DOCUMENT RESUME

ED 381 481 SO 024 945

AUTHOR McBee, Robin Haskell
TITLE Demistifying Research in Education and the Social Sciences. A Primer for LRE Folks.
PUB DATE [93]
NOTE 10p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Citizenship Education; *Educational Research; Higher Education; *Law Related Education; *Qualitative Research; Social Studies; *Statistical Analysis

ABSTRACT

The article argues that the law-related educational (LRE) community should understand education research methods better to demonstrate to policy makers and funding sources the value of LRE and to evaluate alternative educational strategies. To facilitate the reading and conducting of research, the article explains for LRE practitioners research procedures and the format in which research results are presented. The scientific method is described as a process of forming a hypothesis and systematically and objectively testing the hypothesis through observation and experimentation. The article explains the differences between quantitative and qualitative research, experimental and non-experimental research, basic and applied research, and the problems of subject selection, sampling, reliability, and validity. Contains seven references. (JD)

Demystifying Research in Education and the Social Sciences

A Primer for LRE Folks

by

Robin Haskell McBee

DEMYSTIFYING RESEARCH IN EDUCATION AND THE SOCIAL SCIENCES
A Primer for LRE Folks

by Robin Haskell McBee
Virginia Institute for Law and Citizenship Studies
Virginia Commonwealth University

Why Bother?

Most of us are involved in the field of law-related education because we intuitively know that it works. Our own personal experiences with the content and strategies, our informal observations of its "light bulb" impact on teachers and kids, and our anecdotal inventory of success stories all tell us that it works. Unfortunately, not everyone gets our front-row seat in the LRE theater; so not everyone benefits from the full picture we get. Further, the world is full of doubters - people who simply won't take our word for it. Actually, they really shouldn't have to take our word for it; if it works, we *should* be able to prove it. Whenever people in other fields make a claim that something works, the rest of us ask for proof - be it medicine, business, sociology, education, whatever. Even in law we have our own system of gathering factual data in order to prove that something is so - "beyond a reasonable doubt". And it is painfully evident to many of us that the educational decision makers and power mongers require some substantiation before they adopt (AND FUND) a new program or approach. It is, therefore, important for us to begin acquiring some of the rudimentary skills for generating such proof.

One of the "skills" for proving your assertions is being able to understand and use others' research to bolster your own claims or to serve as a starting point for your own investigations. In education, there is a whole "club" of individuals, known as educational researchers, who make their livings generating and reporting on data which substantiate one educational claim or another. They have procedures and a language that are somewhat unique to their field and difficult - if not impossible - for the non-researcher or lay person to understand. Such researchers exist in other fields, as well, and as we move further and further into the violence prevention arena, we repeatedly find ourselves face-to-face with their research reports and claims, too. Even if we never generate our own research, how will we know which of those claims are legitimate and which are not, which to recommend to others and which ones to discard if we haven't got the foggiest idea what those reports say and how to interpret them? How can we at least be informed consumers if we don't understand the language of the research reports? It becomes imperative, therefore, that we at least learn how to understand the talk, even if we never talk it ourselves.

Research in education, psychology, sociology, and other social sciences follows the examples and standards set in the more rigorous natural sciences. However, the language of this research need not leave us fit to be tied. There are some basics which we can acquire rather quickly, and that is the intent of this paper, though this will by no means make us a group of experts. If you want to learn more, you might try taking a course at a local university. All of the social sciences have graduate level courses designed to teach the novice how to understand and interpret research in that field, although sometimes the textbooks are just as dense and difficult to read and understand as the research, itself (another reason why I decided to try my hand at this paper).

The Reason for Scientific Research

Scientific research as we know it today has its roots in the growth of rationality (rational law, rational religion, rational economics, etc.), intellectualism, and cultural stability in the medieval western world. As toleration for scientific inquiry grew and scientists began to seek answers to questions about human existence, they also sought cold, hard, observable proof. In order to prove something, one had to be very logical and methodical - carefully eliminating one explanation after another. Slowly, the process of scientific inquiry (scientific method) evolved into a series of steps, beginning with a hypothetical explanation for a question of importance, followed by systematically and objectively testing that hypothesis, and characterized by the use of observation and experimentation which could be verified.

As sociologist Daniel Chirot (1986) describes it, science is the "crowning achievement" of western rationality.

