1-1-1997

Action Research in Legal Education, 33 Willamette L. Rev. 383 (1997)

Paul T. Wangerin
John Marshall Law School

ACTION RESEARCH IN LEGAL EDUCATION

PAUL T. WANGERIN*

I. INTRODUCTION

Law school teachers constantly talk among themselves about the things they teach their students. Usually such talk is informal conversation, such as, "I teach students to think like lawyers"; "I teach students how to find and use the policy behind rules of law"; "I teach students how to identify and use the notion of the *prima facie* case"; and "I teach skills of analysis and synthesis."

Sometimes such talk moves beyond the informal stage, such as, "I used this teaching technique at my school, and it worked great."[1] Robert Gorman told a large audience at the 1001 Annual Convention of the Association of American Law Schools, "we are pretty good at teaching our students traditional skills of analysis."[2] However, legal educators who say such things fail to provide actual evidence to support their views. Most rely on personal opinions alone.

Ironically, legal educators constantly berate law students who rely solely on personal opinions and/or appeals to simple authority. Yet, when it comes to the effectiveness of educators' own teaching and that of their colleagues, most legal educators seem perfectly willing to rely solely on such information. Apparently legal educators exempt themselves from that which they require of their own students.

---

* A.B., University of Missouri, 1969; J.D., John Marshall Law School, with High Honors, 1978. Currently a Professor at the John Marshall Law School, Professor Wangerin has published extensively in the field of legal education.

1. A recent issue of the *Journal of Legal Education* provides several examples of this kind of report. *See, e.g.*, Margaret Z. Johns, *Teaching Professional Responsibility and Professionalism in Legal Writing*, 40 J. LEGAL EDUC. 501 (1990); James E. Moliterno, *The Legal Skills Program at the College of William and Mary: An Early Report*, 40 J. LEGAL EDUC. 535 (1990); Janice Toran, *Teaching Freedom of Information Law*, 40 J. LEGAL EDUC. 487 (1990). For similar comments in other law-related journals, see Paul T. Wangerin, *Skills Training in "Legal Analysis": A Systematic Approach*, 40 U. MIAMI L. REV. 409 (1986).

2. Gorman made this comment, incidentally, as a prelude to statements alleging that law professors are *not* particularly good at teaching students to write.

Of course, exceptions to this generalization may exist. Numerous people have attempted to gather empirical proof regarding the effectiveness of certain kinds of law school teaching.[3] For example, in 1987, David Leonard published empirical data that, he argued, demonstrated that his school's supplemental teaching program had a small but significant short-term academic impact. The program, offered to students in academic trouble, produced a general improvement in those students' sense of well-being.[4] Furthermore, in 1989, Charles Finke published empirical data suggesting that his school's supplemental teaching program, of-

3. *See, e.g.*, John D. Blackburn & Edward Niedzwiedz, *Do Teaching Methods Matter? A Field Study of an Integrative Teaching Technique*, 18 AM. BUS. L.J. 525 (1981); Harry G. Henn & Robert C. Platt, *Computer-Assisted Law Instruction: Clinical Education's Bionic Sibling*, 28 J. LEGAL EDUC. 423 (1977); Benjamin N. Henszey & Barry L. Myers, *Evaluation of "New" Teaching Methods for the Basic Business Law Course*, 15 AM. BUS. L.J. 132 (1977); Charles D. Kelso, *Programming Shows Promise for Training Lawyers: A Report on an Experiment*, 14 J. LEGAL EDUC. 243 (1961); Edward L. Kimball & Larry C. Farmer, *Comparative Results of Teaching Evidence Three Ways*, 30 J. LEGAL EDUC. 196 (1979); Willard D. Lorensen, *Concentrating on a Single Jurisdiction to Teach Criminal Law—An Experiment*, 20 J. LEGAL EDUC. 361 (1968); Peter B. Maggs & Thomas D. Morgan, *Computer-Based Legal Education at the University of Illinois: A Report of Two Years' Experience*, 27 J. LEGAL EDUC. 138 (1975).

For a discussion of a number of empirical studies of the effectiveness, or lack thereof, of various kinds of law teaching, see generally Paul F. Teich, *Research on American Law Teaching: Is There a Case Against the Case System?*, 36 J. LEGAL EDUC. 167 (1986). For other articles containing empirical data regarding the things students learn in law school, see ALFRED SMITH, COGNITIVE STYLES IN LAW SCHOOL (1979); David P. Bryden, *What Do Law Students Learn? A Pilot Study*, 34 J. LEGAL EDUC. 479 (1984); James M. Hedegard, *Causes of Career-Relevant Interest Changes Among First-Year Law Students: Some Research Data*, 1982 AM. B. FOUND. RES. J. 787; James M. Hedegard, *The Course Perceptions Questionnaire: Development and Some Pilot Research Findings*, 1981 AM. B. FOUND. RES. J. 463; James M. Hedegard, *The Impact of Legal Education: An In-Depth Examination of Career-Relevant Interests, Attitudes, and Personality Traits Among First-Year Law Students*, 1979 AM. B. FOUND. RES. J. 791; Robert R. Ramsey, Jr., *A Subcultural Approach to Academic Behavior*, 35 J. EDUC. SOC'Y 355 (1961); Charles John Senger, Learning Legal Reasoning in Law School: The Differences Between First and Third Year Students (1989) (Ph.D. dissertation, Michigan State University). Many empirical studies regarding the psychological development of law students, and stress in law students, are cited in Paul T. Wangerin, *Objective, Multiplistic, and Relative Truth in Developmental Psychology and Legal Education*, 62 TUL. L. REV. 1237 (1988).

4. David P. Leonard, *Personal and Institutional Benefits of Offering Tutorial Services to Students Experiencing Academic Difficulty*, 37 J. LEGAL EDUC. 91 (1987). Students on or near academic probation after the first semester of law school were paired during the second semester with students in the top quarter of the class. *Id.* at 92. The high-scoring students, under faculty supervision, individually tutored the poor-performing students. *Id.* at 92-93. This tutoring involved three to four hours each week, and consisted primarily of conferences about note-taking, outlines, and examinations. *Id.* at 93. Approximately 100 students participated in the program. *Id.* at 96.

fered to specially admitted minority students, had an extremely positive academic impact.[5] Then, in 1990, Stephen and Sherry Hartwell published empirical data suggesting that three different supplemental teaching programs their school offered produced little or no academic improvement for students.[6]

Unfortunately, most legal educators have neither the time nor the inclination to engage in the complex empirical research like that of Leonard, Finke, and the Hartwells. Furthermore, even if legal educators wished to gather evidence regarding the effectiveness of certain teaching techniques or educational programs, they would be dissuaded by the lack of a workable model for gathering such evidence. Standard statistical analysis is simply too difficult.

