# What Quantitative Research Is and Why It Doesn't Work?



Claudia Krenz Gilbert Sax

Two persistent critiques of quantitative experimentalism are (a) the lack of isomorphism between its measures and "reality" and (b) its failure thus far to produce "truths" useful to educational practice. These critiques have long been commented on. As early as 1918, B. R. Buckingham wrote:

We may labor ingeniously at our analyses of results and may bring from afar the most potent methods which statistical theory has evolved, but we shall accomplish little if our instruments are as grossly defective as some of those which are now being employed appear to be. (p. 132)

Buckingham's concern continues to be echoed by contemporary researchers:

If multiple independent anecdotes are to be trusted, the computers too often have been processing in stolid seriousness worthless data produced by children who were staging mass boycotts, or deliberately sabotaging the process or making jokes out of their answers. Anecdotes of simitar scandals are available for questionnaires, attitude scales and interviews. (Campbell, 1978)

Too often, then, the link between results and "reality" is assumed rather than systematically investigated. Consequently, the empirical bases of educational practice are too frequently half-truths and pure fictions.

American Behavioral Scientist, Vol. 30 No. 1, September/October 1986 58-69.

### Basic problems

We quite agree with the first critique, that quantitative concepts are not isomorphic with quantitative measures. As Bateson (1980, p. 133) noted, "i can, in a sense, see the dog *discriminate*, but I cannot possibly see his 'discrimination.' There is a jump from particular to general, from member to class." As a result,

we have no measuring devices... designed with so perfect a knowledge of all the major relevant sources of variation. In physics, the instruments we think of as "definitional" reflect magnificently successful theoretical achievements and themselves embody classical experiments in their very operation. In the social sciences ourmeasures lack such control. They tap multiple processes and sources of variance of which we are as yet aunaware. :At such a stage of development the theoretical impurity and factorial complexity of every measure are not niceties for pedantic quibbling but are overwhelmingly and centrally relevant in all measurement applications that involve inference and generalization. (Webb, Campbell, Schwartz, Sechrest, & Grove, 1981, p. 36)

The social sciences lack the pragmatic methods of validation available to the contemporary physical sciences (a pity it's easier to design educational programs that backfire than to build bombs that don't work).

Although quantitative researchers usually compromise on the issue of face validity—a measure is valid because it appears so—the abyss between concepts and methods suggests they should not. The methodologically unsound but widely accepted racist conclusions of fifty years ago show how easily the biases of researchers and their times infuse results. Broca's data on cranial capacity, for example, was taken as support for the prevalent notion that men were more intelligent than women, and whites more intelligent than blacks:

Paul Broca is now distant: enough. We can stand back and show that he used numbers not to generate new theories but to illustrate a priori conclusions. Shall we believe that science is different today simply because we share the cultural context of most practicing scientists and mistake its influence for objective truth? Broca was an exemplary scientist; no one has ever surpassed him in meticulous care and accuracy of measurement. By what right, other than our own biases, can we identify his prejudice and hold that science now operates independently of culture and class? (Gould, 1981, p. 74)

And we also agree with the second critique, that quantitative experimentalism does not yield "truth."

The history of Science indicates that most of the avenues explored in the early stages of a branch willead nowhere... We have every reason to believe that most, if not all, present theories will lead nowhere. But that, of course, doesn't mean that they're worthless because wee can never find a fruitful approach without false beginnings. (Kemeny, 1959)

There is only theory-laden perception and thus no grandstand from which anyone, quantitative researcher or otherwise, may review the parade. However, to deduce that we can't know anything from our being unable to know everything is fallacious reasoning. The role of methodology is to chart the "course between the extremes of inert skepticism and naive credulity" (Campbell, 1978, p. 185).

The present essay cannot contribute additional insight into either the lack of isomorphism between quantitative concepts and measures or the attendant failure of quantitative methods thus far to yield "truths" useful to education practice. The abyss between concepts and methods—and the resulting inability of quantitative experimentalism to yield "truth"—is an existential problem for researchers, one that, at best, they can cope with but never solve. By "cope with" we mean that systematic pruning of the untended daisy fields of concepts we've allowed to proliferate must, of necessity, be an integral part of every stage of inquiry. By "cope with" we do not mean ignoring the problem. Yet that seems to be all that's been done.