> *It is based on calculability, on proof, and on empirical observation in a way that no economy, legal or political structure, or religious ethic can be. It is, par excellence, the domain of*

highly trained specialists. It has also taken on a life of its own... (p. 49)

This verifiable proof, based on objective observation, is the essence of rational thinking and lies at the foundation of research in both the natural and social sciences. Over the decades and centuries, it has become accepted procedure to go through certain steps for proving the validity of your research. It is these steps and their explanation which seem so difficult to understand to someone unfamiliar with such scientific procedures and reports.

The scientific method and experimenter objectivity are key to generating proof that is acceptable to the larger world. A successful experiment or study is very specific and limited enough in scope so that the researcher can be confident that the results obtained are logically connected to or caused by the hypothetical explanation that has been offered. The method used to test the hypothesis must have a design which eliminates all possible other explanations (known as *rival explanations*) for the results. There are, therefore, a variety of procedures used in selecting who or what will be studied, how they or it will be studied, and how the results or data will be analyzed (usually involves sophisticated statistical tests). These procedures and their explanations have become the language of social science research, and the remainder of this paper will seek to define and explain some of the basics of that language.

The Scientific Process

The scientific inquiry process begins with a question, problem, or idea which is clearly stated. The question is related to already established conclusions or knowledge in the field. In other words, somebody else has already proven this, but no one has proven that. A hypothesis is then offered as a possible explanation. That hypothesis must be credible to others at the outset. Following the generation of a hypothesis, an experiment or study is designed and conducted. The design must be careful not to leave holes in the process which allow for rival explanations. After the study is conducted, the results are collected and analyzed and conclusions are drawn. The analysis and conclusions must also be logical and credible to the larger world as well as to the researcher.

The Parts of a Research Report

The typical research report in the social sciences follows a similar pattern. After the title and author(s) is(are) given, often an abstract of the study is offered. In most reports this is a paragraph in small print at the beginning of the report which summarizes the research question, design, results, and conclusions. It is similar to the abstract you are asked to write for larger grant proposals.

The body of the actual report usually begins with an introduction, which offers a context for the research and problem to be studied and focuses on the reason for and significance of the study to be taken. Usually, the specific *research question* is posed in the introductory paragraph or paragraphs as well. Following the introduction is a review of literature, which can be anywhere from a few lines to several paragraphs. In it, the researchers outline what kind of related research has already been done by them and others. The review focuses specifically on related questions that have been "answered" and on holes in that research which lead to the question for this study.

Following the literature review, the report offers a specific *hypothesis*, followed by the description of the method or *design* of the study. Often set aside in its own section, the research design specifically describes who or what was studied and how they or it was studied. This includes describing who the *subjects* of the study were and how they were obtained, what was done to the subjects (*procedures*), and what *instruments* (tests, surveys, observations, etc.) were used to measure the impact of the procedures. This section also describes the *statistical test(s)* that was(were) used to analyze the results.

The method section is followed by a report of the results. Commonly, tables and graphs are included in this section. In the final two sections of the research report, the author offers a discussion or analysis of the results and conclusions based on the analysis. The discussion section explains the results and interprets them in light of other research and any weaknesses in the design (where there are possible rival explanations for the results other than the procedure, itself). The conclusion summarizes the answer to the research question based on inferences made from the results, weaknesses in the design, and the relationship of this study to others. The researcher also often recommends areas for further research based on the results of the study.

Types of Research

There are two general types of research: *quantitative* and *qualitative*. Quantitative research comes from the hard sciences. It is more numbers-oriented, requires complete objectivity on the part of the researcher, and calls for analyzing data using deductive and statistical methods. Qualitative research is not tied to numbers at all but rather to observations and descriptions of one or more subjects in natural settings. The researcher often becomes more involved with the subjects and analyzes data inductively, stressing themes and trends. In LRE, a typical quantitative study might be based on the institute evaluations teachers fill out; whereas, a qualitative

study might focus on extended observations of an exemplary mock trial coach in order to determine what it is that mock trial coaches do.