The forthcoming analysis describes "action research," a different model for generating evidence concerning teaching and program effectiveness in law schools.[7] Part I of the Article de-

---

5. Charles L. Finke, *Affirmative Action in Law School Academic Support Programs*, 39 J. LEGAL EDUC. 55 (1989). Finke describes a supplemental learning project that was similar to Leonard's in several ways. Finke worked during three successive years with small groups of specially admitted minority law students who were *expected* to perform poorly in school. *Id.* at 64-65. As noted earlier, Leonard worked with a group of students who already *had* performed poorly in school, a group that presumably included a significant number of specially admitted minority students. Leonard, *supra* note 4, at 92. Finke's program provided greater opportunity for supplemental learning assistance than did Leonard's. Compare Leonard, *supra* note 4, at 92, and Finke, *supra*, at 64-65. Finke's program included an eight-day summer orientation program; extensive exercises in reading, writing, and analysis; an introduction to and overview of the substantive law courses; ongoing tutorial help during the semester of the experiment; help and practice with examination skills; and enrollment in a special section of legal writing. Finke, *supra*, at 62-65. Both students and staff/faculty tutors offered tutorial assistance. *Id.* at 65.

6. Steven Hartwell & Sherry L. Hartwell, *Teaching Law: Some Things Socrates Did Not Try*, 40 J. LEGAL EDUC. 509 (1990). The Hartwells' project, although similar to Leonard's and Finke's, cites Finke only in passing and does not cite Leonard at all. The Hartwells provided several groups of law students with different kinds of supplemental learning assistance. They then analyzed whether each type of assistance produced positive academic gains. *Id.* at 511. They divided a large property class into four groups. *Id.* Three groups were offered hour-long weekly supplemental assistance—short objective quizzes, supervised outside-of-class discussion, and graded essay tests; the fourth group served as "controls." *Id.*

7. *See* WALTER BORG & MEREDITH DAMIEN GALL, EDUCATIONAL RESEARCH: AN INTRODUCTION (1989). For other recent books describing education research methodology, see also LORIN W. ANDERSON & ROBERT B. BURNS, RESEARCH IN CLASSROOMS: THE STUDY OF TEACHERS, TEACHING AND INSTRUCTION (1989); DONALD ARY ET AL., INTRODUCTION TO RESEARCH IN EDUCATION (4th ed. 1990); JOHN BEST, RESEARCH IN EDUCATION (1989); C.M. CHARLES, INTRODUCTION TO EDUCATIONAL RESEARCH (1988); LOUIS COHEN & LAWRENCE MANION, RESEARCH METHODS IN ED-

scribes action research and argues that educational research need not be left only to education research specialists or to those with advanced training in statistical analysis. Rather, almost any teacher or school administrator can conduct simple educational research. However, many people in the educational research community disagree. Part I also attempts to explain this disagreement.

Part II of the Article briefly describes several kinds of backward-looking research. The three models described are correlational research, causal-comparative research, and observational research. Part III shifts to experimental research and discusses problems involving the internal and external validity of the experiments. Part IV describes two kinds of experimental designs:

UCATION (3d ed. 1989); COMPLEMENTARY METHODS FOR RESEARCH IN EDUCATION (Richard M. Jaeger ed., 1988); EDUCATIONAL RESEARCH, METHODOLOGY AND MEASUREMENT: AN INTERNATIONAL HANDBOOK (John P. Keeves ed., 1988); DAVID R. KRATHWOHL, SOCIAL AND BEHAVIORAL SCIENCE RESEARCH: A NEW FRAMEWORK FOR CONCEPTUALIZING, IMPLEMENTING, AND EVALUATING RESEARCH STUDIES (1985); EMANUEL J. MASON & WILLIAM J. BRAMBLE, UNDERSTANDING AND CONDUCTING RESEARCH: APPLICATIONS IN EDUCATION AND THE BEHAVIORAL SCIENCES (2d ed. 1989); MARY W. OLSON, OPENING THE DOOR TO CLASSROOM RESEARCH (1990); ROBERT E. SLAVIN, RESEARCH METHODS IN EDUCATION: A PRACTICAL GUIDE (1984); MARY LEE SMITH & GENE V. GLASS, RESEARCH AND EVALUATION IN EDUCATION AND THE SOCIAL SCIENCES (1987); NORMAN E. WALLEN & JACK R. FRAENKEL, EDUCATIONAL RESEARCH: A GUIDE TO THE PROCESS (1991).

Discussions of research methodology related to more specific topics also exist. For discussions of research methodologies used in connection with the evaluation of thinking skills, see, e.g., JOAN BOYKOFF BARON & ROBERT J STERNBERG, TEACHING THINKING SKILLS: THEORY AND PRACTICE (1987); TOWARD THE THINKING CURRICULUM: CURRENT COGNITIVE RESEARCH: 1989 YEARBOOK OF THE ASSOCIATION FOR SUPERVISION AND CURRICULUM DEVELOPMENT (Lauren B. Resnick & Leopold E. Klopfer eds., 1989). For discussions of research methodologies dealing specifically with students' writing abilities, see HUNTER M. BRELAND ET AL., ASSESSING WRITING SKILLS (College Entrance Examination Board, Research Monograph No. 11, 1987); CHARLES W. BRIDGES, TRAINING THE NEW TEACHER OF COLLEGE COMPOSITION (1986); THE WRITING TEACHER AS RESEARCHER: ESSAYS IN THE THEORY AND PRACTICE OF CLASS-BASED RESEARCH (Donald A. Daiker & Max Morenberg eds., 1990); RESEARCH IN COMPOSITION AND RHETORIC: A BIBLIOGRAPHIC SOURCEBOOK (Michael G. Moran & Ronald F. Lunsford eds., 1984); JANICE M. LAUER & WILLIAM J. ASHER, COMPOSITION RESEARCH: EMPIRICAL DESIGNS (1988); EDWARD M. WHITE, DEVELOPING SUCCESSFUL COLLEGE WRITING PROGRAMS (1989). For discussions of methodological techniques used in connection with education research specifically addressing issues involving student attrition and academic failure, see PETER EWELL, CONDUCTING STUDENT RETENTION STUDIES (1984); INCREASING RETENTION: ACADEMIC AND STUDENT AFFAIRS ADMINISTRATORS IN PARTNERSHIP (Martha McGinty Stodt & William M. Klepper eds., 1987).

control group and single group designs. And Part IV concludes the Article.

I. ACTION RESEARCH IN EDUCATION

Walter Borg, a prominent authority on educational research, suggests that action research "provides the teacher or administrator in the field with objective, systematic techniques of problem solving that are far superior to an appeal to authority or reliance on personal experience . . . ."[8] Although action research is a valuable tool for individual teachers and administrators, it is *not* the same as regular educational research. Borg explains the difference:

> Although [action research] is similar in some respects to regular educational research, action research differs principally in the extent to which findings can be generalized beyond a local school situation. Educational research usually involves a large number of cases in order to reduce some of the random errors that occur in small samples. It involves establishing as much control as possible, consistent with the research goals, over such variables as teaching ability, pupil IQ, and socioeconomic status. Perhaps most important, regular educational research involves more precise sampling techniques than are found in action research. Many action research projects are carried out in a single classroom by a single teacher; others are carried out by all teachers in a school or even a school district. As action research projects become more extensive, they become more similar to other types of educational research. The emphasis in action research, however, is not on obtaining generalizable scientific knowledge about educational problems, but on obtaining knowledge concerning a specific local problem.[9]

