State of New Mexico
LEGISLATIVE EDUCATION STUDY COMMITTEE

State Capitol North, 325 Don Gaspar, Suite 200
Santa Fe, New Mexico 87501
PH: (505) 986-4591 FAX: (505) 986-4338
http://legis.state.nm.us/lcs/lesc/lescdefault.asp

June 28, 2007

MEMORANDUM

TO: Legislative Education Study Committee

FR: Kathleen Forrer K. F.

RE: STAFF REPORT: ASSESSING EDUCATIONAL RESEARCH

As the committee is aware, during this interim there will be a change in the way issues will be presented. As opposed to the practice in recent years, the majority of the presentations to the Legislative Education Study Committee will be the responsibility of the assigned committee analyst(s), with outside experts serving primarily as resources. The staff's job will be to help the committee determine what the appropriations recommended by the committee to support public education in New Mexico have accomplished. To do that, staff will be reviewing research at the national and state level, as well as talking to a educators in the field, to students, and to other stakeholders.

Why should one even be concerned about what is going on at the national level? Although New Mexicans often kid themselves about being different from the rest of the nation, our state is not really that much different. Other states have students who live in poverty, students whose first language is not English, and schools that cannot make AYP (adequate yearly progress). States can learn from each other and from the research that shows how to address common problems. As Benjamin Franklin once said, "We must all hang together, or assuredly we shall all hang separately."

The Problem

In the introduction to *Estimating Causal Effects: Using Experimental and Observational Designs*, published by the American Educational Research Association, the authors reiterate a theme that may sound familiar:

> Among educational leaders and policymakers there has been increasing concern regarding the need for scientifically based evidence on which to base funding decisions for specific educational programs and practices. This concern is fundamentally about having better evidence for making decisions about what programs and practices do or do not work. The need for such evidence leads to causal questions, such as whether particular programs and practices improve student achievement, social development, and educational attainment. ...Nonetheless, among researchers there is a lack of clarity regarding which designs, methods, and analytical approaches are most appropriate for making causal inferences. (p. 1)

If the authors are correct, and even the researchers are confused, how can practitioners and policymakers distinguish research that is useful from the not particularly helpful or even the just plain bad?

Educational Research

In general, educational research may be classified into two basic categories: (1) empirical research that is based on the systematic gathering of information through observation and/or experimentation, or (2) non-empirical research that is based on some other method of securing information, such as a review of historical records. Although both types of research help practitioners and policymakers develop a greater understanding of the "art" of education, only empirical research can show "what works." When the *No Child Left Behind Act of 2001* (NCLB) uses the term "scientifically based research," it is referring to empirical research.

Experimental and quasi-experimental research designs are cited in the literature as the types of empirical, or scientific, research most likely to provide information regarding the efficacy of programs in education. According to the Institute of Education Sciences, the so-called "gold standard" for research in education is the randomized controlled trial, a type of study in which individuals are randomly assigned to both an intervention group and a control group. In *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide*, the Institute provides the following illustration:

> For example, suppose [a researcher] want[s] to test, in a randomized controlled trial, whether a new math curriculum for third-graders is more effective than [the] school's existing math curriculum for third-graders. [The researcher] would randomly assign a large number of third-grade students to either an intervention group, which uses the new curriculum, or to a control group, which uses the existing curriculum. [The researcher] would then measure the math achievement of both groups over time. The difference in math achievement between the two groups would represent the effect of the new curriculum compared to the existing curriculum.

True experimental research is rarely used in education for two reasons—practicality and ethics. Practically speaking, rarely does a researcher have the opportunity to randomly assign students to different groups. And ethically, although in a "true" scientific experiment, the control group would not be provided any treatment, or in this case, math program, no educator would argue for withholding education to any student. The National Research Council of the National Academies of Science flatly states that, "in studying the effectiveness of an educational program, then, comparisons must almost always be made to standard or existing practice."