Research can also be divided into *experimental* and *non-experimental* categories. In experimental research, the experimenters are able to *control* or *manipulate* certain factors which affect the subjects. They are studying cause and effect or *causal* relationships between different factors. In non-experimental studies, researchers are describing or analyzing information and cannot manipulate or control the factors being studied. They are seeking either to define or describe simple information or relationships (known as *correlations*) which tie two factors together or to analyze them (as in historical and legal analysis). An LRE experimental study might analyze the impact of the case study method on students' ability to understand the evolution of civil rights in the United States by using the case study method, lecture, and text reading with one class and only lecture and text reading with another class. In such an example, the instructional strategies are the factors or *treatments* that are being controlled and manipulated. Surveys of LRE prevalence in classroom teaching would be an example of non-experimental research in our field. Here we are finding out what people are doing, but we are not changing any factors that influence what it is that they are doing.

You may also, from time to time, run into references to *action* or *evaluation* research. These are two of four general descriptive categories which refer to the reason or purpose for conducting the research. The other two categories are basic and applied research.

Basic research is intended to develop theories or to understand and explain phenomena. Since this type of inquiry usually lies at the foundation of the learning, delinquency prevention, and violence prevention theories with which we are more familiar, it normally takes place in contrived settings such as laboratories, and it is not commonly found in the educational arena. Examining how one builds a cognitive framework for problem solving, or conditions that promote or detract from memory building, or the age at which children are capable of taking another's point of view are all examples of basic research.

Applied research is geared toward the practical application of theories and ideas, and it normally takes place in the natural setting. This is much more common in the education arena. In law-related education, one might examine the application of bonding theory in delinquency prevention and the use of outside legal resource people (ORPs) by comparing the attitudes and behaviors of a group that works with ORPs and one which does not. Another LRE example of applied research might be testing the impact of student courts as a behavior management technique.

Action research is a type of applied research which is used for a specific classroom setting or educational decision and is usually limited in scope. You might want to know, for instance, whether or not doing a full group debriefing session on a daily basis at your summer institute is more effective than small group debriefing, and you might conduct action research specific to your own situation in order to investigate this question. In another example of action research, you might question the ability of a high tech interactive computer program to promote greater understanding of the law by having one of your classes use Tom Snyder's Decisions, Decisions: Prejudice to prepare an essay on conflicting rights and another class use your lawyer partner's presentation and a case study to do the same.

On the other hand, if you were trying to decide whether or not to do a full-scale, system-wide or state-wide implementation of the Center for Civic Education's Violence in Our Schools program, then you might test it out on a selected group of classes to start and make your decision based on the results of that study. This is a form of *evaluation*, which seeks to make larger, more wide scale decisions about the effectiveness of a past or potential program. In another example, you might analyze the effect of your summer LRE institutes in providing sufficient information and demonstrations to make your teacher participants feel comfortable about teaching the content, or you might, several months later, survey the same teachers to determine what content from that institute has actually been taught. Finally, you might test the teacher participants' students to determine what has been learned as a result of teaching that same content.

Variables

Within educational and social science research there are frequent references to several kinds of *variables* which further describe the study being undertaken. As the common usage of this word implies, variables are aspects of the study whose values, degrees, or categories can or do change. Gender, age, socioeconomic status, attitude, behavior, teaching method, and achievement are all examples of variables. Commonly in research studies there are references to *independent and dependent variables*. Independent variables - usually controlled or manipulated by the researcher - influence, predict, or

cause dependent variables to change in some way. Most quantitative research is trying to determine how one (or more) thing(s) affects another one (or more) thing(s). The "things" are the variables, and those that cause the effect are independent while those that are affected are dependent.

If we were studying the effect of participating in mock trials on knowledge of the criminal justice system, we might deliver similar units on the criminal justice system to twelve classes, with six of the classes or *treatment* groups including a mock trial as a culminating experience in the unit and six classes serving as the *control* groups. We would then look at the posttest results to see if there was a difference between the treatment and control groups. In this example, the independent variable is an instructional strategy - mock trials, and the dependent variable would be knowledge of the criminal justice system.

Sometimes other *extraneous, intervening, or confounding variables* enter into the picture and muddy it up. These variables are not controlled by the researcher or accounted for in the research design, and they raise questions as to whether or not the independent variable caused the change in the dependent variable or the change was caused by one of these other variables. Examples of such intrusions on our study of mock trials might include academic ability, socioeconomic status (SES), and age or grade level. If levels of these variables are not consistent across treatment and control groups, it might be one of these variables - rather than the mock trials - which causes the difference in knowledge measured on the posttest. A teacher's prolonged absence, the arrest of a student in one of the classes, or the uneven use of other interactive strategies (some use them, some do not) might also confound the results.