An additional, and crucially important, point must be made now in connection with this discussion of action research. Many people who themselves have advanced training in education research methods insist that education research of any kind can be conducted only by people who understand how to use complex statistical tools. If this is true, then most teachers and administrators in the law schools simply will not be able to conduct edu-

8. WALTER R. BORG, APPLYING EDUCATIONAL RESEARCH: A PRACTICAL GUIDE FOR TEACHERS 14 (2d ed. 1987).

9. *Id.* at 13-14.

cation research of any kind because most law school teachers and administrators do not understand how to use complex statistical tools. Things, however, are not quite as simple as this. Historians and sociologists who study the development and nature of professions[10] and those who study the development and nature of substantive academic disciplines[11] independently paint a remarkably similar picture. These experts believe that people working in or studying similar fields gradually claim to possess the exclusive knowledge and skills necessary to that field.[12] These groups then begin to develop a specialized language or jargon that only they can understand.[13] Furthermore, these groups begin to insist that people wishing to work in a particular field must obtain specialized schooling that these groups control. Finally, and perhaps most significantly, professionals and academicians constantly attempt to prevent outsiders from working in their field.[14]

Given these facts, it is not surprising that educational researchers insist that only people trained in statistical analysis are equipped to conduct educational research. In effect, educational researchers can use statistical analysis as a weapon to defend their territory from outsiders.

Fortunately, not all education researchers use statistical analysis to repel outsiders. Borg, for example, flatly states that "most action research projects can be analyzed *using the simplest statistical procedures.* In fact, many action research projects

---

10. *See generally* ANDREW ABBOTT, THE SYSTEM OF PROFESSIONS: AN ESSAY ON THE DIVISION OF EXPERT LABOR (1988); MAGALI SARFATTI LARSON, THE RISE OF PROFESSIONALISM: A SOCIOLOGICAL ANALYSIS (1977).

11. *See generally* TONY BECHER, ACADEMIC TRIBES AND TERRITORIES: INTELLECTUAL ENQUIRY AND THE CULTURE OF DISCIPLINES (1989); BURTON R. CLARK, THE ACADEMIC LIFE: SMALL WORLDS, DIFFERENT WORLDS (1987); ROBERT C. WILSON & JERRY G. GAFF, COLLEGE PROFESSORS AND THEIR IMPACT ON STUDENTS (1975); Tony Becher, *The Disciplinary Shaping of the Profession*, *in* THE ACADEMIC PROFESSION: NATIONAL, DISCIPLINARY AND INSTITUTIONAL SETTINGS 271-303 (Burton R. Clark ed., 1987); Sheila Slaughter & Edward Silva, *Service and the Dynamics of Developing Fields*, 54 J. HIGHER EDUC. 481 (1983); Bruce Wilshire, *Professionalism as Purification Ritual: Alienation and Disintegration in the University*, 61 J. HIGHER EDUC. 280-93 (1990).

12. *See generally* NATIONAL ENDOWMENT FOR THE HUMANITIES, HUMANITIES IN AMERICA: A REPORT TO THE PRESIDENT, THE CONGRESS, AND THE AMERICAN PEOPLE (1988).

13. *See generally supra* note 12.

14. For a lengthy discussion of this notion of protecting jurisdiction, see generally ABBOTT, *supra* note 10.

need *no statistical analysis at all* to serve their purpose as an aid to decision making."[15] Furthermore, Borg actually provides step-by-step directions for conducting several statistical procedures, which he calls "easy analysis."[16] Borg's directions are as easy to use as pre-printed legal forms. Anyone who teaches in a law school can follow these step-by-step directions.[17]

Importantly, no one familiar with the practice of law would believe that a nonlawyer could read a short pamphlet or book and thus be prepared to try a case, write a will for Donald Trump, or draft the documents needed to transfer title to the Empire State Building. Only highly trained experts can do these things. However, most would admit that a layperson could read a pamphlet or small book and then make a credible appearance in small claims court or draft a simple will or an agreement for the sale or purchase of a house. The same is true of education research. A nonexpert in education research could not read a pamphlet or small book and then individually create valid mea-

---

15. BORG, *supra* note 8, at 7 (emphasis added).

16. *Id.* at 305.

17. Consider, for example, the statistic called a standard deviation. Standard deviations must be calculated if researchers want to know whether differences in test scores are significant or merely the product of chance. Borg provides an easy-to-follow example for calculating a standard deviation. The example calculates and standardizes 13 students' test scores for two separate groups:

> Step 1: Square each of the 13 Group A scores. [Group A was one of the two groups of students.] For example, 28 [squared] = 784. You can quickly compute the squares using a desk calculator, or you can look them up in a table of squares and square roots. . . .
>
> Step 2: Add up the squares for Group A. This gives you a total of 7662.
>
> Step 3: Divide 7662 by the number of pupils . . . in Group A, that is, 13. This gives us 589.38.
>
> Step 4: Now square the mean of 23.54 that we computed earlier for Group A. [An earlier recipe showed how this was done.] The square of 23.54 is 554.13. . . .
>
> Step 5: Subtract the result of 554.13 found in Step 4 from the result of 589.38 . . . found in Step 3. The result is 35.25.
>
> Step 6: Compute the square root of 35.25. Most hand calculators will give you the square root of 5.94. This square root is the standard deviation for Group A.

Id.

Admittedly, not all of Borg's formulas for statistical procedures are as short or simple as the standard deviation formula just described. Calculating "rank correlation," for example, has 11 steps. *Id.* at 307. Likewise, the "Mann-Whitney U-Test" takes 11 steps. *Id.* at 315. Furthermore, several of Borg's procedures require reference to statistics books. Nevertheless, Borg thinks that any administrator or teacher—from kindergarten teachers to law professors—can perform the analysis he describes.

suring instruments or do factorial research[18] or Aptitude-Treatment Interaction research.[19] Only highly trained experts can do these things. However, no education research practitioner would deny that a nonexpert could read a pamphlet or small book and then conduct simple correlation studies, simple experiments, or the like.

Apparently the only obstacle preventing nonexperts from conducting certain types of educational research and legal work is the parochialism of professionals.

II. BACKWARD-LOOKING RESEARCH TECHNIQUES: CORRELATIONAL, CAUSAL-COMPARATIVE, AND OBSERVATIONAL

Educational research falls into two main categories. Backward-looking research techniques attempt to explain things that already have happened. Three common types of backward-looking education research are correlational research, causal-comparative research, and observational research. Forward-looking research techniques, which involve experiments or quasi-experiments, attempt to predict future events in controlled settings.

Before describing these different types of action research, two commonalities should be noted. First, all education research must begin with a comprehensive review of the pertinent literature. Literature reviews often reveal that questions or issues that seem completely new to individual teachers or administrators have been grappled with by others for years. Second, action research, typically involving formal evaluations of persons or programs, utilizes complex program and personnel evaluation standards promulgated by evaluation professionals.[20] Such standards are designed to assure fairness and accuracy in the evaluation process.