An alternative to true experimental design is quasi-experimental design in which participants are not randomly assigned to comparison groups:

> Suppose [a researcher] want[s] to use a comparison-group study to test whether a new mathematics curriculum is effective. [The researcher] would compare the math performance of students who participate in the new curriculum ("intervention group") with the performance of a "comparison group" of students, chosen through methods other than randomization, who do not participate in the curriculum. The comparison group might be students in neighboring classrooms or schools that don't use the curriculum, or students in the same grade and socioeconomic status selected from state or national survey data. The difference in math performance between the intervention and comparison groups following the intervention would represent the estimated effect of the curriculum. (*Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide*)

The effect would only be estimated because without randomization, the researcher cannot be sure that something other than the new curriculum, such as the preference of the students for a particular teacher, impacted the results. One way for a researcher to mitigate this problem is to match the two groups as closely as possible on aspects other than the curriculum, such as age and socioeconomic status. Another method which has been used more and more often in recent years is regression discontinuity, a quasi-experimental research design in which participants are assigned to program or comparison groups solely on the basis of a cutoff score on a pre-program measure. (This is the method being employed by the National Institute for Early Education Research to evaluate New Mexico PreK.)

In addition to experimental and quasi-experimental research designs, there are other types of research that may prove helpful—and are usually easier to read—but which are not generally viewed as providing evidence of cause and effect. These include meta-analysis, which is a comprehensive review of past studies on a specific topic; and descriptive research, which attempts to answer questions about what, how, or why something is happening:

- Meta-analysis is not simply a review of the available literature on a given program. Rather, meta-analysis is the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings: "In meta-analysis, research studies are collected, coded, and interpreted using statistical methods similar to those used in primary data analysis. The result is an integrated review of findings that is more objective and exact than a narrative review." (*A User's Guide to the Meta-analysis of Research Studies*, p. 2)

- Descriptive research, such as correlational research, lacks the random assignment and manipulation of a treatment present in experimental research. As such, the most that descriptive research can uncover is the correlation or association of factors; it cannot reveal an actual causal relationship. For example, a researcher might examine whether there is a relationship between the level of experience of the teachers in a particular school and the achievement scores of their students. Even if the researcher finds that there is a positive relationship between the number of years a teacher has been in the classroom and the achievement scores of the teacher's students, the researcher cannot claim that experience has "caused" the higher scores. There are too

many other factors unaccounted for, such as the prior educational experiences of the students.

Assessing Research

How, then, can a non-researcher assess educational research, which is often presented in terms only a researcher could love? In *What is Scientifically Based Research? A Guide for Teachers*, the National Institute for Literacy poses three questions that can help the reader distinguish between research that "confirms the effectiveness of an instructional practice and research that does not":

1. Has the study been published in a peer-reviewed journal or approved by a panel of independent experts? If a study has undergone peer review, it has been scrutinized by scientists in the relevant field of specialization. Peer review provides a baseline of quality control because it exposes ideas and experimentation to examination and criticism by other researchers. Its absence should raise doubt about the quality of the research. Presentations at education conferences that make claims about specific educational practices should also be held to this standard. NOTE: Not all journals are peer-reviewed. For example, the *American Educational Research Journal*, the *Journal of Educational Psychology*, and *Reading Research Quarterly* conduct peer reviews and contain empirical evidence about teaching techniques. In contrast, *Phi Delta Kappan*, although a valuable source of opinion—and great cartoons, does not publish peer-reviewed original research.

2. Have the results of the study been replicated by other scientists? To be considered scientifically based, a research finding must be presented in a way that enables other researchers to reach the same results when they repeat the experiment.

3. Is there consensus in the research community that the study's findings are supported by a critical mass of additional studies? A single experiment rarely decides an issue, supporting one theory and ruling out all others. Issues are most often decided when the community of scientists in a field comes to agreement over time that sufficient evidence has converged to support one theory over another. Scientists do not evaluate data from a single, perfectly designed experiment. They evaluate data from many experiments, each containing some flaws but providing part of the answer.

When actually assessing the validity of a particular research study, there are additional questions to keep in mind:

1. What is the research question? A descriptive study might ask if teacher professional development has a positive association with student achievement. An experimental or quasi-experimental study might ask if providing teachers with professional development in reading causes their students to have higher achievement in reading.

