past decade, and what has been the rationale for each?
- C. The evident concept (the actualized concept) of education for the target population. (Note: Regardless of any official formal enunciation of a concept of education for this target population, a responsible data base must include an empirically deduced definition. The formal view and the informal views expressed through practices and policies are often highly divergent.)
 - 1. What in practice, constitutes being a member of the target population?
 - 2. What do they need?
 - 3. What can be done for them?
 - 4. As of now what do we think we are *able* to do for them?
 - 5. How do we know when they are ready to progress from one service to another?
 - 6. What constitutes being an *educated* person?

II. Value Problems:

- A. Interpretation of the needs of the target population
 - 1. What meaning is derived from the descriptive data in I.A. above?
 - 2. What is the political meaning drawn from I.B. in relation to I.A. above?
 - 3. Whose opinions affect and constitute II.A.1. and II.A.2. immediately above?
 - 4. To what extent can a more clear picture of the status quo be derived?
- B. The aspiration and commitment of the establishment
 - 1. What is the model of educational services to which the several voluntary and private agencies are willing to be committed?
 - 2. What is the commitment of the public agencies? To what model of services?
 - 3. What are the differences between evident accomplishment (as assessed by all of the above) and the commitments?

III. Planning Problems:

- A. Deriving meaning from assessment
 - 1. Whose responsibility is it to enunciate evaluative meaning?
 - 2. What changes are suggested by the evaluation?
 - 3. What model for future attainment can be drawn?
 - 4. What agencies and units will be charged with implementing change toward the new model?
- B. Motivating change
 - 1. What agencies are authorized to motivate change?
 - 2. What communication processes can be used to stimulate the relevant segments of the public?
 - 3. What reward systems can be defined for the new model?

Uses of Evaluation

Educational evaluation is expected to serve many different purposes. Many uses of evaluation are professionally motivated, but one suspects that not all the roles evaluation is expected to play are professionally defensible.

In reference to the school curriculum, evaluation historically has been more a tool of administration than a tool of the teacher. Commonly, teachers evaluate students; administrators and supervisors evaluate the curriculum. Perhaps this distance between the teacher and the evaluation of the curriculum should be considered abdication. On the other hand, perhaps teachers haven't been given sufficient training and opportunity to become involved in the evaluation of curriculum. In any case, the purposes that evaluation is made to serve range from truly professionally to essentially unprofessional. It pays both teachers and administrators to recognize the difference.

There are several distinctly different categories of motives underlying evaluation. Since many of these are obviously indefensible, it is only timidly suggested that we should actually classify evaluation projects according to the scheme following. This classification scheme is offered as the informal alternative to self-delusion—against that day when we may be more able to call a turkey a turkey.

Categories of Motives for Evaluation of Educational Programs: Otherwise Known as "Models of Evaluation"

1. Basis for Progress

One of the most acceptable and exciting models of evaluation is that which attempts to define, describe, and quantify the way things are, the goals in sight, and the available resources that would lead toward the goals. The future of education would be more likely a future of progress if evaluation studies were motivated as a *basis* for progress.

2. Accomplishment through Time

Closely related to the first model, this one puts a premium on being able to record a chart of change. It may or may not be goal-oriented; it may or may not be concerned with documentation of the processes by which goals are to be attained. This model is usually intended to be a management tool, allowing easy comparisons between "where we were and where we are."

3. The Squeaky Hinge

Consider the analogy: A door that squeaks; three hinges on the door; two drops of oil left in the can—which two hinges get the oil? Call for an evaluation; put an ear next to each hinge in turn; find the hinge most clearly not squeaking; give a drop of oil to the other two. In this model, evaluation is called in after a problem arises. Its function is to serve management and to assure that resources are efficiently applied “where they will do the most good.”

4. Accountability

Another of the standard management motives of evaluation is to provide a basis for accountability. Evaluation, especially when instrumented through a designed series of reporting moments can be a useful way to find out what people in the structure are doing, what problems are being dealt with, with what frequency and requiring what portions of time. Problems of this sort are the stuff management is made of, and objective evaluation of such factors is certainly more acceptable than guessing.

