

Comparing Instructional Methods: Some Basic Research Problems

GEORGE L. GEIS*

Professors interested in teaching are aware of numerous different "methods" of instruction: e.g., discussion, lecture, presentation by computer. It seems natural and straightforward to ask: Which one is best? Educational researchers, too, have posed this question and have repeatedly, over the years, attempted to answer it, comparing teaching methods, one with another. This paper is a discussion of such research and raises questions not only about the validity of the research but also about the question itself. It suggests that reformulating the question can lead to more productive research and to answers which will be more useful to the practitioner. Such a discussion might seem best directed to researchers.¹ But practitioner educators are consumers of applied research of this kind. The questions they ask and the degree of critical knowledge they demonstrate in judging answers can help raise the level of that research.

Comparison Research

Comparative studies are common in the literature which is addressed to educational practitioners. Extensive summaries of results periodically appear. Some compare variations within a particular method, for example "Learning in discussions: a résumé of authoritarian-democratic studies" (1). Others compare two or more methods (9) or compare the so-called traditional method with all others (5). In almost every case these surveys of many such studies present a picture of ambiguity or contradiction. This may seem to suggest the futility of pursuing such research, reinforcing the view that teaching is simply not amenable to objective study or at least that research will not prove profitable.

A sub-set of comparative studies is concerned with demonstrating the superiority of an innovative method. Typically a three-phase history characterizes this literature: 1) The originators of the new method show that it is dramatically effective. 2) Follow-up studies, including some by acolytes of the innovators, demonstrate less clear-cut superiority and suggest a more cautious approach. 3) Later studies fail to confirm the earlier data and the innovative treatment is rejected as another educational panacea that failed.

* Centre for Teaching and Learning Services, McGill University

¹ A recent example of examination of these problems but directed toward the researcher is: Shaver, James P. A verification of independent variables in teaching methods research. *Educational Researcher* 1983. Vol. 12, 10, 3-9.

Cynical rejection of research into teaching methods (as well as highly touted innovations) is understandable in the light of the high frequency of such results. One cause of the frequent invalidity of such research lies in the self-defeating nature of the questions being asked.

The question

Close examination of the question: "Is method *A* (e.g., lecture) better than method *B*? (e.g., discussion)" suggests several problems to the critical researcher or reader of research. The methods referred to must be completely and precisely defined. If they (technically, the independent variables) are not well defined there is no possibility of *replication* of the study and consequently it does not meet a primary criterion of research.

The word "better" provokes two more concerns. It vaguely suggests the phenomenon being affected (e.g., "student achievement scores are higher" would be one possible meaning of better). But somewhere this phenomenon (technically, the dependent variable) must be as clearly defined as the treatment in order to meet the criterion of replicability. Thus, loosely speaking, both the cause and effect must be well defined or the activity simply does not qualify as research. Nor does it provide the practitioner with information necessary for replicating the method and results. A second subsidiary matter is raised by the word "better." A host of dependent variables could be conjured up at this point. As suggested above, student achievement scores might be one, but equally appropriate might be "time to learn" or "cost of teaching" or "greater effectiveness with heterogeneous populations." While a change in one of these dependent variables may be seen as valuable to some, it would be valueless to others. For example, a treatment that markedly reduced costs might be termed "better" by a budget-conscious administrator, while one that reduced paper-grading work might be hailed as superior by an over-burdened instructor. "Better," then, not only needs to be defined in terms of specified outcomes, its appropriateness must also be re-evaluated once that definition is explicated.

Let us look closely at defining the two parts of the comparison question.

Defining the Method

Usually an educational method is defined in terms of obvious, formal characteristics. Thus, "Computer-Assisted Instruction (CAI)" refers apparently to any instruction delivered by a computer. "Lecture" seems to refer to a classroom format in which a teacher talks to (or at) the students. But we intuitively know that a formal description, say, of medication (a pill, an injection) can be almost irrelevant. One would not ask: "Are pills better than injections?" because we know that the answer depends upon such things as what is in each medication and what is wrong with the patient. Identifying and clustering instructional treatments in terms of one common trait such as physical appearance may prove to be equally simple-minded. If the effects of all pills were compared with the effects of all injections the results might well look like those reported in the

literature on comparing teaching methods. (We should note that at least there would be common agreement concerning what is a pill and what is an injection. Some educational “treatments,” for example “lecture-discussion,” will not readily produce such agreement among observers categorizing methods.)