Common Issues in Research

It is not possible in the limited scope of this paper (or of my own knowledge) to cover all of the possible *threats* or compromises to the validity of a particular piece of research. Indeed, some of these threats involve a sophisticated understanding of educational or other types of research. However, there are some fundamental issues or concerns with respect to research that might help you better understand, analyze, and use research to your benefit.

Subject Selection and Sampling

As earlier indicated, subjects are the people or groups who are selected from a larger *population* to be studied or to participate in the study. They form the sample, and the process of selecting them is called *sampling*. Since backgrounds and characteristics can influence the results and can vary from one group to another, the way in which they are selected is a critical issue in applying the study results to other conditions and situations.

Randomly selected samples are selected in such a way that every member of the larger population has the same chance to be selected. There are several means of accomplishing this. With systematic sampling every nth member is selected. Stratified sampling divides the larger population into predetermined groups (e.g. new and experienced teachers, African-American and European-American, or low, middle, and high SES) and randomly selects equal numbers of subjects from each group. In cluster sampling, subjects are randomly selected in equal numbers from naturally occurring groups, such as school systems in the northern section of the state or groups in certain neighborhoods of the city.

If subjects are selected from a population for a particular reason (they are conveniently available; they volunteered; they are representative of a particular problem to be studied), then they are not randomly selected and, consequently, do not accurately represent the population. Therefore, you cannot generalize the study's results to the population from which that sample came. Further, if subjects volunteer to take part in a study, they may bring particular biases to that study which could further cloud the legitimacy of the results. For example, if you conducted a survey on disciplinary practices of all the teachers in the school system, and the responses were completely voluntary, those teachers who are either very pleased with the school system's and their own practices or very displeased with them would be more likely to voluntarily respond. Therefore, the responses would likely be heavily weighted in one of those two directions. Consequently, your picture of discipline in the system's schools would leave out significant portions of individuals somewhere in the middle.

Reliability

Within a particular study or evaluation, specific instruments (tests, surveys, observations, etc.) are used to measure information or collect data related to the research question. It is assumed that no instrument can perfectly indicate that which is being measured. There is always some degree of error associated with the measurement. Responses may change over time or show a lack of *stability* (that is, subjects who are not part of any study perform differently on the same measure when it is given at two different times). They may vary by or lack *equivalence* depending on the group being questioned (e.g. one group of 4th graders

gives a different set of responses than another similar group of 4th graders). Responses may also lack *internal consistency* or differ within the test itself depending on how the question is asked (e.g. subjects strongly agree that police are helpful in one part of the instrument and disagree with a similar statement in another part).

Since these changes or errors in measurement can have an impact on inferences and conclusions that are drawn, researchers must strive to keep such errors to a minimum. Reliability, therefore, refers to the degree to which the measure is error free. It is tested by comparing scores a) on the same measure taken by the same group at two different times (*stability*); b) on the same measure taken by two different but similar groups (*equivalence*); or c) scores on responses to similar questions in different parts of the same instrument (*internal consistency*). Where the instrument is an observation scale, the reliability is tested by comparing observation or rating scores for the same subject(s) but given by different observers (*interrater reliability*).

Scores on tests of reliability are compared by computing a *correlation coefficient* which can range between .00 to .99. With correlations, we are trying to determine how closely one score is correlated to another. The closer to 1.00 the coefficient is, the more strongly correlated the scores and, therefore, the more reliable the instrument. In general, correlation coefficients above .60 are considered acceptable and above .75 good. When reading a formal research report, you should look for information indicating the reliability of the instrument. If it is not offered, this should raise questions for you as to whether the instrument used to measure change is able to do so accurately.

Validity

Outside of research, when we speak of something being valid, we normally mean that it is solid, well-founded, or persuasive. We might even be referring to it having been tested. However, in research arenas, validity has much more specific meanings. In general, validity refers to the appropriateness and meaningfulness of inferences that are made both within the study and in applying the study's results to larger populations. We must ask ourselves, does the study have *internal validity*, or are the inferences we make from the test data appropriate? Does the test we're using in this study really measure understanding of the court process, for example, or how do we know that this new personality inventory really does measure what we have defined as attitude toward the law? Is the experimental design constructed in such a way as to avoid possible threats to the inferences and conclusions that are drawn? Has the researcher accounted for other legitimate explanations for the results if they exist?

