DOCUMENT RESUME

ED 053 715 JC 710 214

AUTHOR	MacMillan, Thomas F.
TITLE	Experimental Strategies for the Evaluation of Instruction.
PUB DATE	Aug 71
NOTE	26p.; Paper presented at the Danforth-UCLA Economics-Economic Education Institute held at University of California, Los Angeles, August 24-25, 1971
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Educational Research, Evaluation Methods, Experimental Groups, *Experiments, *Instruction, *Junior Colleges, Methods Research, Research, *Research Methodology, Scientific Research

ABSTRACT

Experimental designs address themselves to two concerns: providing answers to research questions and controlling variance. The latter involves threats to internal and external validity. The threat to internal validity constitutes the greatest difficulty because of maturation, history, experimental mortality, and differential selection. Examples of eight basic research design paradigms are made, pointing to the relative strengths of design in maximizing experimental variance and minimizing extraneous and error variance. Of most concern is whether a particular instructional strategy is resulting in measureable results on the part of students. The following are elements of a successful research proposal: (1) the development of a clear statement of the research problem, including an explanation of variate and criterion variables in the study, type of relationship between the variables, and the target population; (2) justification for the research approach; and (3) development of a statement of operational research objectives and/or hypothesis. (CA)

ED053715

EXPERIMENTAL STRATEGIES FOR THE EVALUATION OF INSTRUCTION

A Presentation to the Danforth-UCLA

Economics-Economic Education Institute
held at University of California, Los Angeles

August 24,25, 1971

Thomas F. MacMillan

Chairman, CJCA
Committee on Research and Development

About five years ago, Roger Kaufman re-wrote the fable of the three little pigs to illustrate some points about developing systems for instructional development and evaluation. His version of the story went like this:

One upon a time there were two pigs (a third one had gone into marketing and disappeared) who were faced with the problem of protecting themselves from a wolf.

One pig was an old timer in the wolf-fending business, and he saw the problem right away - just build a house strong enough to resist the huffing and puffing he had experienced before. So, the first pig built his wolf-resistant house right away out of genuine lath and plaster.

The second pig was green at this wolf business, but he was thoughtful. He decided that he would analyze the wolf problem a bit. He sat down and drew up a matrix (which, of course, is pig latin for a big blank sheet of paper) and listed the problem, analyzed the problem into components and possiblities of wolf strategies, listed the design objectives of his wolf-proof house, determined the functions that his fortress should perform, designed and built his house, and waited to see how well it worked. (He had to be an empiricist, for he had never been huffed and puffed at before.)

All this time, the old-time pig was laughing at the planner pig and vehemently declined to enter into this kind of folly. He had built wolf-proof houses before, and he had lived and prospered, hadn't he? He said to the planner pig, "If you know what you are doing, you don't have to go through all of that jazz." And with this, he went fishing, or rooting, or whatever it is that pigs do in their idle hours.

The second pig worked his system anyway, and designed for predicted contingencies.

One day the mean old wolf passed by the two houses (they both looked the same - after all, a house is just a house). He thought that a pig dinner was just what he wanted. He walked up to the first pig's house and uttered a warning to the old-timer, which was roundly rejected, as usual. With this, the wolf, instead of huffing and puffing, pulled out a sledge hammer, knocked the door down, and ate the old-timer for dinner.

Still not satiated, the wolf walked to the planner pig's house and repeated his act. Suddenly, a trap door in front of the house opened and the wolf dropped neatly into a deep, dark pit, never to be heard from again.

Morals:
1. They are not making wolves like they used to.
2. It's hard to teach old pigs new tricks.
3. If you want to keep the wolf away from your door, you'd better plan ahead. (1)

Although all three of these morals could be applied to almost any group of educators who are concerned with the improvement and assessment of instructional strategies, it is perhaps to the last moral that we should pay the most heed. If you want to keep the wolf of faulty conclusions or poor research away from your door, then you'd better plan to spend some time thinking about experimental design.