Take the case of “Lecture” as a treatment category. After a moment’s reflection most of us would agree that the variety of lecture classes we have had as students (or taught as professors) is enormous. We can recall a well-organized, highly expert, humorous, warm lecturer whose presentations were spiced with everyday examples and analogies and who had the knack of unfolding the core ideas in each lecture as if he or she were a detective solving a mystery. We can also conjure up one barely audible, uncertain, painfully withdrawn lecturer who rambled through a thicket of “ers” and “ahs” to an ending that was enforced by the class-change bell ringing, a bell that perversely rang just as he or she seemed to be approaching, however obliquely, a major point. We would be loath to equate all actors as equal simply because they all perform in plays on stage before an audience. Superficially the format of presentation is the same but intuitively we know that what happens within the broad formal boundaries of the activity is what is critically important. Surely it seems preferable to define the method not in terms of superficial similarities but with reference to a set of features which may be present or not in many different “methods” (e.g., feedback to students, consideration of individual differences, pedagogically sound sequence and organization). This point will be elaborated below.

We should further note that since the treatment is so badly defined it is often impossible to carry out a critical step in this sort of research, namely observation to determine if all of the cases of a single treatment class represent in fact similar treatments: e.g., were all the “discussion classes” actually discussion? (Verifying that a prescribed treatment is indeed carried out is the focus of the growing and interesting research literature on what in medical research is called *compliance* (12).

Summary

The problems raised here are not trivial ones nor academic nit-picking. Almost every researcher who has reviewed a set of comparison studies echoes the words of Robert Hohn (8): “Inadequate description of the experimental techniques as well as control conditions, is perhaps the greatest deficiency in the recent literature on teaching innovation. A large majority of the thirty-one studies reviewed for this paper which compared more than one strategy of teaching provided incomplete information about both treatments employed. The typical procedure is to characterize a particular treatment with a label such as “lecture,” “traditional,” “self-paced” or “group” with little or no data or operational terms used to clarify what particular interaction was occurring within these groups.” (p. 3)

The Criterion Problem

We have already briefly looked at the problem of defining the second key word, “better,” in the comparison statement. Success may be greater student achievement, happier learners, wider applicability, impressed government sponsors, etc.

An adequate evaluation should probably consider *several* of these key variables. Practitioners are likely to suffer disillusionment when they buy into a method on the basis of evidence of increased instructional effectiveness only to discover skyrocketing costs. Some common variables besides achievement and cost might be student and teacher attitudes, development requirements (e.g., training teachers, producing materials), and implementation requirements (e.g., specially designed rooms, increased support staff).

Furthermore, we should look for precision in the definition of these variables. Suppose the reports of success of one method over another refer to higher levels of student achievement. Some such studies report merely changes in test scores or grades without describing the tests or the bases and procedures for grading. Teacher-made tests are *notoriously unsophisticated in terms of minimal standards* for valid psychometric instruments. Does the test cover the major curriculum areas? Do the testing situations correspond to those indicated in the objectives of the course (e.g., does the course aim at developing problem-solving behaviors and the tests require fact-recall?). Without a description of or a copy of the test materials and some indications of the soundness of the test we cannot judge the quality of the dependent variable.

The test results deserve similar careful scrutiny. Some studies have reported only gain scores on tests, e.g., the difference between a test score before and after instruction. Not only should we have information about the tests themselves as discussed above, we also should know the test scores. While an average gain of 10 points on a 100 point test may seem impressive and be statistically significant, it is hardly pedagogically satisfying if the gain represents the difference between a pre-test score of 10 and a post-test score of 20. Students have learned far less than half of what was presumably taught.

Following the suggestions for greater precision and clarity will lead to major revision of the question being asked. Method "A" and Method "B" would be operationally defined and the actual implementation of these methods would be confirmed. The impact upon learners would be spelled out in terms of *objective observations or measures*.

