cious insights and suggestions and with steady and unfailing psychological support. I am very grateful to them for their work with me and for being conscientious, scrupulous, and creative editors.

Authors rarely recognize and acknowledge the work and contributions of project editors. Project editors—I prefer to say "super editors"—are those individuals who solve the many technical communication problems encountered in preparing book manuscripts for publication. I would like to mention the extraordinary competence, insight, and aesthetic and technical creativity of the individual who has prepared this book for the printer: Jeanette Ninas Johnson. I am very grateful and here acknowledge her valuable contribution.

It is doubtful that this book could have been written without the sabbatical leaves given me in 1961–1962 and 1970–1971 by New York University. I am grateful to the University for its generous sabbatical policy.

The price a family pays for an author's book is high. Its members put up with his obsession and his unpredictable writing ups and downs. I express my gratitude and indebtedness to my wife and sons by dedicating the book to them. I must say more than this, however. My wife has had to cope with two overseas moves and one transcontinental move, two retirements, and innumerable logistical and temperamental problems. To express thanks and gratitude in the face of this extraordinary example of coping seems pale and inadequate. Nevertheless, I here express both.

Fred N. Kerlinger

Eugene, Oregon
June 1985

Contents

Preface VII

PART ONE
The Language and Approach of Science

Chapter 1 Science and the Scientific Approach 3
Science and common sense. Four methods of knowing. The aims of science, scientific explanation, and theory. Scientific research—A definition. The scientific approach.

Chapter 2 Problems and Hypotheses 15

Chapter 3 Constructs, Variables, and Definitions 26
PART TWO
Sets, Relations, and Variance

Chapter 4 Sets 45
Subsets. Set operations. The universal and empty sets; set negation. Set diagrams. Set operations with more than two sets. Partitions and cross partitions. Levels of discourse. Study suggestions.

Chapter 5 Relations 55
Relations as sets of ordered pairs. Determining relations in research. Rules of correspondence and mapping. Some ways to study relations. Multivariate relations and regressions. Study suggestions.

Chapter 6 Variance and Covariance 69

PART THREE
Probability, Randomness, and Sampling

Chapter 7 Probability 89

Chapter 8 Sampling and Randomness 109

PART FOUR
Analysis, Interpretation, Statistics, and Inference

Chapter 9 Principles of Analysis and Interpretation 125

Chapter 10 The Analysis of Frequencies 147

Chapter 11 Purpose, Approach, Method 174
The basic approach. Definition and purpose of statistics. Binomial statistics. The variance. The law of large numbers. The normal probability curve and the standard deviation. Interpretation of data using the normal probability curve-frequency data. Interpretation of data using the normal probability curve-continuous data.

Chapter 12 Testing Hypotheses and the Standard Error 186

PART FIVE
Analysis of Variance

Chapter 13 Analysis of Variance: Foundations 203

Chapter 14 Factorial Analysis of Variance 225

Chapter 15 Analysis of Variance: Correlated Groups 251

Chapter 16 Nonparametric Analysis of Variance and Related Statistics 265

PART SIX
Designs of Research

Chapter 17 Research Design: Purpose and Principles 279
Chapter 18 Inadequate Designs and Design Criteria 292

Chapter 19 General Designs of Research 302

Chapter 20 Research Design Applications: Randomized Groups 319
Simple randomized subjects design. Factorial designs. Evaluation of randomized subjects designs. Study suggestions.

Chapter 21 Research Design Applications: Correlated Groups 329
The general paradigm. Research examples of correlated-groups design. Multigroup correlated-groups designs. Factorial correlated groups designs. Analysis of covariance. Research design and analysis: Concluding remarks. Study suggestions.

PART SEVEN
Types of Research

Chapter 22 Nonexperimental Research 347

Chapter 23 Laboratory Experiments, Field Experiments, and Field Studies 364

Chapter 24 Survey Research 377
Types of surveys. The methodology of survey research. Two studies. Applications of survey research to education. Advantages and disadvantages of survey research. Study suggestions.