One type of internal validity is *instrument validity*, which is not the same as reliability. Instead, it examines the degree to which the instrument legitimately measures what we say it measures. Instrument validity can be established in one of three ways. We can compare our instruments to other similar, but already established measures (criterion-related evidence for validity). If, for example, there already exists an academic test of knowledge on the criminal justice system whose reliability is proven, then we might administer that test and our own test to a group of students similar to those who'll participate in the study and compare the results. A second way of establishing instrument validity is by having outside experts judge the content of the instrument to determine whether its content truly represents the larger domain which is being considered. Known as content-related evidence for validity, this type of evidence might be exemplified in the LRE field by having several criminal justice experts review our test for the mock trial study to see if it is comprehensive enough. The third means of demonstrating instrument validity is called construct-related evidence. This type of evidence is used with instruments that measure difficult to observe variables, such as intelligence, creativity, or self-confidence. In this validation process, the researcher seeks to tie together certain psychological constructs, such as attitude and motivation or self-concept and hostile interpretation of others' actions. Whichever form of evidence is used to demonstrate the validity of our measures - whether it's proven instruments, judges, or constructs, it is important to do so in a context similar to that of the study (e.g. similar age group, gender mix, academic ability, etc.).

Internal validity relates to the experimental design, itself. How strong is the design in overcoming a variety of challenges to the inferences and conclusions that are drawn? Does the design eliminate the possibility of alternative explanations for our results? We've already seen that the instrument and extraneous variables can be threats to the study's validity. Other typical threats to internal validity include significant differences in the subjects assigned to different groups (known as subject selection); bias on the part of the subjects; maturation of the subjects during the study (that is, the fact that they have aged or that significant time has passed may strongly influence the results);

loss of subjects during the study (known as subject attrition or mortality); subjects changing just because they are a part of the study (they feel special, important, left out, don't want to let the experimenter down, etc.) or changing because of what they have learned or heard from other subjects (known as diffusion of treatment); and subjects changing because they have taken a pretest (from which they may have learned something new or been alerted to certain types of information). Also, what the experimenter says or does may affect subject responses; the experimenter may have a bias which influences observations or interpretations; or the number of times the treatment is tested on different subject or subject groups may be insufficient to draw conclusions other than that the results happened by chance (known as a treatment replications threat).

When it comes time to analyze the results of a quantitative study, the research could suffer from yet another type of validity issue - *statistical conclusion validity*. In quantitative research, standard statistical tests are used to analyze the **statistical** significance of the results. (It is important to differentiate between statistical significance and real-world significance; a study's results may be statistically significant but useless in the real world.) A lay person might ask why we bother with the statistics at all? Why not just look at the results and, if one is more than the other, then just say so? (This is referred to as descriptive statistics.) However, there is an important justification for using statistical analysis. Such an analysis enables us to infer that our results have meaning beyond our specific study. Otherwise, we would not know whether or not our results are just a fluke or a matter of chance. Further, no one is likely to care about our results unless we can somehow demonstrate that these results are likely to happen again and again with groups similar to those who were studied Here, then, is where the field of inferential statistics - which is based on the laws of probability - comes in. Using various statistical tests we can determine how likely it is that our study's results will be true for the larger population from which our sample came. We can also determine the degree to which we can be certain that this conclusion is true. The degree of certainty about the conclusion is referred to as *level of significance*. You will see this reported as $p \leq .10$ or .05 or .001. A p of .05 or less is generally considered to be statistically significant.

There are many types of statistical procedures used for various types of experiments. Failure to use the most appropriate procedures for the particular experimental design can legitimately lead others to conclude that alternative explanations have not been eliminated. While it is not critical for you to understand all of these types of procedures (and it is far beyond my capability to explain them), it is helpful to at least be able to recognize these terms and the fact that they are statistical tests when you read research reports. Typical statistical procedures you will see described in research reports include chi-square, regression analysis, t-test, analysis of variance or multivariate analysis of variance (ANOVA, MANOVA), and analysis of covariance or multivariate analysis of covariance (ANCOVA, MANCOVA).

It is also helpful to recognize that all statistical tests are built on certain assumptions which, if not met in the experiment's design, also threaten the validity of the statistical conclusions. Statistical procedures assume that the sample which is studied is drawn from a population which is normally distributed. In other words, the degree to which a particular characteristic appears in a particular population will be consistently distributed along a normal or bell curve. (This means that most of the population displays the characteristic to an average degree and is clustered around the middle "hill" portion of the curve, while much smaller portions of the population display the characteristic to a much lesser degree or to a much higher degree as represented by the two "valleys" or tails of the curve.) The larger the population from which the subjects are drawn, the more likely it is that this is a normally distributed population.