A. *Correlational Research*

Correlational research involves using either simple or complex statistical tools to identify relationships between different

18. BORG & GALL, *supra* note 7, at 693.
19. *Id.* at 700.
20. *Id.* at 755.

variables.[21] For example, a law school teacher or administrator might wish to gather information to explain why some law students dramatically outperformed their LSAT/UGPA index scores. Such information could be useful in future admissions decisions or in creating academic programs. A simple action research project might easily accomplish this task.

A researcher might identify a group of students who had dramatically outperformed their index scores. The researcher then would administer a battery of tests and questionnaires. "Easy analysis" statistical tools might reveal common traits in students who outperformed their index scores. The researcher then could conclude that these particular characteristics are correlated with better-than-predicted academic performance.

Regrettably, correlational studies like this one cannot prove the existence of causal links between two or more variables because other "uncontrolled" variables may have intervened. For example, a correlational study of a rooster and a sunrise probably would show an extremely high correlation between the rising of the sun and the crowing of the rooster. Nevertheless, such a study could not be used to prove either that the rising caused the crowing or that the crowing caused the rising. Other things such as the movement of the planets, bad food, noise, or an attractive hen also may have caused these events.

Despite the serious limitation just noted, correlational studies can be used, to some extent, to predict the future. This is true because they at least *suggest* the existence of causal links. Thus, if a correlational study indicated a high correlation between good introductions on law school exams and high grades, then the researchers certainly would be justified in concluding that good introductions *might* be related causally to high grades.[22]

21. BORG & GALL, *supra* note 7. *See also* BORG, *supra* note 8, at 12.

22. Interesting data on countless other issues in the law school setting also could be collected using the correlational method of research. A teacher might wonder, for example, whether the students' conscious use of specific issue-spotting techniques, such as the "checklist" approach or the "rule of twos" approach, correlates with good grades. Both of these issue-spotting techniques are described in Paul T. Wangerin, *Learning Strategies for Law Students*, 52 ALB. L. REV. 471 (1988). Or a teacher might wonder whether the display in law school answers of what Kissam calls "short paragraph writing" skills correlates with better grades. Philip C. Kissam, *Law School Examinations*, 42 VAND. L. REV. 433 (1989). A teacher also might wonder what degree of correlation exists between the students' grades on the essay and objective portions of final exams. Correlational studies could produce data on all of these issues.

B. *Causal-Comparative Research*

The "causal-comparative" method of education research, like the correlational method, is a backward-looking method.[23] However, a critical difference separating the two makes the causal-comparative method a considerably more powerful research tool. Unlike the correlational method, which analyzes performance characteristics of a single group of people, causal-comparative method of research analyzes characteristics of *two* or more groups. Consider again a project attempting to discover why some students dramatically outperform their index scores. Recall that the earlier study involved the analysis of only one group, the outperforming group. A related causal-comparative study would examine two groups. One of those two groups, of course, would be the outperforming group, while the other group might consist of students who significantly *underperformed* their index scores. Analysis of the outperforming group may reveal that they had strong study skills or exceptional emotional support at home. On the other hand, analysis of the second group might reveal that its students had poor study skills or relatively little emotional support at home.

As noted earlier, correlational research cannot be used to *prove* causal relationships because variables outside the control of the researcher may have caused the measured effects.[24] The same is true of causal-comparative research. Nevertheless, causal-comparative research is a much more powerful tool because correlational research establishes correlations in only one set of circumstances, whereas causal-comparative research establishes comparable correlations in two related but different sets of circumstances. Although causal-comparative research cannot *prove* causal links, it provides a stronger suggestion than correlational research.

C. *Observational Research*

Observational research is perhaps the simplest kind of research for a person trained in the law.[25] Observational research in the field of education involves gathering observational data about teachers, students, and the educational environment.

23. BORG & GALL, *supra* note 7, at 535-71.
24. BORG, *supra* note 8, at 10.
25. BORG & GALL, *supra* note 7. *See also* BORG, *supra* note 8, at 10.

Consider, for example, an observational research project related to a statement many law school teachers made repeatedly: "I do not teach or test on 'rules' and 'information' in my classes; I teach and test skills of analysis, synthesis, and the like." A law school action researcher, for example, might send a trained observer to her own class and to gather data on what she actually does during the class. This researcher also might ask her own students for data on certain points. Halfway through the course, for example, she might ask the students to complete a short questionnaire: "In your opinion, what percentage of the class time do I devote to the teaching of skills and what percentage to the teaching of information or rules?"

An additional point concerning observational research should be addressed. This type of research frequently leads to other kinds of educational research. The observational data just described regarding the teaching of skills or information, for example, easily could lead to a research project involving a correlational or causal-comparative study. Alternatively, the same data easily could lead to experimental research.

III. Experiments in Education Research: The Quest for Experimental Validity

As noted earlier, the principal problem with correlational and causal-comparative research is that such research cannot be used to prove causal links. Experimental research is different from correlational and causal-comparative research because it *can* be used to prove causal links. To do so, researchers first must rule out, as possible causes of a measured effect, *anything other than an experimental intervention itself.* When researchers do this, their experiments are said to be "internally valid."[26] Next, researchers must rule out, as possible causes, all facets of an experimental intervention *other than the particular facet being examined.* When researchers do this, their experiments are said to be "externally valid."[27] Internal validity issues involve possible causes of effects that are *external* to the experiments. Conversely, external validity issues involve possible causes of effects that are *internal* to the experiments. The terminology used in

26. Borg & Gall, *supra* note 7, at 642.
27. *Id.* at 649.

this context, is confusing, if not downright bizarre. Nevertheless, this terminology is well-established.

A. *Internal Validity*

There are twelve different possible outside causes for measured effects in experiments.[28] Those twelve threats to internal validity can be classified readily into four general categories.

The first category consists of problems incidental to the use of measuring instruments. One of these problems is called "test sensitization."[29] The mere fact that subjects take a test at the beginning of a study (a pretest) can influence their scores on a test they take at the end of the study (a posttest). The pretest itself might have taught subjects things later displayed on the posttest, an effect called instrumentation. If these changes occurred, then differences between the subjects' performance on pretests and posttests may not have been caused by the experimental intervention. Also, statistical regression (regression toward the mean) can occur. People who score very high or very low on tests tend to move toward the middle when they take the same tests a second time. This can occur even if nothing else has changed. If this occurs, however, then effects seen at the end of an experiment might have been caused not by the intervention itself, but simply by this statistical phenomenon.[30]

Fortunately, it is relatively simple for action researchers to deal with these threats to internal validity. These problems occur only when performance is measured more than once. Thus, researchers who measure performance only at the conclusion of experiments need not worry about these threats. Unfortunately, researchers who avoid this problem then run headlong into two other groups of internal validity problems.

The second threat to internal validity, generally called the "differential selection" problem, involves differences existing between the experimental and control groups *prior to* commencement of experiments. If such differences exist, then it is possible that the effects measured after an experiment might have been caused by the pre-existing differences and not by the experimental interventions themselves. This problem can be addressed in

28. *Id.* at 644-48.
29. *Id.* at 644.
30. *Id.* at 645.

part by measuring performance at both the beginning and the end of an experiment. However, if this is done, threats to internal validity involving measuring issues again may become problematic.