Even at this point major research problems remain which threaten the validity of any important conclusions.

"Generalizability"

A statement like "Treatment A is better than Treatment B" implies "always better." In fact there are many specific conditions that prevailed at the time of the research/evaluation. Some of them may have critically affected the results. Controlling for these factors, for example, the subject matter or the physical environment, is a major responsibility of the researcher. One likely candidate for a list of critical variables is the population of learners. Like other important variables it should be carefully and fully described. Generalization to another student population may well depend upon the similarity of the new population to the one in the research. The interaction of instruction with student character-

istics interests many educational researchers. Common sense would suggest that not all learners are alike and therefore for optimal effectiveness different treatments should be designed for different learners. One segment of this literature explores Aptitude-Treatment Interactions (3, 4). Others deal with such variables as time allowed for learning, rewards in instruction, reading level, learning objectives, and student preferences for different kinds of media (2, 6).

Research Strategy and Learning Model

This point of emphasizing the learner suggests another problem with methods-comparisons research. Implicit in all such research is the premise that the method of teaching is *the* critical variable in determining learning. Indeed, the research model suggests that teaching is a one-way process involving the direction of the subject matter to be learned toward the student, like water sprayed from a fire hose. Some techniques of presentation, infusion, or dissemination, are presumably, superior to others. Thus: the search for an optimal method.

What this approach overlooks, of course, is the point raised above, namely that other factors, which lie outside of the method (however precisely that method is defined) may be equally or even more important. In short, the methods-comparison research implies a commitment to a particular model of the teaching-learning process. That "one-way" model may not only produce conflicting and insignificant research results, it may also divert us as teachers from attending to the critical contributions of the learner (and other variables) in the teaching-learning process.

Purpose

The issue of how widely findings may be generalized to other situations is related to the question of the purpose of the comparison research. As a consumer I might ask these questions: Is this used car better than that one given my driving needs, my budget and my ability as a driver? The answer can be extremely useful to me but hardly casts much light on the superiority of one make of car over another.

Professors should be encouraged to carry out such mini-evaluations: Which of two textbooks works better with *my* class? But it is important to realize that the results of such an evaluation may be severely restricted. The matters discussed in this paper merely open the door to the difficulties of conducting research which is both valid and generalizable. If a study is to throw light upon the differential effectiveness of instructional methods, it must be designed with that purpose in mind. While there is nothing wrong with one homeowner recommending a type of water softener to another, the recommendation simply does not, in most instances, qualify as the result of careful research. If individual professors are interested in improving their own courses it is probably not worthwhile for them to become involved in either complex and sophisticated research in education nor to be overly concerned with the method and treatment questions. The best strategy might be to specify course goals and construct some means of determining how well the students are progressing toward those goals.

Then the professor might choose any method which has some degree of legitimacy and with which he or she feels comfortable. Professors, using broad guidelines for developing more effective instruction, can produce closer and closer approximations as they continue to make changes and observe the effects of those changes. This smacks of tinkering and it lacks elegance; the professor may not be able to contribute to the literature involving the comparison of treatments but may end up with a much improved course.

At another level of inquiry the search ought to continue for generalizations which go far beyond the specific question: Should I use Textbook A or B? This is primarily a task for educational researchers who should re-word the comparison question as it was stated at the outset of this article. Instead of comparing "methods" described in terms of formal properties, the contemporary researcher will try to isolate critical functions in teaching methods (e.g., feedback to students, structure or content) and study variations in these. And such work is going on in Psychology and Education.

Critical Features

In what has been called meta-analysis many studies are analyzed in an attempt to factor out, *a posteriori*, what were critical features. To put it another way it involves retrospectively defining independent variables. A good example of this research strategy is the work of the Kuliks (11). Careful examination was made of a sizeable set of studies which involved the Personalized System of Instruction (PSI or The Keller Plan). Instances of success and failure of the treatment were "sorted" and then the two collections were analyzed to reveal differences within the grossly defined independent variable (i.e., the PSI treatment) which would explain the difference in effectiveness. The task was made easier in the case of PSI since the method has been carefully and clearly defined by Keller and Sherman (10). Kulik, et al. were able to determine which components of PSI are critical to its success. Interestingly they turn out to be features commonly emphasized by many instructional psychologists and educational developers (e.g., small unit size, mastery, immediate feedback). This consistency with lore and theory is encouraging since it suggests that even relatively crude research in instruction if properly formulated may reveal, or at least suggest, some powerful variables. Note that the critically important variables are not unique to PSI; they can be incorporated into other teaching formats.