Why should we spend any time at all on research design? After all, most instructional settings are clearly not experimental laboratories, and most faculty members are not particularly interested in generalizing beyond the students with whom they come in immediate contact. Further, like so much of education-ese these days, insisting on experimental design seems to so many of us to be unnecessary because it is doomed to failure before we begin, or, what is worse, even if we get any results from experimentation, the net effect of the results will be simply to belabor the obvious. For an experienced teacher, the "seat of the pants" evaluation of the success of a particular unit or lecture serves to inform him of his strengths and weaknesses, and of the level of mastery enjoyed by his students. Further, there seems to be no reason to be sanguine about the consequences of one instructional method over another, when viewed under experimental conditions. Dubin and Taveggia reviewed 91 studies of instructional method and reported that "we are able to state decisively that no particular method of college instruction is measurably to be preferred over another when evaluated by student examination performances." (2)

One answer to the question of why spend the time on research design is found in McKeachie's more recent review of research on college teaching. McKeachie's real contribution is in the area of experimental method, for while he acknowledges the conclusions of Dubin and Taveggia, he reminds his reader that their study "deals only with the effects of teaching on course examinations." McKeachie continues, "When one asks, however, whether knowledge (1) is remembered after the final examination, (2) can be applied to new problems, or (3) is related to attitudes and motives, we find that class size and teaching method do make a difference." (3) The point is really this: if we are indeed interested in observing and evaluating the consequences of the instructional process, then some real attention to, and understanding of, the best ways to assess those consequences can be of critical importance.

The Purpose of Research Design

One of the best and most direct statements of the purposes of research design is this: "research design has two basic purposes: (1) to provide answers to research questions and (2) to control variance." (4) In Fred Kerlinger's strategic view of research design, the whole point is: "Maximize systematic variance, control extraneous variance, and minimize error variance." To do this, the researcher attempts "(1) to maximize the variance of the variable or variables of his substantive research hypothesis, (2) to control the variance of extraneous or 'unwanted' variables that may have an effect on his experimental outcomes, but in which he is not interested, and (3) to minimize the error or random variance, including so-called errors of measurement." (5)

Let's look at the issue of controlling variance more closely. For the sake of illustration, assume that you are interested in the effect of structuring an Introduction to Economics course in the following way: one large lecture session per week, using appropriate instructional media and reinforcement techniques (distributing printed lecture notes, administering spot quizzes, etc.); one small group discussion per week in a seminar group of not more than 25 students; one hour of independent study using audio tapes and audio-tutorial materials in an independent study lab. Let's further suppose that two sections of the class are offered, and that you decide to compare "traditional" methods with the "non-traditional" methods by observing the final exam performance of both groups of students. On the basis of this experiment you are going to restructure the entire Economics Department. What are some of the problems that may affect your experiment in ways that will interfere with your attempt to maximize systematic variance and minimize error variance?

Threats to Internal and External Validity

Campbell and Stanley (6) have summarized the kinds of problems that may affect an experiment under the two broad headings of Internal Validity and External Validity. Internal validity is concerned with the question "Did the experimental treatments in fact make a difference in this specific case?" Campbell and Stanley have listed a number of specific extraneous variables that may pose a threat to both internal and external validity. On the following page, these are given, with brief discussion on the nature of each threat to experimental conclusions.

THREATS TO INTERNAL AND EXTERNAL VALIDITY
(Campbell and Stanley, Experimental and Quasi-Experimental Designs for Research, Chicago, Rand, McNally & Co., 1963, pp. 5-6)

I. Threats to Internal Validity (Was it the experimental treatment that made the difference?)

A. History - the specific events occurring between the first and second measurement in addition to the experimental variable.

B. Maturation - processes within the respondents operating as a function of the passage of time per se (not specific to the particular events), including growing older, growing hungrier, growing more tired, and the like.

C. Testing - the effects of taking a test upon the scores of a second testing.

D. Instrumentation - in which changes in the calibration of a measuring instrument or changes in the observers or scorers used may produce changes in the obtained measurement.

E. Statistical Regression - operating where groups have been selected on the basis of their extreme scores.

F. Differential Selection - biases resulting in differential selection or respondents for the comparison groups.

G. Experimental Mortality - or differential loss of respondents from the comparison groups.

H. Interaction Effects - effects due to interaction in multiple-group designs which might be confounded with the effect of the experimental variable.

II. Threats to External Validity (Can we generalize the results of this experiment?)

A. Reactive Effect of Testing - the use of a pre-test might increase or decrease the respondent's sensitivity or responsiveness to the experimental variable.

B. Reactive Effect of the Experimental Arrangements - which would preclude generalization about the effect of the experimental variable upon persons in non-experimental settings.