How much basic epistemological problems are ignored may, in part, be due to the dominant philosophy of science of most quantitative researchers, logical positivism/empiricism:

Today empiricism is the professed philosophy of a good many intellectual enterprises. It is the core of the sciences, or so at least we are taught, for it is responsible both for the existence aund for the growth of scientific knowledge. It has been adopted by influential schools in aesthetics, ethics, and theology. . . This predilection for empiricism is due to the assumption that only a thoroughly observational procedure can exclude fanciful speculation and empty metaphysics as well as. . .further the progress of knowledge. .. empiricism in the form in which it is practiced today cannot fulfill this hope...The fight for tolerance in scientific matters and the fight for scientific progress must still be carried on. What has changed is the denomination of the enemies. They were priests... a few decades ago. Today they call themselves. . . "logical empiricists." (Feyerabend, 1963, pp. 3-5)

Because of what Koch (1964) calls a scandalous lag in the history of ideas, logical positivism/empiricism was adopted by the social sciences and other disciplines about the same time it was abandoned by philosophy.

Despite their initial claims, the postivist/empiricists did not solve the problem of induction. The positivist/empiricist attempts to define theoretical terms-including Russell's explicit definitions, the early Bridgeman's operational definitions, and Carnap's reduction sentences and correspondence rules— failed. Craig's and Winnie's theorems in mathematics, which convinced the empiricists that the theoretical terms in theories consisting solely of theoretical and observational terms could be eliminated or their meanings changed without changing the validity of the theory, were a further embarrassment. Other anomalies, which stemmed from the empiricists' misplaced faith in the nonmodal symbolic logic of Principia Mathematica, included being able to confirm a theoretical statement such as "all ravens are black" with an observation such as seeing a yellow pencil. The empiricists were not heartened by such opportunities for indoor ornithology (Brown, 1977).

Attempts to build a satisfactory model of scientific explanation also failed. Scriven's argument, for example (that having syphilis was a legitimate explanation of paresis although only 5% of those so inflicted develop paresis), helped demolish the empiricist model of explanatory relevance. That is, tough paresis, the phenomena to be explained, was logically entailed by the condition of having syphilis, it could not be deduced with lawlike regularity from having syphilis. Overall, the empiricists' attempts to make science logical were either too rigid, gutting any ability to predict and generalize, or too loose, allowing nonsense statements into the corpus of scientific knowledge. It is to their credit that the empiricists explicitly admitted their failure.

Perhaps the tenacity with which the positivist/empiricist philosophy is held is attributable to the paucity with which it is understood:

What seems to have been imparted to the typical psychologist might be characterized as an ocean of awe surrounding a few islands of sloganized information, as for instance, that a theory is an "interpreted formal system"... that a theory makes contact with "observable state of affairs" via specifications of experimental "operation" or by means of a cryptic device known as the "reduction sentence." (Koch, 1964, p. 11)

Poorly learned philosophy lessons are not, however, the only problem affecting quantitative research.

#### Poor research practice

Statistics lessons have also been learned badly. It is distressing 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<sup>1</sup> statistical significance with little prospect that similar results will be found by those who replicate the study.

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 occurring in the theses and dissertations that fortunately never leave the library. We will not bore you with lists of these errors, but they are there 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; they entice students into working mechanically. One student, after entering only 2-digit numbers for the better part of a day, reported a mean of 113, 174 without questioning these astounding results. It is hard for researchers to develop a feeling for the data or for the effects of experimental procedures when they are surrounded by mechanical and electronic gadgets that often serve little purpose— except perhaps to help them exchange what is important for what can be obtained with the least effort and most money.

Researchers have learned their statistical lessons badly, and they carry out their perceived responsibilitis too well. If the null hypothesis cannot be rejected with thirty or forty persons in each experimental and control condition, everyone knows that the "solution" is to increase N until significance is reached. Or, alternatively, significance ("tabular asterisks" as Meehl, 1978, called it) can be bought at the price of trivial hypotheses, thereby reducing experimental logic to a method of answering questions no one is asking.

The motto must be something like Significance no matter what! or, as facetiously mentioned by Kuhn (1961), "If you cannot measure, measure anyhow" (p. 164). This convoluted resoning 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. Maier's Law (1960) states that "if facts do not conform to the theory, the 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 corollaries 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 the problem another good procedure is to rename the facts. Maier provides an example that shows that behavior potentially embarrassing 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 (1960) also suggests that one good way to avoid explanations of events is to give 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 atractive 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.

Statistics are not, however, the only tools in our arsenal. 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. Because these results can be replicated easily, we have a reliable finding; we cannot blame faulty statistics.

Finally, we can also get a lot of mileage out of quantitative nonexperimentalism. Let one example suffice. A researcher administers a personality inventory to a group of subjects and then uses their scores to identify those whose overall agreement and disagreement with the items exceeds the mean (this is easy to do with any variability at all). Having identified, say, the top 5% of the agreers and the disagreers, the researcher could then write about those variables for which there happened to be significant differences between the two groups (this is easy to do with a large number of variables). The problem comes, of course, with the conclusion that the significant variables discriminate the agreers from the disagreers. No doubt, granting the reliability with which the data were analyzed, the researcher is correct about the variables that happen to discriminate between those who happened to agree and disagree more than the sample average. But whether these variables reflect more than chance-level relationships is something that won't be known unless those relationships are evaluated with new data.

