NEW DIRECTIONS FOR COMMUNICATION RESEARCH

● L. P. GREENHILL

This paper makes suggestions for the improvement of communication research and discusses some new directions such research might take.

L. P. Greenhill is associate director, Division of Academic Research and Services, Pennsylvania State University.

THE WORD *communication* has many aspects.

For the purpose of this presentation I propose to interpret the word *communication* to mean the presentation of auditory and visual stimuli to students in such a way as to encourage them to acquire new information, skills, and attitudes of an academic type. Specifically, I shall be referring principally to the use of television and related techniques for the instruction of college students.

In discussing new directions for research in this area I shall divide my comments into two parts. The first part will deal with suggestions for the improvement of the kinds of research that are currently being undertaken; the second will deal with some new areas and approaches to communications research.

IMPROVED METHODOLOGY

Experimental Designs

Let us turn first to a consideration of the improvement of research methodology—especially the improvement of experimental designs.

Need for randomization. Probably few innovations in education have been subjected to as extensive evaluation as has the use of television. By and large, the experimentation has been of high caliber and of relatively rigorous design as compared with earlier educational experiments. However, in looking over much of the re-

search and in attempting to interpret and evaluate it one cannot help but notice the fact that one of the basic requirements for applying a test of significance has been omitted from many experiments —in other words, the subjects have not been randomly assigned to the several treatments.

Administrative considerations often make it inconvenient to use randomization, and, accordingly, intact groups are used and an attempt is made to "equate" these groups by selecting from them subjects who are equivalent in terms of measures of ability, age, sex, socioeconomic status, or whatever else may appear to be important. These efforts to "equate" self-selected groups are rarely successful and one can never be sure that he has compensated for all the important factors that might bias an experiment by using this method. This process of "equating" groups is even more tenuous when comparisons are made across different schools—that is, when students in School 1 are taught by Method A and those in School 2 by Method B. (We were caught in this trap once, and even after equating for age, sex, and intelligence, the experiment still yielded inconsistent differences which defied interpretation. One conclusion was that the several schools varied greatly in morale—a factor which certainly interacted with the experimental variables.)

One encounters similar problems when attempting to compare different treatments with two intact groups of college students. One finds that students in one section of a course may differ considerably in curriculum background from students in another section of the same course, and that these curriculum differences often reflect different abilities which can seriously upset experimental comparisons.

What I have been trying to do is to stress the imperative need for random assignment of students to experimental treatments in order that any biases affecting the experiment will have an equal opportunity of affecting each method. One is then justified in using a test of significance to ascertain the level of confidence that one can have in the results of the experiment.

Randomization presents little difficulty when the several experimental treatments are given at the same hour, but when the same teacher has to teach by two different methods this usually has to occur at two different hours and this presents a more difficult problem for random assignment. At Penn State we have worked out a technique for randomizing students across class sequences which may be useful to others. Prior to the experiment and during registration for the course, we examine each student's schedule in

order to determine which students can be assigned to either of the two sequences to be taught by the experimental methods. If a student has both sequences open, he is randomly assigned to one of them and is regarded as being in the experiment. Students who do not have both sequences open take the one that they wish, but their scores are not included in the experimental comparison.

To sum up this section, I would like to *argue against:*

- Applying tests of significance to groups which have not been randomly assigned.
- Making comparisons between two different schools when Method A is taught in School 1 and Method B in School 2.
- Making comparisons across two school years or semesters when Method A is taught one year and Method B the next.

Of course, comparisons of the last two types may be made and the results presented in the form of descriptive data without the use of tests of significance, but interpretation is usually difficult.

Where the requirement of randomization has been met, the precision of an experimental comparison can often be increased if one or more tests of ability are administered before the experimental treatments are begun. Such tests can then be used as adjusting variables in an analysis of covariance. However, analysis of covariance is not a substitute for randomization.

Need for more complex designs. Let us now consider a different aspect of methodology. Many of us have given thought to the parade of nonsignificant differences that are emerging from current communication experiments. Nonsignificant differences are particularly typical of the results of comparisons between televised and face-to-face instruction where the same teacher has taught by each method. This finding can, of course, be taken as an indication that television can be successfully used to penetrate the quantity and distance barriers to education, because in most of the comparisons larger numbers have been taught over television than in the comparable face-to-face groups, and those participating in televised instruction have often been located at a distance from their teacher.