For the purposes of educational experimentation, it is perhaps the threats to internal validity which constitute the greatest difficulty. We are particularly prone to threats due to maturation, history, experimental mortality, and differential selection. Thus, for example, in the middle of a cooperative research project which was intended to develop a means of predicting individual student attrition in 22 cooperating institutions in Northern California, the passage of Senate Bill 164 allowed the participating colleges to develop and fund major new resources and services to retain exactly the kinds of students that would have been likeliest to withdraw. (7) In that case, history caught with us, and we really couldn't tell whether the model was any good or not, at least not in any experimental sense. In another instance, there had been a long history of investigating the "impact of college" on students, but until Plant and Telford (8) included a control group of students who graduated from high school but did not enter the community college, no one had experimentally asked the question whether the "changes" might also be due to maturation. It turns out, of course, that some of the value attributed to the college experience could also have been attributed to simple maturation, but it took a well-designed experiment to yield the results which proved it. At the classroom level, differential selection and experimental mortality are consistent problems. If we look at the performance of an eight o'clock Tuesday-Thursday section of Economics 1A and compare it with the performance of a ten o'clock Monday-Wednesday-Friday section, the differences are as likely to result from the selection process which placed the two groups of students into those sections as from any experimental treatment we may introduce.

Kerlinger has suggested five ways to control some of the extraneous variables that may serve as a contaminant to experimental findings. The specific suggestions are:

1. To eliminate the effect of a possible influential independent variable on a dependent variable, one can choose subjects so that they are as homogenous as possible on that independent variable.

If we are worried about academic aptitude as a variable in studies of achievement in college classes, for example, one way to control the effect of the variable is to select subjects from a limited range of scores on an academic aptitude test.

2. Whenever it is possible to do so, randomly assign subjects to experimental groups and conditions and randomly assign conditions and other factors to experimental groups.

Kerlinger stresses that "theoretically, randomization is the only method of controlling all possible extraneous variables." If randomization has been accomplished, then these extraneous variables can be assumed to operate with equal strength in all groups, and they can be considered statistically equal.

3. An extraneous variable can be controlled by building it into the experimental design as an assigned variable, thus achieving control and and yielding additional research information about the possible effect of the variable on the dependent variable and about its possible interaction with other variables.

In the study of potential dropout students, it was found that ability clustered by sex was a more potent predictor of attrition than either variable alone: low ability females were about three times likelier to stay in college to complete the semester than low ability males. The findings were further elaborated when race was added as a variable: Black students were liklier to withdraw than any other race. (9) The point is that if we are conducting an experiment in which attrition is one of the dependent variables, we ought to be aware of the influence of race, ability, and sex on attrition, and perhaps include these variables among the independent variables in the design.

4. When a matching variable is substantially correlated with the dependent variable, matching as a form of variance control is profitable and desirable.
5. Before using matching, however, carefully weigh its advantages and disadvantages in the particular research situation. Complete randomization or the analysis of covariance may be better methods of variance control.

The problem of matching is twofold: the variable must be substantially related to the dependent variable, and, when we try to match on more than two variables, we inevitably lose experimental subjects. Imagine the problem of trying to conduct an experiment in two classes matched for race, academic aptitude, and sex. The problem wouldn't be in conducting the experiment, but in the selection and matching process.

Some Experimental Designs

With the background of understanding some of the problems that may arise in experimental design, we may now turn to some paradigms of research design, pointing out in each case the relative strengths of the design in maximizing experimental variance and minimizing extraneous and error variance. There are, of course, many texts available in the field, and the particular two which have seemed most useful to many educational researchers are Fred Kerlinger's Foundations of Behavioral Research (Holt, Rinehart and Winston, 1964) and Donald Campbell and Julian Stanley's Experimental and Quasi-Experimental Designs for Research (Rand McNally, 1963). To both of these sources one would be well advised to turn for a more detailed and clear explanation of the designs that will be considered below. But the purpose of this presentation is to introduce a limited range of possibilities, not the entire universe of research design. For most instructional research, the possibilities really seem to reduce themselves to about eight basic research design paradigms which will be shown below. For each, the convention that "0" means "observation" (testing, classifying, categorizing of experimental subjects) and "X" means "treatment" (experimental manipulation, monkeying around with the instructional process) will be followed. "R" will mean "random assignment" to the experimental or control condition.

Design I: The One Shot Affair

Paradigm: X 0_1

In this design, we introduce something experimentally, and then we measure what has happened to the subjects. Thus, for example, we might decide to do our "non-traditional thing" with large lecture sections, audio-tutorial lab assignments and weekly seminars, and then we might test the students with a final examination to find out how well they learned.

