DOCUMENT RESUME

ED 261 094 TM 850 518

AUTHOR Sax, Gilbert
TITLE Quantitative Methods: A Critique.
PUB DATE 85
NOTE 10p.; Paper presented at the Annaul Meeting of the American Educational Research Association (69th, Chicago, IL, March 32-April 4, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Research; Elementary Secondary Education; Higher Education; Researchers; Research Methodology; *Research Problems; *Statistical Analysis
IDENTIFIERS Causal Inferences

ABSTRACT

This paper addresses several issues in quantitative research that educational researchers should examine with more care. While the purposes of experimentation is to determine causality, the study of causal relations is difficult and problematic. Computational and conceptual errors in statistical analysis seem limited only by the creativity of the researcher. The problem of evidence that contradicts theory is too often solved by throwing out the data or renaming the facts. While researchers have volunteered to improve education, the imposition of a research finding on all children everywhere regardless of the lack of evidence or the presence of questionable evidence is at best a mistake that might not be able to be remedied later. (BS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

© Gilbert Sax 1985

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Sax, Gilbert
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED261094

QUANTITATIVE METHODS: A CRITIQUE

AERA Chicago 1985

Gilbert Sax
University of Washington

My task--and I assure you that it was--is first to critique "quantitative experimentalism," a term I don't readily understand, and second, to do so in no more than fifteen minutes. Although I have been warned to exceed neither the time nor my knowledge limits, only the first restriction can be completely controlled.

However, it is because of time restrictions that I have chosen not to discuss the usual criticisms of quantitative methods with which we are all so familiar: the lack of isomorphism between measurement and "reality," whether reality can ever be known epistemologically, whether any or all educational and psychological constructs are measured by ordinal or interval scales (and whether or not it makes any difference), and whether we should accept the .05, the .01, or the .001 significance level. I will eliminate temptations to discuss both determinism and the uniformity of nature as topics that require more time than we have. Instead, I will try to address some issues that we, as researchers, should examine with more care than we have in the past.

We are told that the purpose of an experiment is to determine "causal" relations. In fact, I have so stated and in print, but I have always included causal in quotes or in italics. I haven't done this because I understand the complexities of that term; rather, I have done so because I don't understand it at all. Let me provide an example

TM 850 518

cited by Robert Morison (1960) some twenty-five years ago. I have referred to Morison on other occasions because he is one of the few people who seems to realize the beauty of quantification when it is combined with theory and just how ugly quantification can be when it tries to pass as disguised scientism. In discussing "cause" and "effect," Morison makes the point that The Cause of a disease has generally been thought to be whatever it is that could--at some given time and place-- ameliorate the disease's symptoms. For example, the medieval physicians believed that malaria was caused by bad air in lowlands (and thus the term mala aria). The lowlands were the cause since malarial symptoms could be reduced or avoided by building on hilltops. That cause remained undisturbed until quinine was introduced into Europe from South America. Since quinine could counter the symptoms of malaria no matter where one lived, quinine must be acting on the body to rid it of that disease. By the end of the nineteenth century, the malarial parasite was discovered in the blood of those suffering with malareal symptoms, and the parasite became the causal agent. Quinine, evidently, helped rid the body of this parasite. Later, it was discovered that the Anopholes mosquito actually transmitted the disease and was, therefore, its cause. The causal chain extended from location (lowlands), parasite, and mosquito.

The story is not quite over. Malarial epidemics rarely occur today even though little has been done to eradicate the Anopholes mosquito. The Boston marshes still produce mosquitoes that are capable of transmitting the parasite, but no local cases of malaria have occurred. According to Morison, it is now believed "that epidemic malaria is the result of a nicely balanced set of social and economic, as well as biological, factors, each one of which has to be present at the appropriate level" (page 194). This conclusion might sound more familiar to us if we substituted a term such as delinquency for epidemic malaria. And since just about everything is "caused" by social, economic, and biological factors that operate together in unknown amounts and ways, that leaves "modern" researchers on about the same level of knowledge as possessed by their great grandparents. Indeed, I once heard research characterized as the search for evidence to prove what your grandmother knew all along.