In summary, then, the two persistent critiques of quantitative experimentalism—the lack of isomorphism between its concepts and its measures and its failure to yield "truth"—are valid. That they must continually be raised is due perhasps to widely held but unarticulated philosophical assumptions, especially those proposed by logical positivism/empiricism. This does not mean, of course, that the empiricist philosophy of sciences "causes" slovenly research practice: Researchers un derstand that philosophical position too poorly for there be be a "causal" connection between the two. Additionally, the meaning of "causality" is generally misunderstood.

### Causality: A poorly understood notion

We are told that the purpose of an experiment is to determine "causal" relations. Careful writers either italicize or put quotation maks around causal. It is not the complexities of the term that require punctuation, but rather that causal may refer to disparate examples. Robert Morison (1960)-one of the few to realize quantification's beauty when combined with theory and its ugliness when mindlessly applied—provided an example more than 25 years ago. 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, medieval physicians believed that malaria was "caused" by bad air int he lowlands (thus the term mala aria). The lowlands were the "cause" because malarial symptoms could be reduced or avoided by building on hilltops. That "cause" remained undisputed until quinine was introduced into Europe from South America. Because quinine could counter the symptoms of malaria no matter where one lived, quinine was thought to be acting on the body top rid it of that disease. By teh end of the nineteenth century, the malarial parasite was discovered in the blood of those suffering with malarial symptoms, and the parasite became the "causal" agent. Quinine evidently helped rid the body of this parasite. Later it was discovered that the *Anopheles* mosquito actually transmitted the disease and was, therefore, its "cause." The "causal" chain extended from location (lowlands), to parasite, and eventually to mosquito.

The story is not yet over. Malarial epidemics rarely occur today even though little has been done to eradicate the Anopholes mosquito. The Boston marshes still produce mosquitoes capable of transmitting the parasite, but no local examples of malaria have been reported. According to Morison (1960). 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" (p. 194). This conclusion might sound more familiar to us if we substituted a term such as delinquency for epidemic malaria. And as just about everything is "caused" by social, economic, and biological factors that operate together in unknown amounts and ways, this leaves "moderns" researchers on about the same level of knowledge as possessed by their great-grandparents. Indeed, research has been characterized as the search for evidence to prove what your grandmother knew all along.

John Stewart Mill, the nineteenth-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, one-third received bourbon and water, and the last group received vodka and water. Every man in every group got rip-roaring drunk, followed by symptoms we all know only toowell. The conclusion: Avoid water when drinking aicoho. The first author once asked students in an introductory course in research methods to critique that hypothetical study. He was more than a little surprised when one studentin 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 "once thing in common," in which case Mill's canon does not apply. All the men had water in addition to alcohol, and we all know that water does not "cause" inebriation. Or perhaps it does. Many years ago the first author was going to school and teaching an introductory psychology class in adult education. At his request, a dentis friend ordered some nembutal placebos for him He didn't realize that he would be dispensing drugs without a license (in which case he had only anticipated a current trend). That evening in class, he randomly assigned half his volunteers to take the placebos and described vividly how student in other classes had fallen asleep on the floor. No one was permited to drive home, and everyone agreed not to sue him or the school district in which he worked. After the coffee break he 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, secience will tentatively accept the simplest. This makes good sense if we could only recognize equally good and simple hypotheses.

In summary, the two persistent critiques of quantitative research- the lack of isomorphism between its concepts and its measures and its attendant failure to yield "truths" useful to educational practice-are valid. The persistence with which they are articulated may be in part due to widely held but largely unarticulated philosophical assumptions. Although falsificationism is not without its problems, quantitative researchers would do well to consider substituting Platt's (1964) "strong inference" for their current confirmatory practices. Another valid criticism of quantitative research is that stastitical analyses and interpretations frequently are done so poorly. The problem here is not with the house of quantitative research but rather with the slovenliness of its inhabitants. A fourth criticism is that notions central to quantitative experimentalism, like "causality," are poorly understood. This lack of understanding can be partially attributed to the infrequency with which researchers think about important epistemological issues. It can also be partly attributed to the complexity of that phenomenon.