However, what we are also looking for is a breakthrough in the *quality* of instruction, and this failure to find significant differences may be a function of the types of experimental designs we are using. In general, these designs involve comparisons between single variables, e.g., TV vs. face-to-face instruction (teacher held constant), large group vs. small group, feedback vs. no feedback, et

cetera. It is quite possible that the effects of most single variables are too small to have any marked influence on learning. In fact, I have had a personal theory for some time that the better the control in single variable experiments, the less likely is one to find significant differences, and that the differences which were often found in early research were frequently the result of inadequate experimental controls!

In the instances where differences have been found in televised instruction, they have usually been between groups taught over television by a carefully selected teacher, and groups taught by other teachers in face-to-face situations. Here we have at least two variables operating and it is impossible to say which is producing the effect.

In such circumstances it would be possible to go to a factorial design in which teacher A teaches by TV and face to face and teacher B does likewise. In this factorial design with two factors and four groups, it would be possible to tell which factor was producing what effect, and whether there is any interaction between the factors. It is therefore conceivable that such relatively complex designs might be used and would yield worthwhile information.

The thought has occurred to my colleagues and myself, however, that what we need is a design in which comparisons would be made between methods involving several variables that are additive in their effects. We have evidence, for example, that repetition of the main ideas in a lesson increases learning, and that certain incentives such as competition increase learning.

What we need then is a design in which patterns of these "positive" variables would be compared. Such a design might be called an "additive" design. In this design the control group would receive the conventional treatment (C). One experimental group would receive this treatment plus, say, some form of repetition (R); a second experimental group would receive repetition plus knowledge of results (K); a third group would receive both (R) and (K), plus some form of social reinforcement (S) in peer group discussions. Such a design would look like this:

TREATMENT	VARIABLES
Control	*Conventional Method* (C)
Experimental 1	C+R
Experimental 2	C+R+K
Experimental 3	C+R+K+S
and so forth	

In such an additive design the preceding treatment would serve as a control for each succeeding one, and the original control group would serve as a base line for comparison of the accumulating patterns of variables.

However, when four or five treatments of this type are in progress, there may be difficulties in having the same teacher teach by each method. At this point television or film recordings can play their part. With closed-circuit television, students in different rooms can be taught by the same teacher at the same time but can be given other differential treatments. Similarly, instruction can be recorded on film and used as a constant stimulus source for several groups whose treatment is varied in other respects.

New Criteria

Let us now turn from the question of designs to the problem of the criteria or tests to be used to evaluate the outcomes of communication research.

One of the criticisms of television research most often heard from the advocates of television is that the tests used to evaluate the televised presentations have been primarily verbal in form and did not therefore adequately test the visualization of concepts that television is so well equipped to present. Undoubtedly there is considerable truth in this point of view. Studies have shown that the effectiveness of a particular method of teaching is to some extent a function of the method of testing. To this extent the use of verbal tests may be biasing the results in favor of a verbal method of presentation.

Television offers unusual opportunities for the presentation of situational tests with pictorial and sound cues that closely approach real-life situations. Thus, lifelike problems can be presented for students' responses in a way that would be almost impossible in the classroom. In fact, this procedure can be carried a stage further and the test problems can be recorded on film for standardized presentation and re-use.

I do not have in mind here the presentation of verbal test items via television although this has been successfully accomplished with great savings in time and the printing of tests. Rather the problems would be more in the nature of laboratory demonstrations from which students would have to deduce the correct solutions. In 1953 The Instructional Film Research Program at Penn State successfully produced such a test on motion picture film for the performance testing of tank mechanics. The test had unusually high reli-

ability. Some beginnings had been made in performance testing via television at Penn State in the courses of Metallurgy, Archeology, and Air Science.

One always has to answer the question whether the performance measured by such test problems is a valid objective of the course of study. Some people claim that the function of most university courses is to develop ability to think in terms of abstract principles, and that verbal tests are the best measure of this particular kind of performance.

The critics of television research frequently assert that the tests used to evaluate the comparative effectiveness of televised and face-to-face instruction are inadequate and test only for recognition of facts. What we really need to do they say is to test for the "intangibles." Pressed to define and defend their view, these critics will usually indicate that by "intangibles" they mean problem-solving behavior, creative thinking, the formation of desirable attitudes, and the like.

Actually, television researchers have gone far beyond most other educational researchers in testing for these types of learning. At Penn State researchers have certainly attempted to build test items which assess a student's ability to apply principles to the solution of problems, or test his ability to interpret data and draw valid conclusions. Penn State also has used such attitudinal tests as the Inventory of Beliefs, the *F* Scale, and the like. Similar special tests have been used by others—particularly at Miami University in Ohio.