2. Does the research design match the research question? In other words, did the researcher design the study in such a way that the question can be answered? Remember that showing that one variable occurs in association with another (correlation) does not show cause.

3. How was the study conducted? The research report should include sufficient details that the study can be replicated, including information regarding who the participants were

and how they were chosen; precisely what treatment was applied and how it was implemented (e.g., did a single teacher teach both the treatment and control groups); what data were collected and by what means (e.g., through a survey or assessment instrument); and how were the data analyzed.

4. Are there rival explanations for the results? If the results can be used to support more than one conclusion, the conclusion offered by the researcher must be viewed cautiously.

The following table, adapted from *A Policymaker's Primer on Education Research*, provides a means of working through these last four questions logically:

Summary: Questions to ask about research validity
Research question and design
Does the *research design* match the *research question*?
Participants
What was the basis for selecting the participants?
How were the participants assigned to groups?
Do participant selection and assignment follow the research design?
Are the results influenced by *extraneous characteristics* of participants and contexts?
Treatment
What is the *operational definition* of the *treatment*?
Is the definition valid?
Was the treatment implemented as planned?
Data collection
What is the operational definition of the *dependent variable*?
Why were the *data* selected?
Was there *pilot testing* or field testing of the *instruments*?
Are the data-collection instruments valid and reliable?
Was there training for data collectors?
What was the *response rate* for questionnaires?
Data analysis
Quantitative:
Were non-significant results (i.e., $p > .05$) discussed as if they were *significant*?
How did *sample size* influence the results?
How did *variability* in the scores influence the results?
Was an *effect size* reported?
Qualitative:
How were the data *coded*?
What procedures were used to *verify* the coding?
Rival explanations
Did any of the following occur that might have affected the results and were not ruled out?
Conclusions about score gains from a treatment without a *pretest*?
Conclusions about score gains from a treatment without a *control* or *comparison* group?
Bias in assigning participants to different comparison groups?
Loss of participants from the study sample?
Spillover of the treatment into the control or comparison group?
Influences from an event that occurred between a pretest and *posttest*?
Effects from participant practice on the measuring instrument?
Extreme scores that could become less extreme on the posttest regardless of treatment?

In the end, it is a matter of balancing all the criteria of usefulness in a way that reflects the local circumstances involved in a particular policy decision. First, it is necessary to determine if the research is empirical and the researcher's conclusions are valid. Next, policymakers must decide how much weight to give to the other criteria of research usefulness. The costs of policy decisions and potentially harmful effects are factors that should always be considered in addition to the information provided by the research. When there is little or no useful research on an education topic related to a policy decision, and a change is needed or mandated, then policymakers should find ways to fund the necessary research. In the long run, a policy decision that is informed by research might be far less costly than one that is uninformed. (*A Policymaker's Primer on Education Research: How to Understand, Evaluate and Use It*, p. 36)

Help (I still don't understand anything I just read!)

Research has its own language that is difficult for the non-researcher to penetrate. Fortunately, there are sources of information, many of which are available on the web, that do much of the work for those who may be short of time and infinite patience:

- According to *Education Week*, the following are websites that "vet the research evidence on promising educational programs and practices," thus providing policymakers and educators with easily understandable information on what works:

  - *Best Evidence Encyclopedia* distills evidence from several different research groups on promising practices and interventions in education. It is produced by the federally funded Center for Data-Driven Reform in Education at Johns Hopkins University in Baltimore. (www.bestevidence.org)

  - *Blueprints for Violence Prevention* at the University of Colorado in Boulder identifies programs found to be effective in reducing violent crime, bullying, delinquency, and substance abuse. (www.colorado.edu/cspv/blueprints)

  - *Child Trends What Works* site summarizes both the research and the professional wisdom on promising programs aimed at improving the lives of children and youths. (www.childtrends.org)

  - *Comprehensive School Reform Quality Center* at the Washington-based American Institutes for Research (AIR) posts research reviews on school-wide improvement efforts and educational service providers. (www.csrq.org)