PART EIGHT
Measurement

Chapter 25 Foundations of Measurement 391

Chapter 26 Reliability 404

Chapter 27 Validity 416

PART NINE
Methods of Observation and Data Collection

Chapter 28 Interviews and Interview Schedules 438
Interviews and schedules as tools of science. The interview schedule. The value of interviews and interview schedules. Addendum. Study suggestions.

Chapter 29 Objective Tests and Scales 449

Chapter 30 Available Materials, Projective Methods, and Content Analysis 468

Chapter 31 Observations of Behavior and Sociometry 486

Chapter 32 Q Methodology 507

PART TEN
Multivariate Approaches and Analysis

Chapter 33 Multiple Regression Analysis: Foundations 526
Two research examples. Simple regression analysis. Multiple linear regression. The multiple correlation coefficient. Tests of statistical
significance. Interpretation of multiple regression statistics. Other analytic and interpretative problems. Research examples. Multiple regression analysis and scientific research.

Chapter 34 Multiple Regression, Analysis of Variance, and Other Multivariate Methods 551

Chapter 35 Factor Analysis 569

Chapter 36 Analysis of Covariance Structures 596

Appendices
Appendix A: Historical and Methodological Research 620
Appendix B: The Computer and Behavioral Research 625
Appendix C: Random Numbers and Statistics 637
Appendix D: The Research Report 645
Name Index 653
Subject Index 659

Foundations of Behavioral Research
PART SEVEN

TYPES OF RESEARCH
Chapter 22

Nonexperimental Research

Among prevalent fallacies, one of the most dangerous to science is that known as post hoc, ergo propter hoc: after this, therefore caused by this. We may joke, with a tinge of seriousness, "If I take an umbrella, it won't rain." We may even seriously say that delinquents are delinquent because of a lack of discipline in the schools or that religious education makes children more virtuous. It is easy to assume that one thing causes another simply because it occurs before the other, and because one has such a wide choice of possible "causes." Then, too, many explanations often seem plausible. It is easy to believe, for instance, that the learning of children improves because we institute a new educational practice or teach in a certain way. We assume that the improvement in their learning was due to the new spelling method, to the institution of group processes into the classroom situation, to stern discipline and more homework (or little discipline and less homework). We rarely realize that children will usually learn something if they are given the opportunity to learn.

The social scientist and the educational scientist constantly face the problem of the post hoc fallacy. The sociologist who seeks the causes of delinquency knows that extreme care must be used in studying the problem. Slum conditions, broken homes, lack of love—each or all of these conditions are possible causes of delinquency. The psychologist seeking the roots of adult personality faces an even subtler problem: hereditary traits, child-rearing practices, educational influences, parental personality, and environmental circumstances are all plausible explanations. The educational scientist, with the goal of understanding the basis of successful school achievement, also faces a large number of reasonable possibilities: intelligence, aptitude, motivation, home environment, teacher personality, pupil personality, and teaching methods.
The danger of the post hoc assumption is that it can, and often does, lead to erroneous and misleading interpretations of research data, the effect being particularly serious when scientists have little or no control over time and independent variables. When they seek to explain a phenomenon that has already occurred, they are confronted with the unpleasant fact that they do not have real control of the possible causes. Hence they must pursue a course of research action different in execution and interpretation from that of scientists who experiment.

Definition

Nonexperimental research is systematic empirical inquiry in which the scientist does not have direct control of independent variables because their manifestations have already occurred or because they are inherently not manipulable. Inferences about relations among variables are made, without direct intervention, from concomitant variation of independent and dependent variables.