5. Find the Mud

In politics it is known as “mud slinging” when the very worst light is thrown on a candidate in order to embarrass him or her for past misdeeds. In educational evaluation, sometimes it seems that the basic motive is to embarrass people. Evaluations so slanted that the only possible result is some sort of exposé give the whole field of educational assessment a bad name. Surely there are weak or faulty aspects of *any* educational enterprise; it seems a waste to engage in evaluation merely to prove that they exist.

6. Punitive

There are few educators so masochistic as to voluntarily engage in continuous evaluation. Many of the aspects of a thorough evaluation are both time-consuming and professionally demanding. Some administrators have been known to use evaluation as a punishment. When you haven’t got any better way to “get” a subordinate, you can always command him to evaluate his or her work. Truly, the commanded evaluation may be well motivated and thus might fit better into one of the models described earlier, but sometimes, just possibly, the major motive is to punish—perhaps to act out some previous threat. This model unfortunately keeps alive the folklore that evaluation is an unpleasant chore rather than an essential element of educational planning and development.

7. The Heavy Hand

One of the popular cries of both the student and the educational practitioner today is, “Nobody cares.” One way for an administrator (or even a teacher) to show that he or she cares—REALLY cares—is to ask for and to review (or pretend to review) lots and lots of evaluative data. The person operating this model of evaluation is thus able to fend off all sorts of criticisms about not knowing what’s going on. At the same time, he or she is maintaining an aura of omnipresence—a very heavy-handed use of evaluation.

8. The Whitewash

The opposite counterpart of the “Find the Mud” model (and equally indefensible) is the Whitewash model. When an evaluation must be done, for some reason, and you really don’t want to participate—or you really don’t want to treat it objectively—you can resort to creating a “snow job,” known in earlier days as a “whitewash.” To bring this off, you need to evaluate only the right things—those that are bound to sound good and show up well. Learning anything is completely beside the point. Your lack of objectivity is compensated by a wide representation of testimonials

(from the right people) that indicate your own opinions are shared by others. At all costs, cover all errors and conceal all failures.

9. Madison Avenue

It is still popular to attempt to persuade through pseudo-science. (This is an odd testimony to the persistence in the American character), since it has long been recognized by behavioral scientists that facts alone don't change people.) Evaluation is sometimes outfitted with white coats and imposing grey-haired thick glasses in order to conjure up an image of credibility. One suspects that much evaluation of federally supported projects falls into this model, by design or by coincidence. The major characteristics are glossy appearance, big names as consultants or authors of the methods employed, and plenty of statistical jargon—whether or not relevant.

10. Mañana

The ancient human tendency never to do today what can be put off till tomorrow takes on a whole new sophistication in the employ of educational evaluation. "Someday we'll have to look at that," is slowly being replaced in the curriculum worker's bag of tricks. Now it's "We'll have to get some data on that." The result is the same—avoidance and procrastination. When the *need* for evaluation becomes the excuse for doing nothing, it is equivalent to employing gasoline to smother a fire. Sooner or later may be sooner than you think! Evaluation has a habit of breaking out of its "someday" box and becoming part of today's scene.

11. Challenging the Status Quo

Sooner or later in any educational enterprise, there is need to challenge the status quo. At such time it is popular to utilize evaluation as one of the instruments to stimulate pressures for change. When the need for change is assumed or accepted as an *a priori*, the particular elements of the system tagged for evaluation are the "sore spots"—the vulnerable factors. Such an evaluation is designed to create instability and to make the structure ripe for alteration. The danger in evaluating under these conditions is the obvious threat to objectivity. Challenging the *status quo* must be considered as one more of the utilitarian and professionally shady uses of evaluation.