  - *International Campbell Collaboration* systematically reviews research evidence from around the world on social, behavioral, and educational interventions. (www.campbellcollaboration.org)

  - *Poverty Action Lab* at the Massachusetts Institute of Technology scientifically evaluates anti-poverty programs in developing countries. (www.povertyactionlab.com)

- *Promising Practices Network* is operated by the RAND Corp., a think tank based in Santa Monica, Calif. It highlights research-based programs shown to be effective at improving outcomes for children, youths, and families. (www.promisingpractices.net)

- *Social Programs That Work*, produced by the Washington-based Coalition for Evidence-Based Policy, is a listing of interventions in education and various social science areas that have been evaluated through randomized studies. (www.evidencebasedprograms.org)

- *What Works Clearinghouse*, funded by the US Department of Education and run by the AIR, contains research reviews and effectiveness ratings on a wide range of educational programs and practices. (www.whatworks.ed.gov)

- In addition, the American Educational Research Association maintains a website with links to a number of electronic journals that are scholarly, peer-reviewed, full text, and accessible without cost. (http://aera-cr.asu.edu/ejournals/)

- The central focus of the Education Finance and Accountability Program (EFAP) at the Center for Policy Research, Syracuse University, is on the tax and state aid programs that fund elementary and secondary education in the United States and on policies to promote efficiency and accountability in school districts. This website describes current EFAP research projects, provides access to working papers, and describes publications from past projects. (http://www-cpr.maxwell.syr.edu/efap/index.htm)

- Finally, two written guides designed to help non-researchers evaluate educational research are *A Policymaker's Primer on Education Research: How to Understand, Evaluate and Use It*, issued in February 2004 by Mid-continent Research for Education and Learning (McREL) and the Education Commission of the States (ECS), and *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide*, issued in December 2003 by the Institute of Education Sciences, US Department of Education.

A Glossary of Selected Educational Research Terms is attached.

GLOSSARY OF SELECTED EDUCATIONAL RESEARCH TERMS

Types of Research

case study: A data collection method in which a single person, entity or phenomenon is studied in depth over a sustained period of time and through a variety of data.

Example: A researcher conducts a yearlong case study of a school district that was awarded a grant to improve teacher quality. The researcher documents the processes used to implement the grant, interviews teachers and administrators, observes staff development, and measures student achievement before and after the grant was awarded.

correlational research: A type of research that seeks to establish an association or correlation between two or more variables. The fact that two or more variables are associated does not necessarily mean that one is a cause of the other(s).

correlational research design: A research design in which data are collected to describe the statistical association between two or more variables.

Example:

Bivariate correlation: In School District X, a researcher collects data on beginning teachers' scores on the state licensing test (variable 1) and data on the achievement gains of each teacher's students (variable 2). The researcher then uses correlational statistics to measure the association between the two variables.

Multivariate correlation (also referred to as multiple regression): In School District X, a researcher collects data on beginning teachers' scores on the state licensing test (variable 1), the number of college courses that each teacher completed in mathematics (variable 2), the amount of time that each teacher spent in school-based field experiences prior to certification (variable 3), and the achievement gains in mathematics by each teacher's students (dependent variable). The researcher uses multiple regression statistics to measure the association between the three teacher variables and student achievement gains and to estimate student achievement gains based on the contribution of each of the teacher variables to that association.

descriptive research: A type of research that has the goal of describing what, how or why something is happening.

descriptive research design, comparative: A research design in which data are collected to describe and compare two or more groups of participants or entities.

Example: A researcher identifies high-poverty schools in the state that have either high or low student achievement. The researcher describes the alignment or match between each school's curriculum and state standards and compares the high- versus the low-achieving schools to determine whether the degree of alignment is different.

empirical research: Research that seeks systematic information about something that can be observed in the real world or in the laboratory.

ex post facto research: Descriptive research that examines the influence of a preexisting independent variable or treatment.