Suppose we learned that the exam scores for the group were higher than any we had observed in the past five years? Would you then be willing to throw out all "traditional" instruction? What about asking some questions first: (1) how do you know that this group of students wasn't just "brighter" than any class in the in the last five years? (2) how do you know whether something like a Presidential announcement of price and wage freezes didn't suddenly create a burning interest on the part of these students to understand more about Economics 1A? (3) How do you know whether or not somebody distributed copies of that same old, yellowed final examination you have been administering for the last twenty years? (4) How do you know (or was it a dependent variable you were interested in) whether the attrition in your class hasn't just skimmed off the lower ability students, leaving you the best students to take the final examinations? Fred Kerlinger minces no words about this one: "scientifically speaking," he says, this design "is worthless; worse, it can be misleading." (10) Day to day, we make a lot of non-scientific decisions, so merely to say that the design has no scientific worth is not to condemn it entirely. But, as we all know, the experienced teacher can intuitively evaluate his own teaching performance with any given class "by the seat of his pants." And furthermore, he is usually right. Everyone has experienced classes which seem to come together as if by magic, in which every lecture and every discussion seems almost to have been choreographed in the cosmos. But if we are evaluating a new method or a new technique or new media with an eye to their adoption, we may want to be sure that the student performance gains we observe are attributable to something more tangible than cosmic choreography. Like so many one shot affairs, this one might leave us disappointed if we tried to recapture it again next semester.

Design II: One Group Pre-Post

Paradigm: $O_1 \quad X \quad O_2$

This design is commonly used in experiments calling for an assessment of attitude or achievement changes as a result of some experimental treatment. Thus, for example, we might be interested in knowing whether attitudes about race, as reflected by performance on the Steckler anti-Negro scale (11), were affected by a six week unit on the nature and roots of prejudice in one of our Ethnic Studies courses. Perhaps the most serious problem with the One group-Pre-Post design is sensitization to the testing process. If, for example, we administer the Steckler scale during the first week of instruction in one of our classes, and then begin formal lectures and discussions on the issues of prejudice in the United States, students are very likely to perform in a sensitized fashion when we administer the test again. Campbell calls measures such as the Steckler scale "reactive measures" because they causesubjects to react, and further to remember their reaction in a similar testing situation. As in the case of Design I, history and maturation are likely to affect the subjects' performance on either test. As Kerlinger summarizes the problem with this design, it is inadequate "not so much because these extraneous variables can operate (they operate whenever there is a time interval between pre-test and post-test), but because we do not know whether or not they have operated, whether or not they have affected the dependent variable measures.

Design III: Two Group Static Comparison

Paradigm:

$$\frac{X \qquad O_1}{\qquad\quad O_1}$$

Two groups are established in this design, one of which receives the special treatment while the other does not. At the end of the experimental period, both groups are observed, and a comparison is made between the two observations. This approach represents a major gain over the first two, since a non-treated group is added to the experiment for comparison purposes. The design could be materially strengthened if the two groups were compared on different independent variables to ascertain their comparability: sex, academic aptitude, age, and other comparison variables might be compared for this purpose. Because the design does not include random assignment to the "treatment" or "control" conditions, there is no certainty that the groups are statistically equal to begin with.

The researcher could, however, randomly assign a treatment condition to one of the two groups, say, by the flip of a coin. The central issue in this design is still how best to be assured that differential selection has not occurred. The relative strength of the design is that it provides a comparison group which, with proper assurances as to comparability, can serve to mediate the effects of history and maturation across the treated and non-treated samples. Two new concerns are introduced in this design which have not been problems in one having only the students with higher academic aptitudes remaining at the end of the semester; further, there may be a reactive effect (a "Hawthorne" effect) if the experimental group is demonstrably and dramatically "treated" differently. For its shortcomings, Design III is the most promising of the three discussed so far. Provided that assurances are possible as to the comparability of the samples, reasonable conclusions can be made about the impact of a certain instructional method or approach on student performance.