Assume that an investigator is interested in the relation between sex and creativity in children. He measures the creativity of a sample of boys and girls and tests the significance of the difference between the means of the two sexes. The mean of boys is significantly higher than the mean of girls. He concludes that boys are more creative than girls. This may or may not be a valid conclusion. The relation exists, true. With only this evidence, however, the conclusion is doubtful. The question is: Is the demonstrated relation really between sex and creativity? Since other variables are correlated with sex, it might have been one or more of these variables that produced the difference between the creativity scores of the two sexes.

**BASE DIFFERENCE BETWEEN EXPERIMENTAL AND NONEXPERIMENTAL RESEARCH**

The basis of the structure in which experimental science operates is simple. One hypothesizes: If x, then y; if frustration, then aggression. Depending on circumstances and personal predilection in research design, one uses some method to manipulate or measure x. One then observes y to see if concomitant variation, the variation expected or predicted from the variation in x, occurs. If it does, this is evidence for the validity of the proposition, \( x \to y \). If not, then the null hypothesis cannot be rejected. To help us achieve control, we can use the principle of randomization and active manipulation of x and can assume, other things equal, that y is varying as a result of the manipulation of x.

In nonexperimental research, on the other hand, y is observed, and x, or several x's, are observed, either before, after, or concomitant to the observation of y. There is no difference in the basic logic: it can be shown that the argument structure and its logical validity are the same in experimental and nonexperimental research.1 And the basic purpose of both is also the same: to establish the empirical validity of so-called conditional statements of the form \( p \to q \). The essential difference is direct control of p, the independent variable. In experimental research, p can be manipulated, which is rather direct "control." When Clark and Walberg had teachers give one group of subjects massive reinforcement and other teachers give another group moderate reinforcement, they were directly manipulating or controlling the variable reinforcement. Similarly, when Bandura and Menlove showed one group a movie with a single model, another group a movie with multiple models, and a third group a "neutral" movie, they were directly manipulating the variable modeling. In addition, subjects can be assigned at random to the experimental groups.

In nonexperimental research, direct control is not possible: neither experimental manipulation nor random assignment can be used. These are two essential differences between experimental and nonexperimental approaches. Owing to lack of relative control of x and other possible x's, the "truth" of the hypothesized relation between x and y cannot be asserted with the confidence of the experimental situation. Basically, nonexperimental research has, so to speak, an inherent weakness: lack of control of independent variables.

The most important difference between experimental research and nonexperimental research, then, is control. In experiments, investigators have direct manipulative control: they have at least one active variable. If an experiment is a "true" experiment, they can also exercise control by randomization. They can assign subjects to groups at random, or can assign treatments to groups at random. In the nonexperimental research situation, this kind of control of the independent variables is not possible. Investigators must take things as they are and try to disentangle them.

Take a well-known case. When we paint the skins of rats with carcinogenic substances (x), adequately control other variables, and the rats ultimately develop carcinoma (y), the argument is compelling because x (and other possible x's, theoretically) is controlled and y is predicted. But when we find cases of lung cancer (y) and then go back among the possible multiplicity of causes (x1, x2, . . . , xN) and pick cigarette-smoking (say x3) as the culprit, we are in a more difficult and ambiguous situation. Neither situation is sure, of course; both are probabilistic. But in the experimental case we can be more sure—considerably more if we have adequately manipulated "other things equal"—that the statement if \( x \to y \) is truly valid. In the nonexperimental case, however, we are always on shakier ground because we cannot say, with nearly as much assurance, "other things equal." We cannot control the independent variables by manipulation or by randomization. In short, the probability that \( x \to y \) is "really" related to y is greater in the experimental situation than it is in the nonexperimental situation, because the control of x is greater.

**SELF-SELECTION AND NONEXPERIMENTAL RESEARCH**

In an ideal behavioral research world, the drawing of random samples of subjects, and the random assignment of subjects to groups and treatments to groups, would always be possible. In the real world, however, one, two, or even all three of these possibilities do not exist. It is possible to draw subjects at random in both experimental and nonexperimental research. But it is not possible, in nonexperimental research, to assign subjects to groups at random or to assign treatments to groups at random. Thus subjects can "assign themselves" to groups, can "select themselves" into the groups on the basis of characteristics other than those in which the investigator may be interested. The subjects and the treatments come, as it were, already assigned to the groups.