#### The Role of Research

Researchers have volunteered to improve education or 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 donot yet understand and that may, in not very many years, give them cause for concern and possible regret. This is aperennial problem:

Since Eden, there have been uncertainties about whether knowledge is good... and there is a social science still to be built that will clarify when and how knowledge is likely to be used to exploit or corrupt or dehumanize... The social scientist is trained to hink thast he does not know all the answers. The social scientist is not trained to realize that he does not know all the questions. And that is why his social influence is not unfailingly constructive. (Cronbach, 1975, p. 13)

To *improve* anything or anyone assumes that we know what we want. We do not have the right to modify behavior (assuming that we can) just because it is convenient 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; we may then choose to do nothing with that knowledge or we may prefer to acto on it. 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 exteme we can find statements glorifying the deity of ATI (aptitude by treatment interactions), even though it has been eight years since Cronbach and Snow warned against believing that we now have (or soon will obtain) instructional guidelines from the ATI research. Unfortunately, there are fewer instances in which solid research evidence has changed the public schools than there are instances in which 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; these can be investigated empirically regardless of whether or not the 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 evidence selected to support some bias—a demand that is only thinly disguised bribery, with the payoff being increases in monye, recognition, addit onal time, more apsce, and new equipment. This misuse of evidence is serious because its widespread occurrence is not recognized as a violation either by the offender who offers the bribe or by the offender who is willing to accept it. Moreover, imposition of aresearch finding on all children, regardless of the lack of evidence or the presence of questionable evidence, may cause irreversible harm.

With our current state of knowledge, we can ask teachers to try new approaches when older "solutions" have not woeked. They might reasonably refuse, and so prevent us from misapplying our own research findings.

#### Note

1. McNemar could have ended the sentence here.

#### References

- Bateson, G. (1980). Mind and nature: A necesary unity. Toronto: Bantam.
- Brown, H. I. (1977). Perception, theory and commitment: The new philosophy of science. Chicago: University of Chicago Press.
- Buckingham, B.R. (1918). Statistical terms and methods, In G. M. Whipple (Ed.), The seventeenthe year book of the National Society of the Study of Education (pp. 114-132). Bloomington, IL: Public School Publishing.
- Campbell, D.T. (1978). qualitative knowing in actionr esearch. In M. Brenner, P. Marsh, & M. Brenner (Eds.), The social context of method (pp. 184-209). London: Croom Helm.
- Cronbach, L.J. (1975). Five decades of public controversy over mental testing. American Psychologist, 30(1), 1-14.
- Feyeraband, P. K. (1963). How to be a good empiricist: A plea for tolerance in matiers epistemological. In. B. Laumrin (Ed.), Philosophy of science: The Deleware Seminar (Vol. 2, pp. 3-40). New York: Interscience.
- Gould, S. J. (1981). The mismeasure of man. New York: W. W. Norton.

Kemeny, J. G. (1959). A philosopher looks at science. New York: D. Van Nostrand.

Koch, S. (1964). Psychology and emerging conceptions of knowledge as unitary. In T. W. Wann (Ed.), Behaviorism and phenomenology (pp. 1-41). Chicago: University of Chicago Press.

- Kuhn, T.S. (1961). The function of measurement in modern physical science. ISIS, 52, 161-193.
- Maier, N.R.F. (1960). Maier's law. American Psychologist, 15(3), 161-193.
- McNemar, Q. (1960). At random: Sense and nonsense. American Psychologist. 15(5). 295-300
- Meadows, A.J. (1974) Communication in science. London: Butterworth.
- Meehl, P.E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow pro-

gress of soft psychology. Journal of Consulting and Clinical Psychology, 46(4), 806-834.

- Morison, R.S. (1960). Gradualness, gradualness, gradualne. American Psychologist, 15(3), 187-197.
- Platt, J. R. (1964). Strong inference. Science, 146, 347-353.
- Webb, E. T., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove, J. B. (1981). Nonreactive measures int he social sciences (2nd ed.). Boston: Houghton Mifflin.

Claudia Krenzz y Gilbert Ray son investigadores de la Universidad de Washington.

## investigación administrativa Cupón de intercambio

Solicitamos pertenecer al programa de intercambio de publicaciones que tienen establecido.

Nombre de la publicación:			Periodicidad:		
Adjuntamos un ej	emplar para su in	formación.			
Institución:		Calle:		No.;Coloni	ia:
Ciudad:	Edo;	País:	C.P:	Deleg:	Tel <u>:</u>

Organo interno de comunicación Científica y Tecnológica de la Sección de Graduados y de Investigación Científica y Tecnología de la Escuela Superior de Comercio y Administración del I.P.N.

Sección de Graduados y de Investigación Científica y Tecnológica: Escuela Superior de Cornercio y Administración, Instituto Politécnico Nacional, Carpio 471, edificio D 3er Piso Col. Sto. Tomás, Delegación Miguel Hidalgo 11340 México, D.F. Tels. 541-62-07 y 547-77-92. Apartado Postal 4025 Sta. María la Ribera Deleg. Cuauhtémoc.

investigación

<u>administrativa</u>

# Cupón de Suscripción

Apartado Postal 4025 Sta. Ma. La Ribera Delegación Cuauhtémoc 06400 México, D.F.