Design IV: Experimental Group-Control Group: Randomized Assignment

Paradigm: R $\quad \dfrac{X \quad 0_1}{\quad\ 0_1}$

Design IV introduced the new element of Random Assignment (capital R, capital A) to the experimental design. Remember that the only way one can be assured of the statistical comparability of two samples is to assign membership to those samples on the basis of random assignment. If random assignment is accomplished, then almost all of the threats to Internal Validity can be met: history, maturation, statistical regression, differential selection, experimental mortality will all be assumed to act by chance with equal impact on the two groups if random assignment has taken place. Kerlinger has summarized his conviction that this is the "ideal design of educational and social research" in four points: (1) it has the best built-in theoretical control system of any other design; (2) it is flexible, being theoretically capable of extension to any number of groups with any number of variables; (3) if extended to the multivariate case, it can

ok

test several hypotheses at one time; and (4) it is statistically and probabilistically elegant. (12)

Design V: Experimental and Control Group, Pre-test - Post-test

Paradigm:

R	0_1	X	0_2	(experimental)
	0_1		0_2	(control)

Subjects are first randomly assigned to experimental or control groups, a pre-test is taken, one group then receives experimental treatment, and both groups are observed. Perhaps the best advantage of Design V or Design II is that it adds the comparison group against which the changes in the two observations of subjects in the experimental condition can be checked. Without such a comparison, the changes might have occurred by history, maturation, or both. If one group matures, or if an event in history occurs which may have an effect on performance, the element of random assignment would be sufficient to assure us that these effects would probably have operated with equal force for both groups.

One problem with this design is the sensitization effect of pre-testing - the reactive measures problem. With random assignment, one could assume that the tendency to be sensitized to the second observation would occur with equal impact in both groups, so this threat to <u>in</u>ternal validity would substantially be met. But perhaps the pre-test had the effect of predisposing the experimental group to the treatment itself - getting it ready for those class discussions and readings on the nature of prejudice, for example. In this case, it would be difficult to generalize any findings to a non-pre-tested group: the most we could conclude is that the treatment would have a certain kind of effect upon the subsequent behavior of pre-tested subjects.

Design VI: Compromise Experimental and Control Group, Pre-test - Post-test

Paradigm:

0_1	X	0_2	(experimental)
0_1		0_2	(control)

This design is a compromise because all of the values of random assignment are lost. There are, however, precious few instances in education in which random assignment of subjects can be accomplished. Since we live in the real world, the alternative to random assignment of subjects must be considered: we may randomly assign the treatment condition, and we may assure ourselves of answers to some of the threats to internal validity by administering a pre-test and comparing the two samples for equivalence. Measures of both central tendency and dispersion should be used as the basis for assuring equivalence: T tests and F tests are possible and useful for the purpose. While one is left without being able to say that the two groups in this design are in fact equivalent, he is at least in the position of being able to say that, for the measures he had used in the pre-test, there is no evidence that the groups are not equivalent.

Design VII: Three Group - Experimental and Control

Paradigm:		O_1	X	O_2	(experimental)
	R				
		O_1		O_2	(control 1)
			X	O_2	(control 2

Design VII introduces two control groups so as to counteract the effect of test sensitization. By comparing the performance of a pre-tested experimental group with a non-pretested group which also receives the experimental treatment, it becomes possible to assess the consequences of the pre-test on subsequent susceptibility to the experimental condition. The design thus makes it possible to assess and meet all of the threats to internal and external validity, and to present conclusions that it would be possible to defend and generalize with great strength. While random assignment is the most difficult condition of this design to meet, if it is at all possible to achieve random assignment, it should as easily be possible to achieve it for three groups as two.

Design VIII: The Solomon Four-Group Design

Paradigm:

	0_1	X	0_2	(experimental)
R	0_1		0_2	(control 1)
	X		0_2	(control 2)
			0_2	(control 3)

About the only way to strengthen Design VII would have been to eliminate any test or treatment sensitization from one randomly assigned group of subjects, and then to observe them along with everyone else in a post-test situation. So complete is this design in responding to all of the threats to internal and external validity that Campbell has hailed it as the new ideal for social scientists. (13).

Each of the designs introduced here has particular strengths and weaknesses. For most of us, it will be impossible to meet the rigid criteria of random assignment and alternative treatment and control conditions of the Solomon Four-Group design, even if it is the new ideal for social researchers. What we are most commonly concerned about is whether the particular instructional strategy we are employing is resulting in measurable results on the part of our students. We may not even be interested in generalizing our findings beyond our own classrooms, and we may not have the time or resources to go beyond the one shot design, despite its lowly status as totally worthless, scientifically.

Designs III and VI may indeed be the most useful to us in the real setting of higher education. Where random assignment may be impossible, it will still be both possible and practical to compare the experimental and control groups on the basis of their performance on some preliminary observation or test. Whichever design is selected, it is critical that the researcher keep in mind the possible threats to internal and external validity, and that at least nodding acknowledgment of these threats be made as the experiment is executed.