The third category of threats to internal validity involves things that occur *during but outside of* experiments. For example, a phenomenon called "mortality" occurs when one or more members of either the control or the experimental groups leaves the experiment prior to its conclusion. In this way, mortality, and not the experimental intervention, might affect the conclusion. The term "history," on the other hand, refers to complications arising from events that occur during but outside of the experiments that might cause the effects measured. Finally, a problem known as "maturation" arises when the subjects of experiments change during the course of experiments due to aging.

Researchers may address these threats by measuring performance at the beginning as well as the end of experiments, using both a pretest and a posttest. When this is done, the researchers can tell if people who left experiments early are different from people who remained in the experiment until the end. If the people who left are not different than the ones who remained, then the representativeness of the group as a whole is the same despite the departure of those people. Likewise, pretesting allows the researcher to compare the maturation of members of the different groups. If few differences exist, this problem is solved. But again, pretesting and posttesting may give rise to problems of test sensitivity and regression.

The last category of internal validity problems is referred to throughout this Article as the "gang of four."[31] This category includes four common problems caused by using control groups in experiments. The first problem, "experimental treatment diffusion," occurs when members of a control group seek out and obtain, on their own, interventions that were provided only to members of an experimental group.[32] Second, "compensatory equalization of treatments" occurs when researchers capitulate to pressure from members of a control group and promise them

31. The author coined the term "gang of four" as it is used in educational research.

32. BORG & GALL, *supra* note 7, at 647-48.

compensatory treatment at some future time.[33]

The third problem, "compensatory rivalry," occurs when members of a control group perform better than average because they feel disadvantaged vis-a-vis the experimental group.[34] Conversely, the fourth problem, "resentful demoralization," occurs when members of a control group perform less than average work because they feel disadvantaged by their failure to receive benefits provided to an experimental group.[35]

All of the "gang of four" problems produce the same threat to the internal validity. Their occurrence might affect the control group's performance. In turn, effects caused by the experimental intervention may be either concealed or exaggerated, resulting in a lack of internal validity.

B. External Validity

As suggested earlier, a research study is externally valid if the results are likely generalizable to people and situations beyond those involved in the study.[36] Conversely, a study is externally invalid if its results are not generalizable to different situations and different people.[37]

Education researchers generally discuss two categories of external validity problems. The first involves population validity. Population validity issues involve differences between the people included in an experiment (the experimental population) and larger groups to whom the researcher hopes to generalize the results (the target population).[38] If the experimental population in an educational research study is different in any significant way from the target population to which the results of the study are supposed to generalize, then the results of that study might not be duplicated if the experiment is conducted with the target population or with a different sub-group of that larger group.[39]

---

33. *Id.*

34. *Id.* (This phenomenon, incidentally, is sometimes also called the "John Henry" effect.)

35. *Id.*

36. *Id.* at 649.

37. *Id.*

38. *See generally id.* at 649-50.

39. Consider, for example, the likelihood of generalizing of the results produced by an experiment involving an academic support program for specially admitted minority students at a law school that has extraordinarily high student admissions standards.

Unlike population validity problems, which deal with the people who are the subjects of a study, ecological external validity issues generally involve the circumstances under which studies are conducted.[40] If the circumstances of an educational study differ significantly from the circumstances to which they are intended to be generalized, then the results might not be duplicated under any but the original circumstances.

Ten factors have been identified as potential creators of ecological invalidity,[41] but this Article discusses only four.[42] First, in many experiments, the written description of an experimental intervention is either vague or general. For example, many reports state something no more specific than "study skills training was provided" or "students received instruction in exam writing techniques." If reports employ such vague descriptions, other people cannot duplicate the experimental interventions precisely.[43] Second, the performance of experimental participants sometimes is affected positively by the fact that they realize they are participating in a research study.[44] Thus, an experimental intervention producing certain results under experimental conditions might

---

See Carolyn J. Inouye, Understanding the Law School Experience: Enhancing Student Achievement and Satisfaction (1989) (unpublished Ph.D. thesis, University of California (Los Angeles)). Even if the experiment was internally valid, serious external validity problems almost certainly would exist because the "population" of elite law schools is vastly different from that of middle-tier or regional law schools. Although the results might be generalizable to other elite law schools, those results almost certainly are not generalizable to middle-tier or regional law schools.

40. *Id.*

41. *Id.*

42. The other six are: multiple-treatment interference; novelty and disruption effects; posttest sensitization; interaction of history and treatment; measurement of the dependent variable; and interaction of time of measurement and treatment effects. BORG & GALL, *supra* note 7, at 651.

43. For example, a report on the hypothetical academic support experiment just described might note generally that "study skills" were taught in the program using a "discussion method." These brief descriptions, however, do not allow other researchers or practitioners to know which particular study skills were taught in this program or how to duplicate the particular discussion method used.

44. BORG & GALL, *supra* note 7, at 651. This phenomenon, known as the Hawthorne Effect, is named after a series of experiments conducted regarding different levels of lighting on workplace performance. *Id.* One group of workers was told the experiment was being conducted to see what effect *higher* levels of light in the workplace had on performance. *Id.* Thereafter, when light levels were raised during the experiment, the workers' performance improved significantly. *Id.* Another group of workers was told the experiment was being conducted to see what effect *lower* light levels had on performance. *Id.* Thereafter, when light levels were lowered during the experiment, these workers' performance also improved significantly. *Id.*

not produce these same results again in a nonexperimental setting.[45]

The third threat to ecological validity involves a subject briefly mentioned earlier. Recall that researchers sometimes use the posttest-only control-group design when they fear pretest sensitization.[46] Pretest sensitization occurs when a pretest affects the performance of the people taking the test.[47] Pretest sensitization also can cause external validity problems. If a pretest has affected an experimental group's performance, a subsequent group that is provided the experimental intervention *but not the pretest* might not show the same gain as the experimental group.[48]

The fourth threat to ecological external validity involves two related but subtly different concepts: experimenter bias and experimenter effect.[49] Experimenter bias occurs when a researcher reports an experimental intervention to be significant simply because that researcher was predisposed to believe such

---

45. Hawthorne Effect issues seem to be related closely to "gang of four" issues, particularly compensatory rivalry. *Id.* at 647. However, a significant difference exists. "Gang of four" problems arise because members of control groups might react to their status as members of *control groups*. *Id.* Conversely, Hawthorne Effect problems arise because all subjects in a study might react to their status as *subjects of a study*. *Id.* Because the Hawthorne Effect potentially influences all subjects of a study, it does not cause internal validity problems because its impact is equal on members of both experimental and control groups. *Id.* Thus, its effects cancel themselves out. *Id.*

46. *See supra* notes 37-40 and accompanying text.

47. BORG & GALL, *supra* note 7, at 652.

48. Pretest sensitization issues also can cause internal validity problems. Consider, for example, how pretest sensitization might affect an experiment dealing with the effectiveness of a law school study skills course. Assume that following the administration of a study skills pretest, several students in the control group were intrigued by the ideas contained in the pretest and obtained study skills books or materials on their own. When such students take the posttest, their performance will be significantly different from their performance on the pretest. Those differences make the performance of the control group unrepresentative of the performance of nonparticipants in the experiment generally. Furthermore, those differences may conceal real gains accomplished by members of the experimental group.