This suggests two things: that a few features may be especially important in *any* form of instruction if it is to prove to be effective and that the contradictory results of comparative methods research may, in part, be due to the fact that elements which are critical to effective instruction are present or absent in particular instances of instruction regardless of "method." Thus, clarity of organization or effective monitoring of student progress can be found in some computer-based sequences and not others, in some discussion classes and not others, etc.

Other research strategies besides meta-evaluation could be cited. For example,

intensive parametric studies involving one or two variables such as time to learn, or the control over responses to text; naturalistic studies which examine real learning environments over a period of time; studies which attempt to make explicit the “cognitive processes” of the learner.

In all of this research the attempt is to isolate and to look closely at key factors affecting learning in contrast to traditional methods comparisons.

CONCLUSION

This discussion has suggested that care be taken in examining studies which compare one method of teaching to another. It has stressed the need for precise identification of the critical variables in instructional methods — variables which cut across traditionally defined “treatments.” It has described some research which has examined methods in order to reveal such attributes. Further it has urged that a similar degree of precision be demanded in the definition of the effect (e.g., “greater student achievement”). In addition the critical contribution of variables other than “method” should be examined, an example being the interaction of student variables and treatment.

The case of method-comparison research illustrates a more general principle: The consumer of research — the educational practitioner — can and should exert controls on that research to make it better and more useful. It will improve to the extent that these consumers are informed and sophisticated about problems such as those discussed here.

In the last decade or so there have been notable advances in applied educational research. In the gap between research and application — between the laboratory and the classroom — there is growing up an area of applied study which is likely to develop more and better means by which we can define our questions and find ways of answering them, (e.g., 7). We can encourage or discourage that development by our stance as consumers.

BIBLIOGRAPHY

1. Anderson, R.C. Learning in discussions: a résumé of the authoritarian-democratic studies. *Harvard Educational Review* 1959, 29, 201-215.
2. Bloom, B.S. *Human Characteristics and School Learning*. New York: McGraw-Hill, 1976.
3. Bracht, G.H. Experimental factors related to aptitude-treatment interactions. *Review of Education Research*, 1970, 40, 627-645.
4. Cronbach, L.J., & Snow, R.S. *Aptitudes and Instructional Methods: a Handbook for Research on Interactions*. New York: Irvington Publishers, 1977.
5. Dubin, R.Y., & Taveggia, T.C. *The Teaching-Learning Paradox: a Comparative Analysis of College Teaching Methods*. Eugene: Center for the Advanced Study of Educational Administration, University of Oregon, 1968.
6. Gagné, R. *The Conditions of Learning*. New York: Holt, Rinehart and Winston, 1977.
7. Glaser, R. *Adaptive Education: Individual Diversity and Learning*. Holt, Rinehart and Winston, New York: 1977.

8. Hohn, R.L. Effectiveness of innovations in the teaching of psychology: a critique. *American Psychological Association Newsletter, Division on the Teaching on Psychology*, December 1973, 3-6.
9. Jamison, D., Suppes, P., & Wells, S. The effectiveness of alternative instructional media: a survey. *Review of Educational Research*, Winter, 1974, 44, No. 1, 1-67.
10. Keller, F.S., & Sherman, J.G. *PSI, the Keller Plan handbook: Essays on a Personalized System of Instruction*. Menlo Park, California: W.A. Benjamin, 1974.
11. Kulik, J.A., & Kulik, C.-L. C. Effectiveness of the personalized system of instruction. *Engineering Education*, 66 (2) December 1975, 228-231.
12. Leonard, W.H., & Lowery, L.F. Was there really an experiment? *Educational Researcher*, 8 (6), June 1979, 4-7.