Some Further Remarks About Experimental Design: Murphy's First Law

Researchers and computer scientists are familiar with an irrevocable law which seems universally to apply. Like "Parkinson's Law," or "The Peter Principle," "Murphy's First Law" seems to be widely known and its applications understood, but almost none has ever identified its original source. Briefly stated, Murphy's First Law is:

Anything that can possibly go wrong, will.

In research, one finds the applications of this principle to be almost endless. What we have been discussing to this point is experimental design, and our purpose has been to introduce possible sources of difficulty, or threats to internal and external validity in experiments, and illustrate how those threats can best be met through the judicious selection of an effective experimental design strategy. But to leave the discussion at that is not to provide the most useful service to fledgling researchers. What needs to be done is to step back from the elegance of possible research designs and raise some of the important questions that will otherwise assure the operation of Murphy's First Law as we begin to evaluate the consequences of new instructional strategies.

The Committee on Research and Development of the California Junior College Association is frequently asked to endorse or review research proposals so that the researcher can use Association approval to gain better access to the field as he attempts to conduct research in California community colleges. To make the process of review more meaningful to the prospective researcher, and more standardized for the Committee, we have developed a one page sheet which lists the six elements of a proposal which need to be developed in order to make clear exactly

What the researcher intends to do, Why he thinks it is valuable and possible to do it, and How he intends to find answers for his research questions. The project approval sheet is included below, but three elements of it may be elevated for special attention here: Statement of the Problem, Justification for the Proposal Approach, and Operational Research Objectives, Hypotheses and/or Questions.

Developing Clear Statements of the Research Problem

Research on the impact of instructional methodology or strategy is particularly difficult if one of its stated purposes is to compare contrasting methods offered to two or more classes, say, of Economics 1A. One of Kerlinger's major points in his discussion of "maximizing" variance due to experimental treatment is that "one of the main tasks of an experimenter is to maximize this variance. . . he must 'pull apart' the methods as much as possible to make A_1 and A_2 as unlike as possible." In conducting research in higher education, what we are often talking about is taking two groups of students, each of which is enrolled in a one semester class on the same topic (although possibly with different instructors), administering one teaching approach to one group and another to the second, and observing the performance of students on a final examination for the purpose of comparing their "mastery" of the subject matter, given "experimental" versus "control" conditions. What we really hope to isolate is the specific instructional method or approach, and its particular consequences for a group of students. In a complete statement of the research problem, then, it is critical to provide the most extensive evidence that subjects are being selected and research questions are being asked in a way that will assure the best "isolation" of the research problem. At the very least, the problem statement should clearly identify:

a. the variate and criterion (or "independent" and "dependent") variables in the study.

b. the type of relationship between the variables.

c. the target population.

The importance of selecting and clearly stating the criterion variables cannot be emphasized too much. Perhaps the greatest contribution of McKeachie's recent review of research on college teaching, to which reference was made above, was his providing evidence that alternative instructional approaches have different consequences in student persistence, student attitude, and retention of learning after the final examination. If the criterion variable for instructional research is the final examination alone, perhaps the real impact of alternative methods of instruction will not be assessed at all. This point is the more relevant because of the current interest in what is being referred to as a "systems approach" to instructional development. Indeed, the advocates of such an approach are almost evangelical in their zeal, and deprecating in their manner. One publication in the field goes as far as to suggest the following:

> . . .to merely afford a scholar more opportunity to prepare a course of instruction is not sufficient condition, in and of itself, to insure the improvement of instruction. This weakness can be overcome through the application of the systems approach. Of course, we will need behavioral technologists trained in the systems approach to interact with and guide the untrained scholar in the instructional development. (14)

Despite the somewhat agressive tone of the statement, it may just be that the influence of Bloom's Taxonomy of Educational Objectives, Mager's Preparing Instructional Objectives, Cohen's Objectives for College Courses and the Teaching

Research Division of Oregon State System of Higher Education's The Contribution of Behavioral Science to Instructional Technology will result in our taking dramatically different stands concerning the goals and objectives of our individual courses. Thus merely to compare student performance on a final examination, or, especially, to compare grade distributions for classes taught by alternative methods would be to miss the crucial point that the real difference between the experimental and control conditions may have been in the area of goals and objectives of instruction as much as in the method or strategy.