Interestingly, a set of complicated designs, called Solomon group designs, addresses this problem. A Solomon four group design consists of two experimental and two control groups. One control group and one experimental group take the pretest and the other groups do not. If differences exist between the performances of the groups who took the pretest and those who did not, pretest sensitization occurred. Complex statistical analysis then can be used to sort out the effect of the pretest. BORG & GALL, *supra* note 7, at 705-08.

49. *Id.* at 656-58, 705-08. Experimenter bias also can wreak havoc with internal validity in an experiment. *Id.* at 657.

significance might occur.[50] In these situations, the researcher may be either lying or simply misrepresenting the data. In either event, the results of the experiment are unlikely to be replicated if different researchers conduct them. Experimenter effect differs subtly. Sometimes deviation in an experimental group following an intervention occurs because of the personality or skill of the intervention administrator rather than because of the intervention itself. If this occurs, the results of the experiment are unlikely to be replicated when another individual administers the intervention.[51]

A subtle but important relationship exists between external validity issues and at least one category of internal validity problems. As noted earlier, "gang of four" problems involve uncontrolled variables that solely affect the performance of *control groups*. When this occurs, an experiment is internally invalid. Similarly, some uncontrolled variables solely affect the performance of *experimental groups*, producing external invalidity.

Borg has noted that people who conduct action research need not be concerned about external validity issues.[52] Typically, action researchers are not interested in publishing their findings in scholarly journals, and are not concerned with others imitating their actions. Action research, after all, is primarily a technique for providing individual teachers and administrators with objective, systematic, problem-solving techniques.[53] Despite these comments, action researchers should not be wholly oblivious to external validity.

First, after conducting their research, action researchers necessarily assume that the results are generalizable at least to

50. In a famous experiment, students were told that many generations of breeding had made some rats maze-bright and some rats maze-dull. Students then were assigned to train the rats described as either maze-bright or maze-dull. In fact, all of the rats were essentially identical. When the experiment was over, the researchers discovered that the rats thought to be maze-bright earned significantly higher scores than those thought to be maze-dull. This experiment is described briefly in BORG & GALL, *supra* note 7, at 656-57. For a full report, see Robert Rosenthal & Kermit L. Fode, *The Effect of Experimenter Bias on the Performance of the Albino Rat*, 8 BEHAV. SCI. 183 (1963).

51. A particularly pernicious example of experimenter effect occurs when a single researcher interacts with both a control group and an experimental group. The researcher may interact with the experimental group in a way that produces a positive change and may interact with the control group in a way that does not produce such a change.

52. BORG, *supra* note 8, at 14.

53. *Id.*

future situations involving different students and teachers *at the same institution*. If not, the results of the research could not justify future action.

Second, action researchers cannot ignore external validity because of cost effectiveness. Regular educational researchers often conduct one-shot studies with financial assistance from sources outside the institution. However, action researchers generally attempt to create educational programs that operate on a continuing basis and are funded by limited internal resources. Experiments lacking external validity easily can lead to the creation of ongoing and costly educational programs. If an action researcher cannot tell which variable of the experiment lead to the measured changes, then he or she cannot pinpoint the specific ways for expending future limited resources.

IV. EXPERIMENTAL DESIGNS

Over the years, researchers have created numerous designs in an attempt to solve the internal and external validity problem. The most common designs are called the control group or multiple group designs.[54] Single group or single subject designs are less known but more useful for action researchers.[55]

A. *Control Group Designs*

When education researchers design control group experiments, they generally do three things. First, they divide all of the subjects of the study into either a control group or an experimental group. Second, researchers provide the experimental group with an experimental intervention that is withheld from the control group. Third, at the conclusion of the experiments, researchers measure the performance of all subjects.[56]

One important aspect of the control group design is the researchers' attempt to divide the subjects into subgroups by random assignment, which creates essentially identical control and

54. *See generally* BORG & GALL, *supra* note 7, at 662.

55. *Id.* at 673-74.

56. Researchers typically use standardized tests to measure performance. *See generally* DANIEL J. KEYSER & RICHARD C. SWEETLAND, TESTS: A COMPREHENSIVE REFERENCE FOR ASSESSMENTS IN PSYCHOLOGY, EDUCATION AND BUSINESS (3d ed. 1991); JAMES V. MITCHELL, JR., THE NINTH MENTAL MEASUREMENTS YEARBOOK (1985). Researchers sometimes measure performance simply by looking at the grades students receive in one or more classes.

experimental groups. This is critically important for two reasons. First, if the groups are identical at the beginning, the problem of differential selection is eliminated. Creating identical groups at the beginning means that any pre-existing conditions that might influence the postintervention performance of the experimental group also will influence that of the control group. Second, if the two groups are identical at the beginning, the problems of history and maturation are eliminated. Furthermore, the postintervention performance of both the control and the experimental groups is equally susceptible to pre-existing external conditions.

There are two types of control group design experiments: the "posttest-only control group design" and the "pretest-posttest control group design." The posttest-only design measures performance of the control and the experimental groups only at the conclusion of experiments. This design is used to avoid the serious problems that arise whenever researchers measure performance more than once, such as test sensitization, instrumentation, and statistical regression. Such problems are particularly difficult for action researchers because addressing them often requires the use of complex statistical tools. The posttest-only design provides a simple method of eliminating such problems.

Education researchers often use a simple diagram to illustrate the posttest-only control group design.

$$\begin{array}{cc} X & O \\ \hline & O \end{array}$$

In this diagram, the experimental intervention is referred to with the letter "X." The observation and measurement of performance is referred to with the letter "O." The experimental group is above the line, the control group below. The solid line in the diagram indicates that the two groups were selected by random assignment.[57]

Regrettably, the posttest-only design is susceptible to mortality and maturation. An alternative that addresses that weakness is the pretest-posttest control group design. This design is diagrammed as follows:

57. Random assignment sometimes is noted by an "R" at the left side of each line of the diagram. *See, e.g.*, BORG & GALL, *supra* note 7, at 663.

$$\begin{array}{ccc} O & X & O \\ \hline O & & O \end{array}$$

Again, the experimental group (above the line) receives an experimental intervention (X), but the control group (below the line) does not. Performance is observed and measured twice, before and after the intervention.

Quasi-experiments can be used to study situations arising in school settings where random assignment to control and experimental groups is not possible because of practical, political, or even ethical problems.[58] Quasi-experiments (studies not employing random assignment) work best when they use the "Nonequivalent Control Group Design." This design is very similar to the pretest-posttest control group design.[59] Recall the diagram for the pretest-posttest design:

$$\begin{array}{ccc} O & X & O \\ \hline O & & O \end{array}$$

The diagram for the Nonequivalent Control Group Design is:

$$\begin{array}{ccc} O & X & O \\ \hdashline O & & O \end{array}$$

As these two diagrams reveal, both designs use an experimental group (above the line) and a control group (below the line), and both provide an experimental intervention (X) only to the experimental group. Further, each uses pretests and posttests (Os). The only difference in the diagrams involves the dividing line. The solid line indicates random assignment. The broken line indicates nonrandom assignment.