Another assumption behind statistical tests is that subjects have been randomly selected and randomly assigned to test groups. The randomness is critical because it assures that we have *variability* or that any variations that typically appear in the population also appear in the sample. Because of the randomness, it is also assumed that the level that any characteristic varies in a group will be similar across all groups in the experiment.

Finally, it is important for the number of subjects or subject groups not to go below critical levels in order for the conclusions to be valid. If a whole group of people, like a class or a reading group, receives the treatment at one time, this is considered one subject group or one *treatment replication*. The conventional wisdom for sample size is 15 treatment replications or subjects for experiments and 30 for non-experimental studies. Now that you know all of this, you should be wary of the validity of results when studies are conducted on small numbers of subjects or on subjects not randomly selected from the larger population.

Even if the study is internally valid - that is, the results are statistically significant and there are no significant alternative explanations for the results, it may not have *external validity*. This refers to our ability to generalize about the results or to apply them to more generalized populations. For example, results from studies conducted on populations that only represent one socioeconomic status (SES), such as the very poor, are not necessarily generalizable to middle and high SES groups. Similarly, studies of in-school 12th graders may not be generalizable to dropouts, or responses by juvenile delinquents may not be generalized to the larger teenaged population.

Credibility of Qualitative Research

There are issues of credibility on the qualitative side of research, as well. However, they have a very different orientation. Here, the major questions are the integrity and validity of the study, the accuracy of the findings, and the credibility of the researcher.

The purpose of qualitative research is to describe a phenomenon, not to explain what caused it or to make generalizations about it to other groups. Therefore the approach to conducting qualitative research is very different from that for its counterpart. Sample sizes are small and *purposeful*. That is, a few highly representative or informative cases are selected for in depth study rather than large numbers. The study is conducted in the natural setting, and the report is very detailed in describing the context specifics of the phenomenon being studied. Also, the study uses multiple means (methods and resources) of generating information on the question being investigated. It characteristically involves the researcher conducting detailed observations or interviews. The researcher is expected to get much more involved with those who are being studied (called *participants* not subjects) since this involvement is considered important to knowing what to ask and how to ask it. Scientific objectivity is, therefore, out of the question. Also, the researcher often becomes the instrument of data collection, so credibility of the results often hinges on the background, qualifications, and style or approach of the researcher.

When reading qualitative research, ask how rigorous the researcher has been in gathering the data and *triangulating* it or using multiple methods for such data gathering and analysis (i.e. repeated observations; interviews coupled with surveys and document or record reviews; interviews of people from varying perspectives; using different people to analyze the data). Triangulation ensures an information richness, accuracy, reliability, and validity. With regard to the researcher, ask if he or she has adequately accounted for any personal or professional information which may have affected data collection, analysis, and interpretation? Are there any biases on the part of the researcher or evaluator? Has that individual changed? Has that individual been properly prepared and trained? Have the participants had or were they likely to have had strong reactions to the researcher which may have biased the way in which they acted or responded? Finally, ask if there is ample data to support any conclusions that are drawn and if the researcher has adequately considered negative cases or exceptions to the trends that appear in the data.

Case Study: Analysis of A Well Known LRE Experimental Study

Many of us are familiar with the results reported from a federally funded, 3-year, national study conducted in the early 1980's. In fact, that study's results have served as our flagship in the relatively foggy void of studies generated by our field over the last twenty years. Unfortunately, some of the methods used in the study raise serious questions about the internal and external validity of its reported results. The purpose of this brief case study is not to denigrate the research so much as it is to offer a familiar example with which we might apply some of the principles described earlier.

There were several stages to this study, with various approaches used at each stage. In the first year's small study of LRE and control classes at ten sites, researchers observed seven of the LRE classes and developed predictions as to which ones would result in greater student knowledge of the law and greater improvements in self-reported attitudes and behavior toward the law. In all ten sites, students were pre- and posttested using a 41-item measure developed by the group conducting the study and scored independently of the classroom observations. Sites for the study were not randomly selected (instead, they were selected based on previous involvement and interest in LRE); nor was student assignment to LRE and control classes random. Overall test scores for LRE classes improved in four classes, declined in four classes, and remained the same in two. The predictions of success, however, were found to favorably compare with those groups who showed the most gains on the test, and it was out of this study that the "six prescriptions for successful LRE instruction" were born.