Example: A researcher conducts a study to compare two reading programs. The participants are students in School A, which has been using Reading Program A for three years, and students in neighboring School B, which has been using Reading Program B for three years. This study is ex post facto because the research concerns effects from a preexisting treatment.

experimental research: A type of research that has the goal of determining whether something causes an effect.

experimental (true) research design: A research design in which (1) an independent variable is directly manipulated to measure its effect on a dependent variable, and (2) participants are randomly assigned to different groups that receive different amounts of the independent variable. (Also referred to as randomized field trials or randomized controlled trials.)

meta-analysis: A comprehensive, systematic, quantitative review of past empirical research studies on a specific topic. Most meta-analyses examine only quantitative studies. Effect-size statistics are calculated to produce an overall conclusion about the various studies on the topic.

Example: A researcher conducts a meta-analysis of computer-assisted instruction in reading. The researcher examines 40 studies and calculates an overall effect size of $d = .25$, indicating a small positive effect of computer-assisted instruction on reading achievement.

pretest-posttest research: Research in which participants take a pretest that measures the dependent variable prior to the administration of a treatment and a posttest that measures the dependent variable after the treatment is completed. The most valid approach to implementing pretest-posttest research is to randomly assign participants to two or more groups, one of which receives the treatment. The pretest-posttest difference scores are then compared for the groups.

Example: A researcher randomly assigns middle school students to participate in either an inquiry-based science unit or a traditional science unit. The students complete a test on problem solving before and after the unit. Because the problem-solving skills of the students in the inquiry-based group improved more than those of the students in the traditional group, the researcher concludes that inquiry-based units facilitate problem-solving skills.

qualitative research: Research in which the data are narrative descriptions or observations. In most qualitative research, there is an emphasis on the influence of context.

Example: A researcher observes how teachers deliver instruction related to different reading curricula in two different schools. The researcher also interviews the teachers to understand their approaches to the different curricula and how their approaches might be influenced by school characteristics.

quantitative research: Research in which the data are numbers and measurements. In quantitative research, there is an emphasis on control of the variables in the study.

Example: A researcher randomly assigns students to different reading curricula. At the end of the school year, the researcher examines the students' scores on a reading achievement test to determine whether the different curricula had different effects on reading.

quasi-experimental research design: A research design in which (1) an independent variable is manipulated to measure its effects on a dependent variable, and (2) participants are not randomly assigned to comparison groups.

Example: A researcher assigns 15 teacher preparation candidates who have a seminar on Wednesdays to participate in eight weeks of student teaching. The researcher assigns 15 teacher preparation candidates who have a seminar on Tuesdays to participate in 16 weeks of student teaching. After the candidates graduate, the researcher compares their scores on a performance-based teacher-licensing test. The amount of student teaching is the independent variable, and candidate performance on the teacher-licensing test is the dependent variable. The researcher does not randomly assign candidates to the comparison groups. As a result, differences between the groups' performance on the test could be due to the amount of student teaching or due to other characteristics of the teacher candidates. The researcher should demonstrate that the candidates in the two groups do not differ in characteristics that are related to teaching performance.

randomized trials: A "true experimental" research design in which (1) an independent variable is directly manipulated to measure its effect on a dependent variable (i.e., the treatment trial), and (2) participants are randomly assigned to different groups that receive different amounts of the independent variable (i.e., the treatment). (Also referred to as randomized field trials and randomized controlled trials.)

regression-discontinuity design: A quasi-experimental research design in which participants are assigned to program or comparison groups solely on the basis of a cutoff score on a pre-program measure. Unlike its randomized or quasi-experimental alternatives, the regression-discontinuity design does not require potentially needy individuals to be assigned to a no-program comparison group in order to evaluate the effectiveness of a program. Those conducting the study must be able to show that individuals just above and below the cutoff point for program entry have similar characteristics and probabilities of being accepted into the program.

