DOCUMENT RESUME

ED 141 514 CE 011 097

AUTHOR Scissons, Edward H.
TITLE Representative and Quasi-Representative Designs for the Improvement of Adult Education Research and Evaluation Studies.
PUB DATE Apr 77
NOTE 16p.; Paper presented at the Adult Education Research Conference (Minneapolis, Minnesota, April 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Adult Education; *Educational Research; Research Criteria; *Research Design; *Research Methodology; Research Problems; *Validity

ABSTRACT

Traditional research literature distinguishes between two general types of threats to the generalizability of experimental findings: Internal validity and external validity. Relatively minor importance has been attached to external validity in educational literature. Bracht and Glass elaborate on external validity and deal with two types: Population validity and ecological validity. Snow indicates that the biggest threat to external validity occurs when the experiment does not fit the phenomena being studied. Traditional experiments are nonecologically oriented, tend to manipulate the subject for experimental convenience, and yield results which are non-representative. Historically, basic research has emphasized internal validity and applied research has been concerned with immediate problem solving (external generalizability). Generalizability must be the sole factor in judging all good research, and the fundamental design principle governing generalizability is representativeness. There are a number of strategies educational researchers may use to build significant and generalizable studies, for example, imbedding experimental procedures in existing structures in as nondisruptive a manner as possible, replicating experiments, using placebo treatment groups, examining the utility of proposed research, and not overgeneralizing from results. Adult educators also must be concerned with the generalizability of their research. They must be applied researchers, emphasizing external validity over internal validity. (LMS)

Representative and Quasi-Representative Designs for the Improvement of Adult Education Research and Evaluation Studies

EDWARD H. (ED) SCISSONS, PH.D.
ASSOCIATE PROFESSOR
UNIVERSITY OF SASKATCHEWAN

PRESENTATION TO THE ADULT EDUCATION RESEARCH CONFERENCE, MINNEAPOLIS, MINNESOTA, APRIL, 1977.

UNIVERSITY OF SASKATCHEWAN
SASKATOON CANADA
EXTENSION DIVISION
S7N 0W0

Edward H. (Ed) Scissons, Ph.D.
Registered Psychologist
Associate Professor

Introduction

The purposes of educational research are to explain and predict human behavior. It is usually unwieldly to undertake research about people using everybody as subjects and everything as variables (although this has not stopped many from trying!). If I may be permitted an analogy, educational researchers must be satisfied with biting off a miniscule piece of a delicious pie and attempting to comment of the hidden qualities of the whole pie having tasted only a small mouthful. To continue the analogy, most work in educational research has focused on two main areas; a) how and from what part of the pie should the mouthful be taken?, and, b) what can you legitimately say about the whole pie having ravished only a small portion. This paper addresses both of these important questions.

The Problem: A Theoretical Perspective

Traditional research literature on experimental design and the evaluation of programs or processes (Campbell and Stanley, 1963) distinguishes between two general types of threat to the generalizability of experimental findings; internal and external validity. Internal validity (interpretability) is affected by actions such as the violation of statistical assumptions or laboratory design for the purposes of treatment. External validity, which subsumes internal validity, is governed by the broader concept of generalizability, i.e., the representativeness or appropriateness of experimental findings when applied to other populations. Traditionally, the education researcher has focused on internal validity as a means of increasing overall validity: Experiments have been made "tight", subjects and subject behavior have been closely controlled. Unfortunately, the result has often been the formation of "lab-wise" but "world-naive" principles.

The problem with such research is not that it is not good. It is--within the restricted parameters within which it is designed. The difficulty comes, (and it is a difficulty that all researchers are confronted with), when lab or "tight" research findings must be used as describers or predictors of real, non-laboratory behavior.

In the main, relatively minor importance has been attached to external validity in the educational literature. In seeking to emulate their more "rigorous" older brothers in the research game, educators have fallen prey to the use of techniques which may be of use in differentiating between crop yields in response to fertilizer, but are of questionable use in the explanation of human educational processes. Bracht and Glass (1968) elaborate considerably on the principle of external validity in applied experimental design. They deal with two different types of external validity; population validity and ecological validity. Under population validity, they deal with the inherent difficulties in generalizing from samples (usually student "volunteers" who, if they are nothing else, are accessible) to target populations which often differ markedly from these students. They offer the possibility that the interaction of subject characteristics with experimental variables (subject by variable contamination) is a major limitation to the generalizability of current research findings. Ecological validity is characterized by the problems inherent in the environment of the experiment; description of independent variables, multiple effects of multiple treatments, Hawthorne effects, and errors due to dependent variable measurement, etc.