Self-selection occurs when the members of the groups being studied are in the groups, in part, because they differentially possess traits or characteristics extraneous to the research problem, characteristics that possibly influence or are otherwise related to the variables of the research problem. Examples of self-selection may aid understanding.

In the well-known research on cigarette-smoking and cancer, the smoking habits of a large number of people were studied. This large group was divided into those who had

---

LARGE-SCALE NONEXPERIMENTAL RESEARCH

Research examples will, as usual, help us to understand the nature of nonexperimental research. Instead of summarizing only individual studies, as we have to now, we describe both individual studies and sets of studies centered around some phenomenon or variable of interest. Nonexperimental behavioral research often focuses on large problems of social and human importance: social class, political processes, segregation and desegregation, public attitudes, school achievement, for example. The importance or "relevance" is the fashionable word—or of the subject of these studies should not obscure our understanding of their nonexperimental character. Because nonexperimental research has inherent weaknesses, however, does not mean that experimental research is more important. As said earlier, the experiment is one of the great inventions of all time, an ideal of control toward which we aspire. This does not mean that experiments are necessarily "better" than nonexperimental studies. On the other hand, nonexperimental research is not necessarily "better" than experimental research because its content and variables seem to be socially important. This would be like saying that psychological research is "better" than sociological research because psychologists more often use an experimental approach and sociologists a nonexperimental approach!

Authoritarianism and Ethnocentrism

One of the most important and influential studies of the century was the set of investigations into ethnocentrism and authoritarianism reported in the book The Authoritarian Personality. The general hypothesis of the study was that political, economic, and social beliefs are related to deep-seated personality characteristics. Another hypothesis was that adult personality is derived from early childhood experiences. In short, attitudes and beliefs were related to underlying personality trends. The investigators, among other things, studied anti-Semitism as part of a general characteristic called ethnocentrism. Later, the investigators extended their thought and work to a still larger construct, authoritarianism, which they conceived to be a broad personality syndrome that determines in part ethnocentrism, social attitudes, and certain other behaviors. The authoritarian personality was conceived to be conventional, cynical, destructive, aggressive, power-centered, and ethnocentric.

While this is an inadequate summary of the basic problems of a complex study, it is sufficient for the present purpose. The study had to be nonexperimental because authoritarianism and ethnocentrism, as defined, are nonmanipulable variables, as are most of the variables related to authoritarianism. One of the major results of the study, for instance,

---

1 Careful scientific investigators will usually not say "cause." The word "cause" is used here to make the point more emphatic and because authoritative sources so use it: see The New York Times, Dec. 6, 1959, p. E-11, where the Surgeon General of the United States Public Health Service was directly quoted as saying: "the weight of evidence at present implicates smoking as the principal etiological factor in the increased incidence of lung cancer."


3 For evidence that the authors' theory about the characteristics of the authoritarian personality and its measurement was in general well-conceived, see F. Kerlinger and M. Rotch, "The Factorial Nature of the F and D Scales," Journal of Personality and Social Psychology, 4 (1966), 391-399.
was information on the relation between authoritarianism and prejudice. It is obvious that when one studies such variables one is studying already existing sets of personality char-
acteristics and attitudes. The subjects are ready-made authoritarians or nonauthoritarians
(with gradations between) and come to the research with already well-formulated atti-
dues. One can conceive, somehow, of manipulating such variables, but the manipulation,
as indicated previously, changes their nature. Whenever one studies the relations between
variables that “already exist” in the individuals studied, or whenever one studies the
determinants of such variables, one is deeply embedded in nonexperimental research and its
problems.