12. Defending the Status Quo

It works both ways! Evaluation is not inherently a force for change. Evaluation, when slanted toward selectively highlighting the strong points in an education organization, process, or system, can serve to protect the status quo against forces of change. Since forces of change are not necessarily devoted to change for the better, administrators learn how to assure that evaluation allows them to put their organization's "best foot" forward. Again, the threat to objectivity is obvious. The reflex action of an attacked organization is to defend itself. Evaluation that begins as a defense (so long as it is not purely a "whitewash") can be converted into a more objective basis for improvement of the organization or educational structure. Thus defending the *status quo*, although utilitarian, can become the trigger for a more professional use of evaluation.

13. The Rose-Colored Glasses Model

When the motive to be served is more a matter of self-satisfaction, pride or boastfulness than it is a matter of self-defense, the evaluation is restricted to highlighting the good features and seeing the bad features in some sort of overly-generous or hopeful light. The "Rose-Colored Glasses" evaluation is usually easy to spot in the slanting of its final report. The data may be sound enough and the range of factors examined reasonably adequate, but what is read into the data is rather starry-eyed and even "Pollyanna." Statements of this sort show up: "Although in the first two years,

Project Headstand has been plagued by serious problems of public misunderstanding and professional apathy, once the problems associated with the manufacture and delivery of the long overdue instructional equipment can be solved Project Headstand should attract national attention.” One must hope not.

Curriculum Research Exercises

Research and Practice (Circle the appropriate response in the following multiple choice options)

1. Which of the following is *not* usually considered as a goal of educational research?
 - a. Description of student behaviors in a classroom.
 - b. Explanation of student behaviors in a classroom.
 - c. Prediction of student behaviors in particular classroom situations.
 - d. Prescription of classroom behavioral objectives.
2. When researchers say that investigating problems is more difficult in the behavioral sciences than in the physical sciences, they usually mean that in the behavioral sciences:
 - a. Variables cannot be satisfactorily controlled.
 - b. Less precise measurements can be made.
 - c. Objective measurements are not possible.
 - d. Exactly two of the above.
3. For experimental results from the laboratory to be used confidently as a basis for classroom procedures, it is necessary that:
 - a. Evidence on procedures suggested by the laboratory findings be collected in the classroom.
 - b. The study be confirmed in other laboratories.
 - c. The classroom procedures be derived from learning theory.
 - d. The laboratory conditions be duplicated in the classroom.
4. If a researcher conducts a study to establish generalizations applicable to other samples of a population, his or her research is most likely to be called _____ research.
 - a. Field.
 - b. Basic.
 - c. Survey.
 - d. Action.
5. A classroom study of an educational method tends to be superior to a laboratory study of it in this respect:
 - a. A classroom trial indicates how much the method accomplishes under realistic operating conditions.
 - b. In the classroom it is easier to identify the specific elements that cause the method to succeed or fail.
 - c. Classroom studies tend to be less costly.
 - d. A classroom experiment uses real subject matter rather than artificial stimuli such as nonsense syllables.
6. If a researcher attempts to maintain close control of the research situation in order to obtain scientific precision, his or her research is most likely to be called _____ research.
 - a. Basic.
 - b. Applied.
 - c. Action.
 - d. Field.

7. A research study in education:
 - a. Suggests what it might be wise to try in the classroom.
 - b. Is inadequate unless it produces an effective technique.
 - c. Is expected to yield a definitive answer to a practical question.
 - d. Can hardly be expected to give relevant information if carried out only in a psychological laboratory.
8. To solve practical professional problems educators typically rely on:
 - a. Authoritative opinion.
 - b. Personal experience.
 - c. Relevant research.
 - d. Both a. and b.
9. To counteract the widespread disillusionment with experiment in education:
 - a. Educators must be made to realize that only experimentally produced practices should be adopted.
 - b. Complex designs should be avoided to reduce threat to educators.
 - c. Researchers must accept the attitude that cumulating sound knowledge is a slow tedious process.
 - d. All of the above.