Because the nonequivalent control group design uses both a pretest and a posttest, it allows researchers to compensate for the lack of random assignment in two ways. Both ways require a comparison of the personal characteristics and the performance of the control and experimental groups. This is done in light of results produced on the pretest. Often the comparison will reveal that the two groups are remarkably similar. If this occurs, the researchers can proceed as though conducting a regular control group experiment and assume that pre-existing conditions will cancel themselves out. However, the pretest may reveal that

58. *Id.* at 669-70.
59. *Id.* at 690.

the two groups are quite dissimilar. If this occurs, the researchers can obtain help from statistical experts who may compensate for the pre-experiment differences with the use of complex statistical tools.[60]

While these designs can address internal validity problems, they cannot rectify "gang of four" problems. However, one technique has been developed to address "gang of four" problems. This technique, borrowed from the field of medical research, requires researchers to administer simultaneously an experimental intervention to the experimental group and a placebo intervention to the control group.[61] This technique requires researchers using a pretest-posttest design to administer *two* pretests and *two* posttests. Some of these pretests and posttests measure the performance targeted for change by the experimental intervention while others are placebo tests.

The various versions of placebo designs can be diagramed as follows. In these diagrams, the "Ps" represent placebo interventions, and the double "Os" represent the two pretests and the two posttests. The new designs are:

Posttest-Only Control Group Design with Placebo

X OO
P OO

Pretest-Posttest Control Group Design with Placebo

OO X OO
OO P OO

Nonequivalent Control Group Design with Placebo

OO X OO
OO P OO

Two important conditions must be satisfied before placebo designs can work. First, all of the subjects of the experiment must think the placebo intervention is as beneficial as the experi-

60. For example, the statistical tool, known as "analysis of covariance," can compensate for pre-experiment differences between members of control and experimental groups. *Id.* at 693.

61. This concept is described in BORG & GALL, *supra* note 7, at 196-99, 665-66. *See generally* John G. Adair et al., *The Placebo Control Group: An Analysis of Its Effectiveness in Education Research*, 59 J. EXPERIMENTAL EDUC. 67 (1990); K. Daniel O'Leary & Thomas D. Borkovec, *Conceptual, Methodological and Ethical Problems of Placebo Groups in Psychotherapy Research*, 33 AM. PSYCHOLOGIST 821 (1978).

mental intervention. In this situation, "gang of four" problems are not likely to arise. Second, the placebo must not affect the performance of the control group on the measured variable.

Refer back to the study involving the effects of study skills training on specially admitted minority students who ultimately displayed better-than-expected grades. Assume this teacher fears a pretest sensitization, instrumentation, or a statistical regression problem. To remedy these problems, she would use a basic posttest-only control group design, requiring random assignment of subjects into control and experimental groups. If this teacher also fears "gang of four" problems, she simply would add a placebo intervention for the control group, such as library skills training. This placebo likely would be considered as beneficial as the study skills training, and it would be unlikely to affect the grades the control groups received in their substantive classes. If the experimental group performed significantly better than the control group, the teacher could conclude that the study skills training program had caused the better-than-expected performance because all other possible causes had been eliminated.

Unfortunately, the study just described has one very serious flaw. Although the researcher could conclude that the experimental intervention caused the difference in overall academic performance, she could not conclude that the study skills themselves caused the difference. Some other aspect of the intervention might have caused the difference. It is likely, for example, that students in the experimental group got together informally after sessions of the study skills training and went out for beer and pizza. Perhaps these beer and pizza parties, and not the study skills training, caused the better-than-expected academic performance. Given the design of the experiment, the researcher cannot isolate the variable causing the better-than-expected performance. This experiment, in short, lacks external validity.

Several things can be done to increase the external validity. For example, the population and ecological validity can be bolstered by conducting the same study simultaneously in law schools. Unfortunately, this approach to external validity is not a particularly workable solution for educational and particularly action researchers. It is difficult for people to conduct empirical studies of teaching and program effectiveness at individual schools, much less at multiple schools. Thus, a solution is needed that individual teachers within their institutions can use. Fortunately, a partial solution exists.

Recall again that the design used in the study skills experiment was a posttest-only control group with placebo design. Such a design controls for virtually all internal validity problems.[62] If a researcher also wanted to increase the external validity, while conducting the study at a single school, the experiment would be rather complex. This experiment could employ a pretest-posttest control group with placebo design.[63]

$$\begin{array}{ccc} \mathrm{OO} & \mathrm{X} & \mathrm{OO} \\ \hline \mathrm{OO} & \mathrm{P} & \mathrm{OO} \end{array}$$

As the diagram indicates, this teacher would: (1) randomly assign the subject students to both control and experimental groups (the solid line); (2) administer two pretests to both groups, one measuring study skills and the other measuring placebo skills (the first sets of double OOs); (3) simultaneously provide an experimental and a placebo intervention to the appropriate groups (the P and the X); and (4) administer two posttests to each group, measuring study skills and placebo skills (the second sets of double OOs).

Additionally, the teacher would collect data both on the performance of the two groups *and on the performance of individuals in the two groups.* Initially, she would determine if the experimental group demonstrated greater improvement in study skills between the pretest and the posttest than the control group. If so, she could conclude that the study skills program had improved the study skills of the subject students. No other explanation for the difference in study skills performance is possible. She would look for relationships between the posttest abilities of individual students and general academic performance. If students *in both groups* who displayed good study skills on the posttest also recorded good grades generally, then the researcher could conclude that good study skills correlated with good grades.

62. Mortality is the only internal validity problem that this design does not address.

63. The design for the study now described, except for the placebo element, is drawn from Philip Beal & Ernest T. Pascarella, *Designing Retention Interventions and Verifying Their Effectiveness*, *in* STUDYING STUDENT ATTRITION: NEW DIRECTIONS FOR INSTITUTIONAL RESEARCH 73 (Ernest T. Pascarella ed., 1982).

Finally, this teacher would merge these conclusions. If the program improved the study skills of the participating students *and* if good study skills correlated positively with good grades for individuals of both groups, then the training aspect of the program caused the difference. Other possible internal explanations—e.g., beer and pizza parties—are eliminated as possible causes. Therefore, the results of this study are highly generalizable.

Finally, in many situations, educational action researchers will be unable to use control group designs because of ethical, political, or other problems. In addition, sufficiently appealing placebo interventions may be unavailable. Despite their ability to prove causal links, such designs are worthless if they cannot be used in real-life educational settings.

B. Single Group Designs

Fortunately, other designs are available for use in educational research. For instance, single subject or single group designs allow educational researchers to gather valuable data without using control groups.[64]

Several single group designs exist. Two such designs, the "one shot case study" design and the "single group pretest-posttest," fail to produce meaningful data.[65] Time-Series designs are

---

64. BORG & GALL, *supra* note 7, at 673-74.

65. In one-shot case studies, researchers provide one group of subjects with an experimental intervention and then administer a posttest. *See generally* BORG & GALL, *supra* note 7. The one-shot case study design barely qualifies as a research design. *Id.* Researchers cannot determine whether pre-existing conditions influenced performance. This internal validity problem is known as test sensitization. Furthermore, researchers cannot determine whether external conditions influenced poststudy performance. The lack of internal validity, in turn, makes this design externally invalid.