Determinants of School Achievement

A large preoccupation of educational researchers has been a search for the determinants of
school achievement. What are the factors that lead to successful achievement in school—
and unsuccessful achievement? Intelligence is an important factor, of course. While
measured intelligence, especially verbal ability, accounts for a large proportion of the
variance of achievement, there are many other variables, psychological and sociological:
sex, race, social class, aptitude, environmental characteristics, school and teacher charac-
teristics, family background, teaching methods. The study of achievement is character-
ized by both experimental and nonexperimental approaches. We are here concerned only
with the latter since it clearly illustrates problems of nonexperimental research.

In 1966 the now famous Coleman report was published. 6 As its title indicates, it was a
large-scale attempt to answer the question: Do American schools offer equal educational
opportunity to all children? Equally important, however, was the question of the relation
between student achievement and the kinds of schools students attend. This study was a
massive and admirable effort to answer these questions (and others). Its most famous and
controversial finding was that the differences among schools account for only a minor
fraction of the differences in school achievement. Most achievement variance was ac-
counted for by what the children bring with them to school. There was much to question
about the study's methodology and conclusions. 7 Indeed, its reverberations are still with
us. The principal dependent variable was verbal achievement. There were, however, more
than 100 independent variables. The authors used relatively sophisticated multivariate
procedures to analyze the data. Much of the core of the analytic problems, the interpreta-
tions of the findings, and the subsequent critiques inher in the nonexperimental nature of
the research.

The controversial conclusion mentioned above of the relative importance of home
background variables and school variables depends on a completely reliable and valid
method for assessing relative impacts of different variables. In experimental research, one
is safer drawing comparative conclusions because the independent variables are not corre-

352 • Types of Research

lated. In the real educational world, however, the variables are correlated, making their
unique contributions to achievement hard to determine. While there are statistical methods
to handle such problems, no methods can tell us unambiguously that X1 influences Y to
this or that extent because the real influence may be X2, which influences both X1 and Y.
The “correct” interpretation of the findings of Equality and studies like it is always
unattainable. While there are powerful analytic methods to use with nonexperimental
data, unequivocal answers to questions of the determinants of achievement are forever
beyond reach.

Participation in Political Processes

How does the participation of citizens in a democracy influence governmental processes?
This difficult question was fundamental to Verba and Nie’s large, complex, and sophisti-
cated study of political participation in American democracy. 8 A representative sample of
over 2,500 residents of the United States was asked questions about political participation,
information on political and local leaders, opinions on political efficacy of citizens, and
many others. The researchers found that participation does indeed make a difference.
Political leaders are more responsive to citizens who participate more. But participants
were not a representative group: the participant population tends to be the more affluent
and better educated people in higher status occupations. Verba and Nie found, then, that
a socioeconomic model of participation and the influence of participation on government
was appropriate.

This fine study’s findings are probably correct, but the ever-present possibility in all
research and especially in nonexperimental research is that independent variables other
than the one under study are the “real” source of the variance of the dependent
variable. I deliberately selected what I thought was a major study with excellent method-
ology to underline the difficulties of nonexperimental research. For the obtained important
relation found between citizen participation and political leader responsiveness, for exam-
ple, might the substantial relation be due not to citizen participation and social status but,
say, to the (presumed) fact that citizens who participate more are also upper social status
people and so are the political leaders? That is, the leaders are responsive not so much
because citizens participate at a high level but because higher social status citizens partici-
pate more than lower status citizens, and leaders respond to the social status—and its
accompanying education, influence, and attitudes—rather than the participation as such.
The participation, in other words, is a variable that “helps” make social status visible to
leaders. 9 Another reservation, as the authors themselves bring out, is that the concurrence
of leaders and citizens was not measured in urban cores. Is it possible that in such cores
the relation is negligible?