Attitudes (Check “Agree” or “Disagree”)

10. One reason there are very few studies which can serve as models of curriculum research is because curriculum research workers have refused to adopt a holistic definition of curriculum. Agree___ Disagree___
11. Curriculum research will become more useful to educators and researchers if more attention is paid to adequate assessment of immediate and intermediate outcomes of particular subsystems within the curriculum. Agree___ Disagree___
12. Curriculum theory at present produces numerous researchable hypotheses. Agree___ Disagree___
13. One of the difficulties of performing meaningful curriculum research is the lack of specific behavioral objectives. Agree___ Disagree___
14. The methodological focus of *curriculum evaluation* should be precise and reliable measurement, while the methodological focus of *curriculum research* should be principles of research design. Agree___ Disagree___
15. The most influential factor in the curriculum innovations of the past decade has been the development of curriculum theory. Agree___ Disagree___
16. The problem of cultural deprivation has had a major impact upon curriculum planning and research. Agree___ Disagree___
17. Federal aid to public education has stimulated much experimental curriculum research. Agree___ Disagree___

18. One of the major difficulties in planning curriculum research is the lack of relevant principles of research design. Agree___ Disagree___
19. The aim of curriculum research is to prove that a given principle is true.
Agree___ Disagree___

The four-choice objective items that follow deal with research terminology, concepts, and skills. The intent is to get some idea of the research knowledge and skills already available to seminar participants. What is now available will become the foundation for extending your knowledge and skills, and what is lacking will become objectives for future activities. For each item choose the alternative you believe would be chosen by a group of professional researchers, and circle the appropriate alternative on your answer sheet.

Correlation

6. A correlation coefficient:
- Is a measure of efficiency that compares achievement to potential achievement.
 - Tells which of two variables in an experiment has greater influence.
 - Indicates which variables in an experiment have been adequately controlled.
 - Tells how closely standing in one variable is related to standing in another variable.
7. In a study of several hundred girls, “desire to help others” correlated $-.27$ with “the number of years a girl has been in training for nursing.” The size of the correlation implies that:
- There is only a chance relationship between the two variables.
 - A measure of desire to help others would be of considerable use in selecting nurses.
 - There is a relationship that nursing educators should try to understand.
 - Nurses’ training causes girls to become less benevolent.
8. Suppose X and Y are continuous variables having interval scales. As a measure of association between X and Y the rank difference correlation tends to be a:
- Slight under-estimate.
 - Large under-estimate.
 - Slight over-estimate.
 - Large over-estimate.
9. If Y is a continuous variable and X is a discrete variable that can have exactly two values, the product moment correlation between X and Y is called a _____ coefficient.
- Tetrachoric correlation.
 - Phi.
 - Biserial correlation.
 - Point biserial correlation.

Measurement Statistics

10. Which of the following is a type of standard score?
- Grade equivalents.
 - Intelligence quotients.
 - Percentile ranks.
 - Mental-age norms.

11. The standard error of measurement is closely related to the discrepancy between student score and:
 - a. What they could do if they worked to their full capacity.
 - b. The teacher's estimate of their ability.
 - c. The average score earned by students of their mental ability.
 - d. The score he would have if their performance were sampled thoroughly.
12. The major problem of validity associated with educational achievement tests is _____ validity.
 - a. Concurrent.
 - b. Predictive.
 - c. Construct.
 - d. Content.
13. The major problem of validity associated with aptitude tests is _____ validity.
 - a. Concurrent.
 - b. Predictive.
 - c. Construct.
 - d. Content.
14. A high reliability coefficient for a test means that:
 - a. The test has been well standardized.
 - b. Students would earn similar scores if the test were repeated.
 - c. The test measures what it is intended to measure.
 - d. Each student is exhibiting his or her full ability.