John Stuart Mill, the 19th century philosopher, proposed five methods for studying causality. His method of agreement shows the difficulty in studying causal relationships:

> If several instances of an event have only one thing in common, that thing is the cause of the event.

Although this proposition at first seems reasonable, it is not without its problems. Consider an experiment in which ninety men had volunteered to participate in a study on the effects of alcohol. One-third were given scotch and water, an equal number were given bourbon and water, and the last group received vodka and water. Every man in every group got hip-roaring drunk followed by symptoms we all know only too well. The conclusion: avoid water when drinking alcohol. I once asked students in an introductory course in research methods to critique that hypothetical study. I must admit that I was more than a little surprised when one student--in all seriousness--argued that the study was poorly designed because it should have been replicated using school-age children.

Obviously the alcohol study was flawed by having more than "one thing in common," in which case Mill's canon does not apply. All men had water in addition to alcohol, and we all know that water does not cause inebriation. Or perhaps it does. Many years ago I was going to school and teaching an introductory psychology class in adult education. At my request, a dentist friend ordered some nembutal placebos for me. I didn't realize that I would be dispensing drugs without a license in which case I had only anticipated a current trend. That evening in class, I randomly assigned half of my volunteers to take the placebo, and I described vividly how students in other classes had fallen asleep on the floor. No one was permitted to drive home, and everyone agreed not to sue me or the school district in which I worked. After the coffee break I returned to the room to find the experimental group snoring peacefully on the floor. Evidently, even placebos have an effect as more recent studies have suggested. Whether placebos are causal agents or not, we can always resurrect the law of parsimony which argues that of several equally good hypotheses science will tentatively accept the simplest. That makes good sense if we could only recognize equally good and simple hypotheses.

Perhaps we should describe just one more experiment that can be conducted under careful laboratory conditions. In this study, the experimenter wanted to know if fleas could be conditioned. Fleas, by the way, have six legs, and for the purpose of this experiment it was necessary to remove their wings. In classical conditioning the conditioned stimulus precedes the unconditioned stimulus so the experimenter quite properly rang a bell and cut off one leg of the flea. It jumped. The bell was rung again, and again the flea jumped, and another leg was removed. This procedure was repeated four more times, and at the end of the experiment the conclusion was reached that ringing bells cause fleas to become deaf. Since these results can be replicated easily and without the need for any high-powered statistics, we have a reliable finding that we cannot blame on faulty statistics.

Statistics forms an important model in education, and it is distressing in the least to observe how poorly statistical analyses can be performed. Some years ago Quinn McNemar (1960) reported on what he called "an astoundingly fallacious significance level":

> a....psychologist inflated his sample size 36 fold: that is, he had 36 observations on each of 25 cases, leading to 900 observations which were then treated as independent for the chi square analysis. This is one way of getting high* statistical significance with little prospect that similar results will be found by those who replicate the study [note: unless, of course, this becomes standard practice].

*McNemar could have ended the sentence there.

McNemar was right in being astonished regarding the statistical analysis of these data. So many statistical errors can be found in published studies that one can only imagine the number that occur on doctoral dissertations that fortunately never get out of the library. I will not bore you with lists of these errors, but they are there and in large numbers. Computational and conceptual errors seem limited only by the creativity of the "researcher." In part, computers can be blamed for some of these problems by enticing students into working

ERIC

mechanically. One student, after entering only 2-digit numbers for the better part of a day, reported a mean of 113.74 without questioning these astounding results. It is easy to disregard any feelings for the data or for the effects of experimental procedures when researchers are surrounded by mechanical and electronic gadgets that serve little purpose except perhaps to help them exchange what is important for what can be obtained with the least effort and most money.