On the other hand, it is probably fair to admit that more work is needed on the development of appropriate criteria for educational research, and that we should perhaps think in terms of using multiple criteria for each experiment. My colleagues and I at Penn State have been discussing this problem recently, and a list of criteria for assessing academic achievement has been proposed in order to assess a broader range of educational objectives. Such a list of criteria is as follows:

Criterion	*Assessment Function*
1. Objective multiple-choice items	Recognition of facts, solving problems
2. Essay questions	Integrative abilities
3. Oral examination	Personality traits
4. Term paper	Library research, integration
5. Interview	Academic attitudes and interest
6. Research project	Creativity
7. Laboratory performance	Manipulation skills
8. Election of future courses	Performance over time in future courses

Some of these criteria may be difficult to quantify reliably but they are suggestive of the direction in which evaluation techniques might move.

NEW AREAS FOR RESEARCH

Let me now turn from the subject of improved methodology to a consideration of some new areas for research in the auditory and visual methods of communication with special reference to televised instruction.

New Areas in TV Research

It appears to me that much of the research now under way is repetitious. That is, the same kinds of comparisons are being made over and over again in different places with the same subject matter, and this is especially true of the comparison of televised with face-to-face instruction. Admittedly, some research of this type is necessary for public relations purposes. Many people (especially teachers) just cannot accept the idea that other teachers could possibly be as effective over television as in a face-to-face situation. Also, every once in a while a new subject matter is tried out on television and the effectiveness of this procedure needs to be evaluated to be sure that the use of television is not adversely affecting learning. Such research I refer to as "quality control" research.

I would think that there should be progressively fewer of these TV—face-to-face comparisons, and that in place of such studies we could begin to concentrate on developing more effective ways of using television and to evaluate the relative effectiveness of these different ways.

We might, for example, investigate more thoroughly certain production variables. There is much debate between educators and TV producers as to what makes for effective televised instruction. We need to know what production techniques actually enhance learning, and what ones have no effect or even interfere with it.

We might compare various ways of encouraging overt participation on the part of students during a televised presentation, or we might compare overt with covert participation. If we bear in mind the fact that we are not only teaching *subject matter* but that we are also teaching *students,* many such variables come to mind. For example, we could compare the effectiveness of various kinds of supplementary activities—written themes, discussion sessions, additional reading, and the like. We could compare methods of teaching

students how to learn from televised presentations. At Penn State we are currently conducting a project in which homogeneous grouping of students is being compared with heterogeneous grouping in TV classrooms.

We need to investigate the applicability to TV of some of the variables studied in film research: e.g., black and white vs. color, camera angle (objective vs. subjective), the use of various kinds of repetition, and different rates of development.

We certainly need to get into the delicate area of comparing various instructors on television in order to identify the most effective television teachers or combinations of teachers (teaching teams) and to learn what makes some teachers more effective than others, or whether a team is more effective than a single teacher on television.

We need to find out why there is resistance on the part of faculties to the use of television.

We certainly need to compare recorded presentations (on film or video tape) with live presentations. If research shows them to be effective, the recording of instruction offers enormous possibilities for re-use, and for interchange of high-caliber instructors between institutions and even countries.

Need for Basic Research

However, in many of these comparisons I fear that we may continue to find few significant differences. In this connection, we at Penn State have developed an hypothesis. It is that students in general do not study to learn as much as possible. They study to pass a course or to get a particular grade. Thus, if the teaching in a given course is appreciably improved in some way or another, the students may do less work in that particular course, and concentrate on other courses that are not being taught as well. Thus, this law of "compensatory effort" may make it very difficult to secure differences in effectiveness between varying instructional methods.

This hypothesis, if true, suggests the need for even more basic research on methods of motivating students so that they will tend to put more effort into their academic work. At Penn State we are currently conducting a project which involves regular meetings of students in work-study groups. These activities are intended as supplements to large TV classes. In these groups students not only discuss course-related subjects, do library research, and the like, but they also discuss their roles and responsibilities as students in

a university and study problems of higher education in general. Such activities are intended to influence students' academic motivations by social reinforcement with the object of making scholarship a more respectable and desirable pursuit.

Along this line of basic educational research, studies might be conducted in the area of interactions—interactions between student and student, student and teacher, and student and subject matter. The various reward systems for academic work might be studied and new and effective incentives for academic work developed. In addition, studies of the programming of instruction might yield worthwhile results: studies of order of topics within a lesson, the ordering of lessons in a course, and the ordering of courses in a curriculum.

Finally, some really basic research needs to be conducted which would attempt to develop information about how communication occurs through auditory (verbal and nonverbal) and visual (verbal and nonverbal) methods of communication and what kinds of meaning are communicated in various situations to different types of people by these several modes.