Hypotheses

15. The purpose of a statistical hypothesis is to:
 - a. Describe some frequency distribution for a given population.
 - b. Describe the relationship between two or more statistics.
 - c. Describe the researcher's expectations for the study.
 - d. None of these.
16. An investigator reports results significant at the .05 level. This means that the probability of rejecting a true statistical hypothesis is _____, and the probability of accepting a false statistical hypothesis is _____.
 - a. .95, .05
 - b. .05, .95
 - c. .05, unspecified.
 - d. Unspecified, .05
17. Which of the following is *not* characteristic of all useful research hypotheses?
 - a. They are linked either to theory or available empirical data.
 - b. They are capable of being tested by objective means.
 - c. They are concisely and clearly stated.
 - d. They are stated in the null form.
18. A school board is considering installing computer equipment as an aid to instruction, but feels that the expense cannot be justified unless at least 70% of the students show improvement. If they make a wrong decision, taxpayer resentment will hinder school programs for several years. If a research study is carried out, which set of statistical hypotheses is most appropriate? (H is the null hypothesis and A is the alternate hypothesis.)
 - a. H: $p = 70\%$
A: $p \neq 70\%$
 - b. H: $p \geq 70\%$
A: $p < 70\%$
 - c. H: $p \geq 70\%$
A: $p < 70\%$
 - d. H: $p \geq 70\%$
A: $p < 70\%$

- b. $H: p \leq 70\%$
A: $> 70\%$
- d. $H: p = 70\%$
A: $p < 70\%$
19. For a specific alternate hypothesis other than the null hypothesis, the probability of committing a type II error is given by the _____ curve.
- Operating characteristic.
 - Significance.
 - Acceptance.
 - Power.

Frequency Distributions

20. Suppose the variance of a set of scores is equal to 16. If three points are added to each of the scores the new variance is equal to:
- 16
 - 19
 - 48
 - 144
21. Suppose the mean of a set of scores is equal to 10. If two points are added to each of the scores the new mean will be equal to:
- 10
 - 12
 - 20
 - None of these
22. For a symmetric distribution the mean is a better measure of central tendency than the median because:
- The value of the mean is less influenced by the sampling fluctuations.
 - The value of the mean is less influenced by extreme scores.
 - The value of the mean is easier to compute.
 - At least two of the above.
23. Suppose the standard deviation of a set of scores is equal to 3. If each of the scores is multiplied by two, the number 36 is the value of the new:
- Range.
 - Standard deviation.
 - Variance.
 - None of these.

Statistical Tests

24. In which of the following situations is chi-square analysis *not* appropriate?
- An observed frequency distribution is to be compared with a theoretical frequency distribution.
 - The research objective is to determine whether high scores on one continuous variable are associated with high scores on another continuous variable.
 - The research objective is to determine whether there is a definite relationship between categories of two continuous variables.
 - The research objective is to determine whether there is a definite relationship between two discrete variables.
25. Suppose three randomly selected instructional treatment groups achieve mean scores on an achievement test: (1) 6.08 (2) 5.83 (3) 9.52. The appropriate F-ratio is significant at the .05 level. From this information, it can be justifiably concluded that:
- Overall differences among the three sample means are statistically significant.
 - The sample mean for group 3 is statistically significantly greater than the sample mean for group 2.

- c. The sample mean for group 3 is statistically significantly greater than the sample mean for group 1.
 - d. All of the above.
26. If one group is randomly drawn from each of a group of normal populations having different means and equal variances, the expected value of the between-groups variance estimate will be _____ the population variance.
- a. Less than.
 - b. Greater than.
 - c. Equal to.
 - d. Greater than or equal to.

Research Procedures

27. If a sample is chosen from a population such that all possible samples have the same probability of being chosen the sample is called a _____ random sample.
- a. Simple.
 - b. Stratified.
 - c. Systematic.
 - d. Clustered.
28. Which of the following maxims is least likely to lead the researcher astray?
- a. The researcher should always assign subjects to treatment groups at random.
 - b. The researcher should always use a simple random procedure for selecting his subjects.
 - c. The researcher should always use as large a sample of subjects as administrative conditions permit.
 - d. The researcher should always sample from a homogeneous population.
29. If two observers can independently measure the same thing and agree on the number to be assigned to that thing, the measurement procedure is called:
- a. Objective.
 - b. Unbiased.
 - c. Stable.
 - d. Valid.