Ironically, many legal educators regularly use this design informally to evaluate their own personal teaching effectiveness. As for teaching reasoning skills in their classes, many law teachers offer the student exams as convincing evidence that at least some of their students possess the skills taught. Law students who display complex thinking skills at the end of a given course may have possessed those skills before they started the course. Conversely, law students who display such skills at the end of a course may have learned those skills during the course but completely outside of the course itself.

Another single group design, the "single group pretest-posttest," addresses the possible effect of prestudy influences but not during-but-outside-the-study influences. *Id.* at 670. This test establishes a baseline of performance, describing the subjects' performance before the study begins. Researchers using this design then provide an experimental intervention. Finally, they administer a posttest. If the posttest shows a change considerably better.[66]

When education researchers use Time-Series designs, they often conduct what are called "A-B" experiments. During the A phase of such experiments, they withhold an experimental intervention from a single subject or a single group of subjects. While doing this, they repeatedly measure the performance of the subjects on a specified variable. The performance of the single subject or single group during this phase creates a baseline representative of the target population. During the B phase, researchers provide the single subject or group with experimental intervention. After or during the intervention, they measure performance repeatedly. If a change in performance occurs between the A and B phases, the researchers may conclude tentatively that the intervention caused the change.

The A-B single group design is diagrammed as:

O O O O X O O O O

The Os in the diagram represent observation and measurement, and the X represents the experimental intervention. As the diagram illustrates, repeated observation during a nonintervention phase precedes a repeated observation following an intervention.

The A-B single group design does not produce compelling evidence. The possible impact of uncontrolled variables that exist during-but-outside-of studies is problematic. Nevertheless, the A-B design produces strongly suggestive, if not compelling, proof of causal links because of the repeated observation and

---

in performance from the pretest, the researchers can be relatively confident that something during the study caused the change. However, it is impossible to determine the variable responsible for the change.

For example, it is entirely possible that students who displayed positive change between the pretest and the posttest learned the target skills during the study from a teacher other than the one conducting the study, or that they learned the skills entirely on their own.

The foregoing reveals the critical difference between single group pretest-posttest designs and control group pretest-posttest designs. Control group designs control for events occurring during but outside of studies. They do because the control group cancels out such during-but-outside influences on the experimental group.

66. The following discussion relies heavily on DAVID H. BARLOW & MICHEL HERSON, SINGLE CASE EXPERIMENTAL DESIGN: STRATEGIES FOR STUDYING BEHAVIOR CHANGE (1976). *See generally* BORG, *supra* note 8. For a lengthy discussion of this topic, see also LOUIS COHEN & LAWRENCE MANION, RESEARCH METHODS IN EDUCATION 204 (1989). *See generally* BORG & GALL, *supra* note 7.

measurement. When researchers conduct A-B studies, they can identify precisely the time when the subjects' performance changes. If the time of the change corresponds with the experimental intervention, then strong circumstantial evidence indicates that the intervention caused the change.[67]

Interestingly, the A-B-A-B single group design provides stronger evidence of causation than A-B designs despite their similarity. With both designs, a single group serves as both the control group and the experimental group. Furthermore, both design researchers repeatedly measure performance. However, in A-B studies, an intervention is withheld and provided only once. Conversely, in A-B-A-B studies, an intervention is withheld and provided, then withheld and provided a second time.

The diagram for A-B-A-B single group designs is:

O O O O X O O O O E O O O O X O O O O

In this diagram, once again, the Os represent observation and measurement, the Xs represent experimental interventions, and the E represents "extinction."[68]

The power of the A-B-A-B design lies in the unusual pattern and timing of change that it can produce. In connection with studies using this design, performance may be relatively low and stable when the researchers initially withhold an intervention. Then, when the researchers provide the intervention, performance may increase significantly and remain relatively high while the researchers continue to provide the intervention. Next, just when or shortly after the researchers again withhold

67. Consider an example involving a law school teacher who wants to increase the general level of classroom participation. In particular, this teacher wants to know whether an announcement that he plans to factor class participation into their final grades actually increases the level of participation in large classes. During the A phase of an A-B single group design, the researcher would take single baseline measurements of participation preceding any announcement. This would be the pre-intervention observation and measurement. Experimenter intervention would occur halfway through the semester, when the teacher announced to the class that he would be recording students' classroom participation and counting it toward the students' final grades. Thereafter, a researcher would measure classroom participation repeatedly. At the conclusion of this study, the teacher would graph the results. If such participation remained at a relatively low and constant rate prior to the intervention and then rose after the intervention, the teacher could conclude that the intervention caused the change.

68. BORG & GALL, *supra* note 7, at 713. The second withholding in this design is called "extinction." *Id.*

the intervention (extinction), performance may fall, although it may not necessarily fall as low as the initial baseline. Finally, when the researchers again provide the intervention, performance rises.

Clearly the A-B-A-B design is more suggestive of causal links between experimental interventions and measured performances than is the simpler A-B design. The down/up pattern and the timing of change in an A-B study could be either coincidental or the result of uncontrolled variables. However, mere coincidence or uncontrolled variables could not have caused the down/up/down/up pattern of change and the timing in an A-B-A-B study.

As noted earlier, correlation studies cannot be used to *prove* the existence of causal links because they cannot eliminate the possibility that extraneous variables caused the measured change. The same is true of A-B studies. Thus, these two studies merely *suggest* the existence of causal links. Just as causal-comparative studies produce much stronger suggestions of causal links than do correlation studies, so also do A-B-A-B studies. They both produce a repetition or pattern of change unlikely to result from external variables or mere coincidence.[69]

V. CONCLUSION

Action research is by no means an easy task. Its use in educational research is just as complicated as regular research because it employs the same designs. This is true whether they are single group or control group designs. Furthermore, the same validity problems plague both regular and action research. However, action researchers need not be as concerned with generalizability (external validity) as are regular researchers. Finally, action researchers can conduct observational, causal-comparative, and correlational research. Admittedly, law school teachers and administrators are not suited to do all types of educational

69. Another link between control and single group studies also can be seen in this context. Ethical or political restraints sometimes may prohibit education researchers from withholding experimental interventions from control groups. Human beings, after all, are not laboratory rats. Such restraints also may prohibit researchers from withholding interventions from single groups during the second A phase of A-B-A-B studies. However, the ethical problems are somewhat ameliorated in A-B-A-B studies because the intervention is provided again in the second B phase of the study. BORG & GALL, *supra* note 7, at 717.

research, just as nonlawyers are not able to do all types of legal work. But the inability to do everything in a field surely should not prohibit people from doing something in that field.

It should be clear now that this Article concerns something much larger than educational research methodology. It is about the underlying nature of professions and academic disciplines, and the inherent unwillingness to share jurisdiction. Professionals and academic insiders often conclude that outsiders cannot do any of the work that a profession or discipline claims for itself. As of now, relatively little sharing occurs between the profession of law and the discipline of education. Consequently, students suffer.