Example: A researcher wants to compare two groups of children who select, and are selected by, a state pre-kindergarten program, taking advantage of the stringent age (birth date) cutoff that states use in defining enrollment eligibility to define groups. This design is conceptually easier to understand by taking the extreme case: consider two children who differ only in that one was born the day before the age cutoff and the other the day after. When both are about to turn 5 years old the slightly younger child will enter the preschool program and the slightly older child will enter kindergarten having already attended the preschool program. If both are tested at that time, the difference in their scores can provide an unbiased estimate of the preschool program's effect. Obviously, if only children with birthdays one day on either side of the age cutoff were included in a

study, the sample size would be unreasonably small. However, the approach can be applied to wider age ranges around the cutoff. In fact, all children entering kindergarten from the state preschool program, and all children beginning preschool in the same year can be included in the study using regression-discontinuity statistical techniques.

Other Terms

confidence interval: A range of values that indicates the confidence or probability of observing a particular score or value in a population, usually expressed as standard deviation units above and below the mean. The wider the interval, the greater the confidence or probability that a particular value will be observed.

Example: Based on a random sample of 4th-grade reading scores, a researcher calculates the following 90% confidence interval for the mean of the population of 4th-grade reading scores: 67 ± 3.2. This indicates there is a 90% probability that the mean reading score of the population is between 63.8 and 70.2.

correlation coefficient: A number that indicates the strength and direction of the statistical association between two or more variables. Correlation coefficients vary between -1.00 and $+1.00$. The higher the numerical value, the stronger the association. A correlation of 0.00 indicates the absence of an association. A positive sign means that as one variable increases, so does the other. A negative sign means that as one variable increases, the other variable decreases.

Example: A correlation coefficient of $+.63$ between the number of education courses and teacher test scores means that the more education courses that a teacher candidate completed, the higher the test score. A correlation of $-.63$ means that the more education courses that a teacher candidate completed, the lower the test score. Neither correlation coefficient, however, can support the existence of a causal relationship between courses and test scores because correlation is not causation.

error: Inaccuracies in implementing a research study, including during sampling, treatment delivery, data recording, or data analysis. Errors increase the variability of the data and threaten the validity of research conclusions.

inferential statistics: Statistics used to make inferences about a population based on the scores obtained from a sample. Inferential statistics are based on the mathematics of probability theory.

Likert Scale: A response scale in which participants respond to questionnaire items about their beliefs and attitudes by indicating varying degrees of intensity between two extremes such as like/dislike and agree/disagree.

random assignment: The assignment of participants to comparison groups using chance procedures so that every participant has the same probability of being selected to a group.

random sample: A sample that is randomly drawn from a population so that each member of the population has an equal probability of being chosen for the sample.

regression analysis: A statistical technique that uses the relationship between two variables, X and Y, to predict the value of X based on observations of Y.

regression analysis, multiple: A statistical technique that determines the linear association between a set of predictor variables and a dependent variable and identifies the combination of predictor variables that best estimates the dependent variable (also referred to as the criterion variable).

Example: In School District X, a researcher collects data on beginning teachers' scores on the state licensing test (predictor 1), the number of college courses in mathematics that each teacher completed (predictor 2), the amount of time spent in school-based field experiences prior to certification (predictor 3), and the achievement gains in mathematics by each teacher's students (criterion variable). The researcher uses multiple regression statistics to measure the association between the three teacher variables and student achievement gains and to estimate student achievement gains based on the contribution of each of the teacher variables to that association.

regression toward the mean: The tendency for extreme scores to move toward the average or mean score when a test or other measure is repeated. Regression effects threaten the validity of research conclusions in studies in which participants are chosen because of their extreme scores on a measure.

Example: Researchers often study schools in which students have extremely low achievement scores. If these students improve their achievement following a treatment or intervention, the improvement could be due to regression effects instead of treatment effects. In such studies, it is important to have comparison schools of students who also have extremely low achievement scores but who do not receive the treatment.

statistically significant: A result that has a low probability (e.g., 5%) of occurring by chance. Because it is unlikely that a statistically significant result has occurred by chance, the result is said to reflect non-chance factors in the study, such as the effects of a treatment.

test, criterion-referenced: A test for which a score is interpreted by comparing it to levels of performance established for the test by professionals in the field that the test addresses.