In particular, there seems to be a new interest in defining goals or objectives in the area of "affective" learning, along with "cognitive" learning. To assess cognitive mastery through the medium of standardized final examination may be appropriate, but it is much more difficult to define a criterion variable in the affective domain. One reference that may be particularly valuable in this quest is Shaw and Wright's Scales for the Measurement of Attitudes (New York; McGraw Hill, 1967), which is an amalgam of actual scales, their uses and scholarly histories. Of particular interest to this group might be the sections on "Attitudes Related to Practices in Economics", "Attitudes Related to Economic Issues," and "Attitudes Related to International Economic Issues."

Perhaps the point that establishing clearly the criterion variable does not need to be stressed so disproportionately, since it is also critical to provide some indication of the independent variables (what it is you intend to manipulate experimentally, or control in such a fashion that assigned variance can be identified), and of the target population. These concerns are more directly related to the discussion of specific research designs, and may adequately have been covered in an earlier section of this paper. Clearly, it is necessary, particularly in Design III and Design VI, which have been suggested to be the most practically useful of the eight reviewed above, to define the target population in such a way that questions of comparability of the samples are answered beyond reasonable doubt. The most direct way of providing such answers is to indicate the independent variables which were included beyond the treatment variable, and to show specific evidence of the comparison between the various groups of subjects in the experimental design on each of the independent variables.

Providing a Justification for the Research Approach

When one has listed the criterion variables, the independent variables, and the nature of the target population, he has essentially portrayed the results of a process of elimination for full justification of the research approach, an acknowledgement of the variables that were not included in the design is desirable. Of particular interest in this justification are the criteria for selection which led to the particular list of variables in the study. Thus, for example, the researcher may have considered a "love of Economics" criterion variable in the affective domain as one of the possible measurements at the end of the experimental period, but may have been forced to abandon this consideration because no such instrument could be found in the literature, and the researcher did not have the competencies or sufficient interest in the question to develop a scale of his own. More important, however, is the acknowledgement of variables which were considered, but over which the researcher could achieve no control. A prime example of such a variable is the "instructor" variable. In studies, it is impossible to select a sample of students enrolled in two comparable sections of a particular course, taught by the same instructor, but using different instructional strategies. One must acknowledge that instructor variance is a factor influencing student performance on the criterion variables. In fact, instructor variance may be of a sufficient magnitude to overshadow variance due to instructional method or medium. The failure to control for such variance will of course be reflected in the findings of the research, and a justification or acknowledgement of the possible impact of this weakness needs to be made.

The same kind of comment may also be made about the need to clarify possible experimental manipulations of the treatment variable. Thus, for example, the researcher may be interested in the impact of video tape recordings of small group seminars on specific issues in Economics on the subsequent performance of students on a criterion variable. Because of campus limitations, it may be the case that VTR units are available only in the campus multi-media center. The researcher must decide, then, how to manipulate the experiment in such a way that all experimental subjects hold taped discussions, and, perhaps, have the opportunity to view the tapes immediately. Problems of scheduling, room size and space, immediacy of playback, and other innumerable instances of the possible operation of Murphy's First law need to be considered, and a justification for the specific manipulation strategy needs to be indicated.

<u>Operational Research Objectives, Hypotheses and/or Questions</u>

If the researcher has managed to stay with the process to this point, he has clearly in mind what the criterion variables, independent variables, population, and treatment variables are. The central questions of this aspect of developing the research proposal are "How will this whole ballet be choreographed?" and "How will I know whether the dancers have understood and followed the choreographer?" Thus, the question "Does the use of video taping have any impact on the authoritarianism of students?" is really not specific enough. Alternatively "Does the use of video tape to record and play back small group seminar discussions on specified issues in Economics have an impact on student attitudes, as reflected in pre-test - post-test comparisons of scores on the Rokeach Dogmatism Scale?" is a better statement of the question, since it includes more information about the specific treatment and criterion variables. The question could be improved further, to add more specification about the treatment manipulation, for example, but the point may have been made already.

Some Concluding Remarks

The purpose of this paper has been (1) to introduce the concerns to which experimental designs address themselves (internal validity); (2) to introduce several alternative experimental designs that may have applications for instructional research; (3) to introduce the CJCA Committee on Research and Development proposal components list. The real substance of learning to design and execute research projects is in the actual experience of doing so: knowing that Murphy's First Law can operate is not the same as watching it operate when you discover you have gathered the wrong data, or neglected to control some critical independent variable. That's an experience every young researcher should have, and will.