Descriptive Research

40. Which of the following is not a characteristic of a sound observation form used in a descriptive study?
- a. The items are based upon available theoretical and empirical knowledge.
 - b. The items are designed to test specific behavioral hypotheses.
 - c. The items require the observer to make a minimum number of inferences while observing.
 - d. The items make provision for recording every specific behavior that can occur.
41. Which of these aims of instruction in a French class is stated in behavioral terms?
- a. To grasp a message written in French by a pen pal.
 - b. To attain a mark of at least 80 per cent in a French spelling test.
 - c. To cover the first sixteen chapters of Jone's *Introduction to French*.
 - d. To appreciate the contributions of French culture.
42. Which one of the following statements is true of questionnaire techniques?
- a. Thorough pre-testing of the questionnaire is necessary for a major study, but usually adds little to a Master's thesis.

- b. Provision should be made for a follow-up procedure before the initial questionnaire is mailed.
 - c. A letter containing a detailed explanation of the origin and rationale for the study should accompany the questionnaire.
 - d. To avoid pressuring the subject no time limit should be set for returning the completed questionnaire.
43. Which of the following procedures for recording responses obtained from an interview is most likely to yield accurate, meaningful results?
- a. Writing a summary of each response as it occurs during the interview.
 - b. Writing a summary of each response after the interview has been completed.
 - c. Recording as many of the responses as possible into predetermined categories during the interview, with the remaining responses being recorded as supplementary notes.
 - d. Recording as many of the responses as possible into predetermined categories after the interview has been completed, with the remaining responses being recorded as supplementary notes.

Relational Research

44. Correlational research in which less expensive and less time-consuming measures are developed for measuring a particular variable is called _____ research.
- a. "Shotgun." b. Concurrent validity. c. Predictive. d. Selection
45. A research design in which subjects found with a given characteristic are compared with subjects found without that characteristic is called a _____ design.
- a. Situational. b. Experimental. c. Descriptive. d. Causal-comparative.
46. Usually the most important problem a researcher must face in conducting a causal-comparative study is:
- a. Defining appropriate populations of individuals.
 - b. Developing treatments to be administered to the subjects.
 - c. Training observers to record their observations accurately.
 - d. Constructing appropriate measuring instruments.
47. The major difference between correlational research and causal-comparative research is that:
- a. Correlational studies use correlation coefficients and causal-comparative studies do not.
 - b. Variables are manipulated by the experimenter in causal-comparative studies while variables are studied as they naturally occur in correlational studies.
 - c. Causal-comparative studies compare scores on one variable for two groups and correlational studies compare scores on two variables for one group.
 - d. Correlational research is much more time-consuming to conduct than causal-comparative research.

Experimental Research

48. Two groups of students are randomly selected from a homogeneous population, pre-tested and then randomly assigned to a control and treatment group. After the treatment has been

- completed, a post-test is administered to both groups. Which of the following competing variables could account for differences in mean scores for the two groups?
- History.
 - Test-treatment interaction.
 - Selection.
 - Maturation.
49. A statistical technique which permits one to test for significant differences among means that have been adjusted for initial differences on an uncontrolled variable is the:
- Randomized block design.
 - Analysis of covariance.
 - Simple linear regression.
 - Multiple linear regression.
50. A large homogeneous group of students is divided into two groups by random assignment. A highly reliable mathematics achievement test is administered, and the two group means are found to be equal. The two groups are each taught the same basic concepts of geometry by the same method. When a highly reliable geometry test is administered, after this treatment one group mean is substantially higher than the other. This difference should be attributed to:
- Group error.
 - Sampling error.
 - Measurement error.
 - Replication error.
51. Training method A is applied to high- and low-aptitude students, whose average final scores are 60 and 40 respectively. When training method B is applied to similar students, there is evidence for an interaction of aptitude and training if the high-aptitude students average _____ and the low-aptitude students average _____.
- 50 . . . 50
 - 40 . . . 20
 - 60 . . . 40
 - 70 . . . 50