Example: Scores on the New Mexico Standards Based Assessment Program are assigned to the following categories based on the proficiency that students demonstrate in relation to state content standards: beginning step, nearing proficiency, proficient, and advanced.

test, norm-referenced: A test for which a score is interpreted by comparing it to the scores of a comparison or norming group of persons who took the test. The similarity of an individual to the persons in the comparison group influences the accuracy of interpretation.

Example: The SAT, which students take to gain admission to institutions of higher education, is a norm-referenced test. A score on the SAT is interpreted with reference to the scores of other students who took the test. A score of 500 on the SAT is considered average because that is the average score of the comparison or norming group of students.

validity, content: The degree to which the items on a measuring instrument (e.g., test or questionnaire) adequately cover the content that the instrument is designed to measure.

validity, construct: The degree to which variables in a research study are considered by the education and research communities as acceptable representations of the constructs that the study concerns.

Example: One-on-one instruction is a valid representation of the construct of tutoring, while whole-class instruction would not be considered valid. Student scores on a standardized mathematics test are a valid representation of the construct of student achievement, while student scores on a survey about attitudes toward school would not be considered valid.

validity, face: The extent to which the contents of a test or procedure "look like" they are measuring what they are supposed to measure.

validity, internal: The degree to which the conclusions of a research study are supported by evidence and can be trusted.

variable, dependent: The variable that is measured in a study. In an experimental research study, the dependent variable is affected by the independent variable. In a correlational research study, the dependent variable is associated with one or more other variables.

Examples:

Experimental research study: A researcher randomly assigns teachers in a large elementary school to receive one of three types of professional development: (1) a class on instructional strategies, (2) a training program on how to increase student motivation or (3) a teacher discussion group. The researcher measures the differences in achievement gains among the students of the three teachers. The dependent variable is student achievement gains.

Correlational research study: A researcher collects data on beginning teachers' scores on the state licensing test (variable 1) and data on the achievement gains of each teacher's students (variable 2). The researcher then uses the association between the two variables to estimate student achievement gains. The dependent variable is student achievement gains.

variable, independent: In experimental research, the variable that the researcher varies or manipulates to determine whether it has an effect on the dependent variable.

Example: As part of an experiment, a researcher randomly assigns teachers in a large elementary school to receive one of three types of professional development: (1) a class on instructional strategies, (2) a training program on how to increase student motivation, or (3) a teacher discussion group. The researcher measures the differences in achievement gains among the students of the three teachers. The independent variable is professional development.

variable, intervening: An unmeasured variable that is assumed to intervene between a treatment or independent variable and a behavior or dependent variable. Most intervening variables are internal and cannot be observed. Their existence is inferred based on external measures.

Example: Learning is an intervening variable because it cannot be observed but is assumed to occur between instruction and performance based on measures such as tests.

variables, extraneous: Variables in a research study that are not intended to influence the results or conclusions. Researchers use various methods to control the influence of extraneous variables.

Example: A researcher conducts a study of the effects of two different reading curricula on 1st-grade reading achievement. Extraneous variables in this study include students' verbal abilities and teachers' characteristics. The researcher needs to control the influence of these extraneous variables on achievement, possibly by having one teacher instruct both curricula and by randomly assigning students to the curricula.

Sources:

Barnett, W. Steven, Cynthia Lamy, and Kwanghee Jung. *The Effects of State Prekindergarten Programs on Young Children's School Readiness in Five States*. Princeton, NJ: The National Institute for Early Education Research, Rutgers University, December 2005.

Lauer, Patricia A. *A Policymaker's Primer on Education Research: How to Understand, Evaluate and Use It.* Denver, CO: Mid-continent Research for Education and Learning (McREL) and the Education Commission of the States, February 2004.

Schneider, Barbara, et al. *Estimating Causal Effects: Using Experimental and Observational Designs.* Washington, D.C.: American Educational Research Association, 2007.

Trochim, William M. *The Research Methods Knowledge Base*, 2nd Edition. Internet WWW page, at URL: <http://www.socialresearchmethods.net/kb/> (version current as of October 20, 2006).