The second year's larger study of 30 sites across the nation again used observations and the 41-item measure of student knowledge and self-reported attitude

and behavior. Again, sites were chosen because of their involvement in LRE, and most LRE and control classes had students who had not been randomly assigned, thus resulting in a lack of equivalence between groups. At one junior high school in Colorado, there were six randomly assigned classes (3 LRE, 3 control). Researchers report that there were 14 out of 32 (44%) possible main effects (statistical analysis procedure) for the LRE classes, and all were in a favorable direction.

The third year's study focused only on the Colorado junior high school where they could get a true experimental design with random subject assignment. Teachers received summer LRE training (enhanced by the study's earlier findings of what leads to successful LRE), and eleven classes of students at the school received either LRE or civics during the fall semester's instruction. Out of 41 possible measured effects of LRE (related to achievement, delinquency theory, and delinquent behavior), 24 (59%) showed statistically significant favorable effects for LRE. The conclusion was that LRE in that school "had dramatic favorable impact on the students who participated" (Johnson, 1984, p. 11) and that the program had reduced the students' delinquent behavior.

The first question about the 3rd year study that jumps to mind regards the conclusions themselves. Only 24 out of 41 or three fifths of the overall measures and less than one third of the specific delinquent behavior measures showed LRE to have a statistically significant impact. Another way of looking at the results is that there was no significant impact on 40% of the responses! (Results from the 2nd year were even weaker.) This is hardly a "dramatic favorable impact" and should cause you to wonder about the bias of the researchers.

Beyond the questionable conclusions, the study, itself, had several challenges to internal and external validity. For example, sites were not selected randomly for participation in any year of the study. Sites were selected because they were already involved in LRE, and so the results could only be compared to sites with similar previous involvement. (In other words, we have no idea how the LRE would have influenced students and teachers in a school where LRE was a brand new phenomenon.) During the first two years, subjects were not randomly assigned, so accurate reliable comparisons between the experimental and control classes could not be made. We just can't know if the results were caused by the characteristics of who took part in the LRE and control classes or by the approach, itself. While this problem was solved in the third year, there were only eleven treatment replications - not enough for any *valid* statistical conclusions. Further, no evidence of independent tests for reliability was given for either the achievement portion of the student test or the attitude portion of that measure. Nor was any corroborating data (such as teacher or principal reports or police reports) given to support the students' self-report of delinquent activity. Finally, generalizability or external validity is highly questionable given the limited number of treatment replications offered in this particular portion of the study and the likelihood that the school's population only represents one socio-economic status rather than all (no data provided).

While these are not all the questions that have been raised about this study, it is easy to see from what little you've learned from this article that there are some very serious questions about the study's internal and external validity. The study's results are, at best, encouraging - not conclusive. We cannot rely on these particular results as being truth, and we should be careful to look with a critical eye at the results of *all* studies prior to adopting their fiats or promoting the products which they evaluate. Further, we can never expect to be taken seriously as a field by the larger educational community and its decision makers, until we generate strong and valid data to support what we believe and assert to be true about LRE.

Sources

Chirot, D. (1986). Social change in the modern era. New York: Harcourt Brace Jovanovich.

Cook and Campbell. (1979). Quasi-experimentation: Design and analysis issues for field settings. Chicago: Rand McNally; pp. 37-84.

Johnson, G. (1984). When LRE is a deterrent to delinquency Paper presented at Rocky Mountain Regional Conference of the National Council for the Social Studies, Phoenix. (ERIC Document Reproduction Service No. ED 252 459)

McMillan, J.H. (1992). Educational research, fundamentals for the consumer. New York: HarperCollins.

Muraskin, L.D. (1993). Understanding evaluation, the way to better evaluation programs. [A publication of Westat, Inc., developed with funds from the U.S. Department of Education, contract number LC89089001].

Patton, M.Q. (1990). Qualitative Evaluation and Research Methods, (2nd ed.). Newbury Park, CA: Sage.

Shaver, J.P. (1984). LRE project: A methodological critique. Paper presented at Rocky Mountain Regional Conference of the National Council for the Social Studies, Phoenix. (ERIC Document Reproduction Service No. ED 252 460)