Snow (1974) in carrying on the work of Bracht and Glass (1968) indicates that the biggest threat to external validity in all types of educational experiments occurs when the experiment does not fit the nature of the phenomena being studied. In short, he is concerned with the imposition of experimental convenience on real non-experimental phenomena. Snow (1974) indicates that a particular problem develops when no method is included in the experimental format to discover this type of error factor. Snow (1974) supports his essay by reviewing the work of Brunswick (1956) in which it was demonstrated that variables associated in time and space cannot be legitimately treated as independent. In other words, if one views that in some way the behavior of an individual is contingent on the environment in which he operates, (a proposition very difficult not to accept), the description of that subject's behavior in an alien environment (independent from the host environment) and an attempt to generalize those laboratory findings to the real world situation, is folly. Many variables, which are positively associated in nature, ill serve the process of educational research by being separated arbitrarily in factorial research design.

Given the agricultural roots of classical analysis of variance and factorial design experiments, it is easy to see how the experimental subject has come to be viewed as a passive object (with the treatment or educational program, the active evaluated entity). Traditional experiments are non-ecologically oriented, tend to manipulate the subject for experimental convenience, and yield results which, although "tight", are non-representative of the world-at-large or the population of the world-at-large.

Applied versus Basic Research

Basic research refers to "...the activity whose immediate aim is the quantitative formulation of verifiable general laws, and whose ultimate aim is establishment of a system of concepts and relations (the so-called nomothetic net) in which all specific propositions are deducible from a few general principles. Basic research seeks eternal verities. Its hallmark is the carefully designed and well controlled experiment whose conclusions are rigorously tested for statistical significance (Ebel, 1967, p. 81)." Historically, basic research has involved a search for internal validity, a search reflected by Campbell and Stanley's (1963) treatise.

Applied research is concerned more with immediate problem solving. The solution of current relevant problems or issues, and the formulations on which they are based are temporary (Ebel, 1967).

On the basis of these two definitions, it would seem that basic research is the crown prince of all research with applied research a poor country cousin. But, is such really the case? The honored position of basic research has rested largely on the foundations inherent in the internal validity work summarized by Campbell and Stanley (1963), a search for generalizability based on a high degree of internal experimental rigour. The demise of applied research has rested on the apparent unconcern of the applied researcher with factors such as statistical assumptions, criterion-predictor contamination, variable independence, and data homogeneity.

Aside from the obvious differences in the two foundations, how far apart are applied and basic research within the perspective of generalizability? Generalizability, it should be noted, must be the sole factor in judging all good research. Are the results reported in experiment A (be that a basic or applied experiment or program evaluation) generalizable to an analgous natural situation? There is no point in researching a truly unique situation. Given the criteria of generalizability, the bulk of educational research, be that research basic or applied, fails to meet the mark. Basic research, while maintaining a high degree of internal rigour, often sets up an alien environment which largely ignores typical subject by environment interaction. It is not sufficient to undertake educational research the results of which are generalizable solely to other experiments. Conversely, applied research often attempts to answer all questions, for all time, for all people. By attempting a global whitewash approach to subject selection, criteria, and variable selection, the subject and the results are often described in such a way as to preclude generalization. It is not so much that the bulk of applied research is "soft"--it is that--but rather that its objectives are ill defined, diffuse, and often unspecified.

The fundamental design principle which must govern generalizability in educational design is representativeness. By this criteria, basic research has failed, since most experimental design represent only other experimental designs, and applied research has failed in its ubiquitous description of the interaction between subject and environmental characteristics. In short, most basic research fails miserably when viewed within the perspective of representativeness and quite probably most applied research does no better (with most applied research programs, so little information is known that it is not possible to make such a qualitative decision). Generalization from educational research studies is largely a two-step procedure. In step one, we generalize from the sample of subjects involved in the experiment to the restricted accessible population from which the experimental sample was drawn. In step two, and it is here that most studies are on shaky ground, we generalize from the accessible population (students, program participants, etc.) to a larger population with which one is ultimately concerned. Errors in making this two-step generalization are additive.