SMALLER SCALE NONEXPERIMENTAL RESEARCH

To illustrate nonexperimental behavioral research studies or series of studies is not easy:
there are many, many of them, but few of them satisfy my criteria of methodological
soundness and substantive interest. I have chosen the following three studies for three


7See ibid., pp. 336–337, especially Table 20-1, where the above argument actually turns out to be in part correct.


reasons. One, they each represent a unique, original, and interesting approach to an important sociological, psychological, or educational problem. Two, each contributes significantly to scientific knowledge. And three, each is nonexperimental.

Kounin: The Management of Classrooms

Again, this example is really a set of researches all directed to the same general question, which can be loosely expressed: What effects do teachers' classroom management procedures have on children in classrooms? The research of Kounin and his colleagues has been characterized by original and significant variables, both independent and dependent, by systems of extensive observations of the behavior in classrooms of teachers and pupils, specifically aimed at measuring the variables, and by careful operationalization of the variables. In one such study, for example, Kounin and Doyle analyzed videotapes of 596 formal lessons taught by 36 teachers in a preschool aimed at measuring the variables lesson continuity and task involvement. The hypothesis tested was simple: the more continuous and unlagging a lesson, the greater the task involvement of children.

During lessons, observers coded children's behavior for appropriate involvement every six seconds. Percentages of involvement scores were calculated. Continuity of lessons was measured by noting and timing child recitations, which were thought to be more discontinuous than the "official" teacher reading and demonstrating. In other words, if a lesson had a high proportion of child recitations, it was considered discontinuous. High task-involvement and low task-involvement lessons were defined by observers, who were required to use percentages of child-recitation times. The mean percentages of child recitation for high task-involvement lessons was 8.40; the mean for low task-involvement lessons was 20.20. The difference was statistically significant. A similar analysis of demonstration lessons yielded similar results. The authors concluded: "measured degrees of continuity within lesson types distinguished between those lessons that had high task involvement when manned by the same occupants." (p. 163)

This is interesting and potentially important research—and all the Kounin studies are characterized by an imaginative yet objective approach to teacher and pupil observation. Unfortunately, none of the studies is experimental. How much more convincing the relation between lesson continuity and task involvement, for example, if lesson continuity had been experimentally manipulated and task involvement had been substantially affected thereby? Questions can and should also be raised about the selection of lessons for analysis and about the analysis of the data. (It seems, again, that independent variable measures were analyzed rather than, as one would normally expect, measures of the dependent variable.) This possible error, incidentally, could probably not have happened if the research had been experimental since there would have been k experimental groups in which different amounts of continuity (or discontinuity) had been engendered and the analysis performed of the task-involvement scores and means of these groups.

McClelland: Protestantism, Capitalism, and Achievement

There has been much thought and speculation on the relations among capitalism, Protestantism, and achievement. Max Weber, for example, wrote an important book on capitalism and Protestantism. 11 His basic hypothesis was that Protestantism led to the spirit of capitalism because the Protestant ethic—self-reliance, deference of enjoyment, asceticism, emphasis on achievement, and so on—produced individuals with the qualities necessary to capitalistic enterprise and development. McClelland, in a remarkable book, has described his many studies on the relation between Protestantism and capitalism. 12 Actually, McClelland's main interest has been on motivation and its measurement, and his research can safely be called one of the successes of psychology. The variable of his principal interest and work has been achievement motivation, commonly called n Achievement, or n Ach, which he has measured by asking individuals to write brief stories suggested by pictures shown them for a few seconds, the pictures representing situations related to work. The stories were then content-analyzed using a complex coding system of several scores, as an index of Achievement for each individual. In the present study n Achievement was an independent variable among several independent variables used to predict the economic growth of nations.