Students have learned their statistical lessons badly, and they carry out their perceived responsibilities too well. If the null hypothesis cannot be rejected with 30 or 40 persons in each experimental and control condition, everyone knows that the "solution" is to increase N until significance is reached. The motto must be something like significance no matter what! This convoluted reasoning begins with the premise that no two populations are ever identical; therefore, there must be a difference between them that should be reflected in the magnitudes of the treatment means. If that reflection happens to be missing, some ingenuity is needed to force the results to come out as they are supposed to do. Maier's Law (1960) states that "If facts do not conform to the theory, they must be disposed of." I am reminded of some types of test scaling procedures that must have invoked the latent spirit of that law.

Like all good "laws," Maier's has corollary attacks that get right to the heart and can be invoked should some evidence be allowed to contradict a pet or petty theory. Besides throwing out the data, which is one approach to a problem, another good procedure is to rename the facts. Maier provides an example showing that potentially embarrassing behavior to learning theorists who insist that reinforcement is necessary for learning to occur can be handled quite easily by calling the unlearned behavior "imprinting" and not learning. In this way, whatever fails to support some favored position can be retained without having to accept "innate behavior." Maier also suggests that one good way to avoid explanations of events is to given them a title:

> For example, a lecturer in describing the habits of people living near the North Pole told his audience how children ate blubber as if it were a delicacy. Later a questioner asked the speaker why these children liked a food that would not be attractive to children living here. The lecturer replied that this was so because the children were Eskimos. The questioner replied "Oh, I see" and was satisfied. In a similar manner the word "catharsis" explains why we feel better after expressing pent-up feelings. (p. 209)

Another good method for gaining consensus among researchers is to express some position mathematically--as a formula. It may say no more or no less than what could be said in understandable English, but the very appearance of mathematical symbols will do much to quash controversy.

Researchers have volunteered to improve education or they have been persuaded to do so for the most humane of reasons. Nonetheless, it is not the business of researchers to change a world they do not yet understand and which may, in not very many years, give them cause for concern and possibly regret. To <u>improve</u> anything or anyone assumes that we know where we want to go, and I am not convinced that we have the right to modify behavior (assuming that we can) just because it is convenient to do so or because we believe that we have consensus or superior knowledge to fall back on to justify our actions. The purpose of research is to obtain reliable knowledge, and we may choose to do nothing with that knowledge or we may prefer to act on it. In either case it will not benefit our cause to make sweeping generalizations that supposedly apply to all children. The old "new math" was perpetrated on schools and students all over the country before it was tested at all. At the other extreme we can find statements glorifying the deity of ATI (aptitude by treatment interactions). It has been eight years since Cronbach and Snow warned us against believing that we now have (or will soon obtain) instructional guidelines from the ATI research. Unfortunately, I can think of few examples where solid research evidence has changed the public schools; I can think of numerous examples where research has been used to defend or to argue against the wholesale application of an innovation. Quantitative research provides a meeting ground for differing positions that can be investigated empirically regardless of whether or not they provide any amelioration of some applied problem. Educators can refuse to implement innovations regardless of their efficacy if those innovations might lead to social injustice, excessive costs, or perceived negative effects. What should not be demanded of the quantitative researcher is <u>selected evidence</u> to support some biased position--a demand that is only thinly disguised bribery with the payoff being money, recognition, additional time, more space, and new equipment. This misuse of evidence is serious because it is so widespread and because it is not recognized as a violation by either offender--the one who offers the bribe and the one who is willing to accept it. The imposition of a research finding on all children everywhere regardless of the lack of evidence or the presence of questionable evidence is at best an ethical mistake that might not be able to be remedied later. With our current state of knowledge, we can ask teachers to try new approaches when older "solutions" have not worked. That they might refuse to do so is not only reasonable, but it could prevent us from misapplying our own research findings.

REFERENCES

Maier, N.R.F. "Maier's Law." The American Psychologist. 15, No. 3, 1960, pp. 208-212.

McNemar, Quinn. "At Random: Sense and Nonsense." The America. Psychologist. 15, No. 5, 1960, pp. 295-300.

Morison, Robert S. "Gradualness, Gradualness, Gradualness." The American Psychologist. 15, No. 3, 1960, pp. 187-197.