The funadmental assumption which is necessary to be met, and which governs the degree of representativeness achieved, is one of sampling randomness. Such randomness, however, is not easily come by. It is difficult to achieve in step one; it is impossible to achieve in step two. People, unlike the kernels of wheat so at home with such techniques, do not compartmentalize so easy.

Several researchers (Bracht and Glass, 1968; Snow, 1974) have addressed this problem of two-step generalization. Their findings indicate that, although one cannot often generalize to the intended population with any degree of certainty, it is possible to generalize to a "hypothetical accessible population like those involved (Snow, 1974, p. 270)." The crux of this generalization, whether it be done in steps one or two, is a thorough description of subject and environmental characteristics. Generalization then, is attempted cautiously, based on a thorough evaluation of the similarity of characteristics between the two or more groups. Clearly, this is the job of an educational craftsman.

3 Interactions of subjects characteristics, as would be readily noted in the real world, are also an important consideration from the perspective of limitations to generalizability. A thorough description of subject characteristics--at both levels of generalization--facilitates the investigation of such interactions. Snow (1974) notes this point in discussing referent generality (the range of characteristics investigated in any given experiment). Design, he argues, is strengthened by increasing the range of traits under consideration and the interactions between those characteristics. These points are echoed in an important paper by Clark and Snow (1975) in which they present six alternative designs for instructional technology research, all of which require increased specificity in subject description as a precursor to multiple regression-like techniques.

Strategy Concerns for Educational Researchers

What course is open to the educational researcher in general and the continuing education researcher specifically if he or she is to build significant research studies based on a consideration of the limitations to generalizability inherent in many traditional designs? As outlined below, there are a number of significant considerations:

1. Experimental Imbedding. If one is to truly evaluate a program, of fundamental concern must be the maintaining of the reality of that program during the course of the research study. To the extent that the research study itself alters the fundamental nature of the program being evaluated, the research findings from that program are non-generalizable. Experimental procedures need to be imbedded in existing structures in as non-disruptive a manner as possible.
2. Experimental Replication. The branching out of the program being evaluated (e.g., systematic replication in time or consequence) provides one index of possible generalizability. Most researchers know that any given point, no matter how silly, can be supported "by the research." Replication might cause a number of questionable ideas to die in the bud. What is needed in educational research is not breaking of new ground but the systematic investigation of what has already been seeded. Glass (1976) makes this point well in his description of meta-analysis (analysis of analyses) as a necessary and important form of research.
3. Internal\Experimental Investigation. Most often, in describing an experiment, a researcher will describe only terminal subject behavior in comparison to entering subject behavior. Process variables are often ignored. Internal documentation and observation of subject

behaviors (particularly important in program evaluation) can do much to explain conflicting or so called "negative results."

4. External Experimental Investigation. Do all experimental subjects decease once a research study has been completed? Do all educational programs cease operation on the publication of the research paper? Such questions are unanswerable in a conventional research design. The increase in information due to post-experimental investigation would undoubtedly lead to an increase in generalizability due to an increase in referent generality.

5. Subject Description. As noted earlier, a thorough description of subject characteristics, including aptitude variables and interaction effects, is paramount if an increase in generalizability is to be achieved. It is important to know why things work as well as that they work. A binary (effective versus non-effective) approach to educational research ill serves the process of education.

6. Systematic Treatment Description. Frequently, the literature reports educational studies, particularly program evaluations, where very little, other than the effects of the programs, is described. Thus, we see participants subjected to a "Life Skills Program", "T.A. Program", "Community Planning Workshop", etc., with little told the reader about the content or process of such a program. Often, when the reader attempts to duplicate such a program, the results are disheartening. As has already been indicated, the interactive effects of a student by learning program by learning environment paradigm is probably the most important consideration in program

evaluation. However, interactive effects are largely ignored in a simplistic evaluation of main effects, an evaluation which hides as much as it shows. The mere reporting of program labels, without a thorough description of the interactive process of such programming, is wholly inadequate.

7. Placebo/Treatment Groups. Most frequently, programs and/or educational experiments are evaluated in terms of a no-treatment control group. What do the results of such evaluations really tell us? The results could probably be summarized as "something is better than nothing." The truly important question in relation to program evaluation is their efficacy vis à vis existing and/or alternate programs. This is not to say that the use of no-treatment control groups is wrong--it is simply inadequate.