Attending to the major points in this presentation may provide at least a frame of reference within which the hard tasks of designing and conducting research may make more sense. Briefly summarized, those points are:

1. The major function of a research design is to control variance, and to enable the researcher to provide answers to specific research questions.
2. There are identifiable threats to the "internal" and "external" validity of an experiment which can be understood and controlled by careful selection of an appropriate research design.
3. There are a variety of experimental designs which may be applicable to the instructional process, but the researcher's ability to execute any given design will be particularly influenced by the amount of control he can exercise in the selection of subjects, the random or non-random assignment of subjects to treatment conditions, the nature of the experimental manipulation, and the nature of the criterion measurements.

4. In the development of a research proposal, the researcher can and should specify as clearly as possible all of the specific research questions that are addressed by the experiment, the reasons for selecting criterion and independent variables, the methods and procedures for measuring the criterion variables and controlling the independent variables, and the nature of the sample which is to be the subject of the experiment.
5. Since "anything that can possibly go wrong, will," the researcher must learn to differentiate between the things he can control or correct and those he cannot. Knowing about and selecting a particular design for research won't matter much if one critical variable has been ignored or forgotten in the process: errors or omissions in data collection cannot be corrected after the samples have dissipated, matured, or changed because of historical accident or event.

Let these points serve as a welcome and introduction to the field. While not all of the practitioners of institutional research practice as if they were aware of these points, the quality of the research that we should expect in such a critical field as the evaluation of our instructional methods and practices might be improved if only we behaved as if we knew that these points were indicators of meaningful problems which the researcher can acknowledge, and, within the limitations of his project, control.

The component titles that should comprise a sound research proposal include:

1. Project Indentification (Title)

   A good project title should very concisely identify: 1) the variables included in the study; 2) the type of relationship that may be inferred between the variables; and 3) the population to whom the results may be applied.

2. Statement of the problem to be solved or situation to be improved

   A good problem statement clearly identifies:
   1) the variate and criterion variables (or independent and dependent variables in an experimental study)
   2) the type of relationship between the variables
   3) the target population

3. Justification for proposal proposal approach
   1) the more important relevant variables considered for possible inclusion
   2) the criteria for selection of variables included in the study
   3) the ways in which the variable might be manipulated
   4) criteria for selection of specific manipulations

4. Operational research objectives, hypotheses and/or questions

   To assure that the proposal is presented in sharp form it is necessary to state the hypotheses, objectives, or questions in operational terms -- that is, the procedures and/or behavioral outcomes must be clearly specified and observable.

5. Sequence of operations and procedures to be used in solving problem

   This component of a research proposal is variously labeled method, procedure, or in one instance method or procedure. The basic function of this component is to describe the operations that will be performed to solve the problem of concern.

6. Evaluation of data

   The description of the data analysis and evaluation procedures should include the following elements:
   1) Indication of consistency of method of analysis with research objectives and design.
   2) Specification of method of analysis including:
      a) Use of special analytical tools, computer, card-sorter, etc.
      b) Use of graphic techniques
      c) Specification of types of tables to be constructed
      d) Specifications of statistical and/or other analytical procedures to be utilized.
   3) Indication of the types of statements that may be validly made from analyzed data.

BIBLIOGRAPHY

1. Kaufman, Roger "Systems Approach Applied to Instructional Development: A fable" in The Contribution of Behavioral Science to Instructional Technology (Oregon: Teaching Research Division, 1970)

2. Dubin, R. and Taveggia, T.C. The Teaching Learning Paradox (Eugene Oregon: University of Oregon Press, 1968)

3. McKeachie, Wilbert J Research on College Teaching: A Review (Washington, D.C., ERIC Clearinghouse on Higher Education, November, 1970)

4. Kerlinger, Fred N. Foundations of Behavioral Research (New York: Holt, Rinehart and Winston, 1964)

5. Ibid, Chapter 15

6. Campbell, Donald and Stanley, Julian Experimental and Quasi-Experimental Designs for Research (New York: Rand McNally, 1963)

7. MacMillan, Thomas F. "Norcal: The Key is Cooperation," Junior College Journal (May, 1970)

8. Plant, Walter and Telford, Charles Changes in Personality for Groups Completing Different Amounts of College Over Two Years (Genetic Psychology Monographs, No. 74, 1966)

9. MacMillan, op. cit.

10. Kerlinger, op. cit, p. 264

11. Steckler, G. "Authoritarian Ideology in Negro College Students", Journal of Abnormal Social Psychology V 54, 1957, pp. 396-99

12. Kerlinger, op. cit. p. 302

13. Campbell, Donald "Factors Relevant to the Validity of Experiments in Social Settings" Psychological Bulletin, LIV (1957) pp. 297-312.