8. Nipping in the Bud. Stated bluntly, there is a great deal of research that is not worth doing; it is simply a waste of time. It is not sufficient to "do the best we can within the constraints given us." If proper attention cannot be paid to important research considerations, the program ought not to be bothered with from a research perspective. The problem with research is that, once completed, poorly defined and operationalized studies often receive the same or more weight as well conceived and operationalized ones. Good research suffers in this "lowest common denominator" approach. Universities particularly are to blame in this area in coercing their graduate students to "go through the motions" in undertaking research that is both unimportant and uninteresting.

More attention needs to be paid to the utility of proposed research rather than attempting to patch-up a sick study which should have been aborted on conception. The difficulty may be that we often look at a study, commit ourselves to it in an evaluation sense, and then attempt to do the best possible within the constraints given. Early on in program evaluation, basic go-no go decisions need to be made. That is not to say that the program itself should not go on, but it should not be "glorified" by research associated with an evaluation process.

9. Too Much of a Good Thing. The urge to overgeneralize is a frequent obsession of educational researchers. A professor that I once know many years ago made what he thought was a world shattering observation. He said that the human organism seems largely incapable of carrying in consciousness more than two alternate hypothesis. Thus we have hot-cold, good-bad, efficient-inefficient, sick-well, etc. The result is that we often take the findings from a very restricted study and overgeneralize to populations that are largely dissimilar. Too much is often made of too little.

The Continuing Educator and the Research Dilemma

The continuing educator is in largely the same position with respect to research programming as his more traditional educator cousin. Needs research, program evaluation, community studies, alternate educational strategies, all are taken as important, albeit crucial aspects of the continuing educator's role. Within the framework adopted by Ebel (1967), the continuing educator must be an applied researcher; he is concerned with the solution of immediate problems, ones largely ephemeral in nature. Educational programming, however, does not occur in a vacuum. Situations are seldom wholly unique. The continuing educator is , must be, concerned with the generalizability of his findings, an area Ebel (1967) reserves for basic research.

However, given the criteria of generalizability of experimental findings based on the representativeness of the research undertaken, distinctions between basic and applied research become less clear. For, if any research is to be good, it must be generalizable. To be generalizable is to include elements of both of the traditional concepts of applied and basic research strategies. Simply to undertake research in order to evaluate a single program is insufficient, if that research cannot be legitimately applied to similar programs. However, to do research on programming conceptually is also insufficient since applicability is largely ignored. The continuing educator/research must be concerned with the problems inherent in the specific program being evaluated, but he must also be aware of what Ebel (1967) calls the "nomothetic net (p. 81)."

There are some that will say that the approach I have proposed in this paper, which emphasizes an external rather than an internal experimental orientation to research, is anti-scientific, even anti-education. Let me assure you that such is not the case. I am a thorough-going rational/logical empiricist. I still believe in the "magic" of research and the design of educational systems based on other than rhetoric. Where I differ from many of my colleagues in the research game is a belief that in emulating our "hard" science brothers, we are seeking to catch the wind. To be an educational researcher is to live with confusion, but it is also to seek to dispel a portion of that confusion through a scientific attitude of inquiry suited to our discipline. The hope, the only real hope as I see it for educational research, is contained in the notions of generalizability and representativeness as I have attempted to outline them in this brief paper.

To continue the analogy I posed at the start of this paper, it may be not so much that we have been overly careful in selecting the piece of pie we wish to taste but that we have either been reluctant to really taste the pie at all or that we say too much about the miniscule piece we have tasted.

References

Bracht, G.H., and Glass, G.V. The External Validity of Experiments. American Educational Research Journal, 1968, 5, 437-474.

Campbell, D.T., and Stanley, J.C. Experimental and Quasi-Experimental Designs for Research on Teaching. In N. L. Gage (ed.), Handbook of Research on Teaching. Chicago: Rand McNally, 1963.

Clark, R.E., and Snow, R.E. Alternative Designs for Instructional Technology Research, 1975, 4, pp. 373-394.

Ebel, R.L. Some Limitations of Basic Research in Education. Phi Delta Kappan, 1967, 49, 81-84.

Glass, G.V. Primary, Secondary, and Meta-Analysis of Research. Educational Researcher, 1976, 10, pp. 3-19.

Snow, R.E. Representative and Quasi-Representative Designs for Research on Teaching. Review of Educational Research, 1974, 44, 265-292.