McClelland's hypothesis was that countries whose population is predominantly Protestant will emphasize achievement. Protestant countries should thus grow capitalistic enterprise than Catholic countries, other things equal. "Capitalistic enterprise" is reflected in economic advance, growth, or development. "Other things equal" means controlling such things as natural resources. Some countries, for example, may be economically more advanced simply because they have greater natural resources. The measure of economic growth, the dependent variable, was electricity production. McClelland's analysis is too complex to explain here. Suffice it to say that he found that Protestant countries were economically more advanced than Catholic countries, as reflected in electricity production, thus supporting the hypothesis. He also found that n Achievement was significantly correlated with deviations from expected growth. (The correlation was .43: ibid., p. 100, Table 3.6. See pp. 88-89 for the reason he used deviation scores, or measures of over- and underachievement, rather than the electricity production scores themselves.)

The research reported in The Achieving Society is an excellent example of sophisticated approaches to theoretical and empirical problems of large consequence. The difficulties of collecting, analyzing, and interpreting nonexperimental empirical data are well illustrated in this remarkable book. One emerges from its study convinced of the validity of McClelland's theoretical explanations and the soundness of his analyses and interpretations of the findings. I know of no more ambitious and competent studies of complex phenomena and the testing of plausible alternative hypotheses. Yet we must constantly bear in mind the nonexperimental nature of the research and the enormous difficulty of

11J. Kounin, Discipline and Group Management in Classrooms. New York: Holt, Rinehart and Winston, 1970. Other interesting studies have been published since this book's appearance.


chapters, we can also ‘confirm’ and ‘disconfirm’ hypotheses under study by trying to show that alternative plausible hypotheses are or are not supported. First consider alternative independent variables as antecedents of a dependent variable. The reasoning is the same. If we say ‘alternative independent variables,’ for example, we are in effect stating alternative hypotheses or explanations of a dependent variable.

In nonexperimental studies, although one cannot have the confidence in the ‘truth’ of an ‘If A, then B’ statement that one can have in experiments, it is possible to set up and test alternative or ‘control’ hypotheses. (Of course, alternative hypotheses can be and are tested in experimental studies, too.) This procedure has been formalized and explained by Platt who, influenced by Chamberlin, called it ‘strong inference’.18 Chamberlin aptly called the procedure the ‘method of working multiple hypotheses,’ and he outlined how the investigator’s own ‘intellectual affections’ can be guarded against. He said: ‘The effort is to bring up into view every rational explanation of the new phenomena, and to develop every tenable hypothesis respecting their cause and history. The investigator thus becomes the parent of a family of hypotheses; and, by his parental relation to all, he is forbidden to fatten his affections unduly upon any one.’

Let \( x_1, x_2, \) and \( x_3 \) be three alternative independent variables, and let \( y \) be the dependent variable, the phenomenon to be ‘explained’ with a statement of the form: ‘If \( x_i \), then \( y \).’ Assume that \( x_1, x_2, \) and \( x_3 \) exhaust the possibilities. This assumption cannot actually be made—in scientific research it is practically impossible to exhaust all the possibilities. Still, it is assumed for pedagogical reasons.

An investigator has evidence that \( x_1 \) and \( y \) are substantially related. Having reason to believe that \( x_3 \) is another relevant factor, he holds \( x_1 \) and \( x_2 \) constant. He is assuming that one of the three factors is the factor, that either \( x_1 \) or \( x_2 \) or \( x_3 \) is the ‘true’ independent variable. (Again, note the assumption. It may be none of them or some combination of all three.) Suppose that the investigator succeeds in eliminating \( x_2 \), that is, he shows that \( x_2 \) is not related to \( y \). If he also succeeds in eliminating \( x_3 \), he can then conclude that \( x_1 \) is the influential independent variable. Since the alternative or ‘control’ hypotheses have not been substantiated, the original hypothesis is strengthened.

Similarly, we can depend on alternative independent variables, which also imply alternative hypotheses. We shift the alternatives to the dependent variable. This is illustrated in a study by Alper, Blane, and Abrams of the different reactions of middle- and lower-class children to finger pains as a consequence of different child-rearing practices.19 The general question asked was: Do social-class differences in child-training practices result in class differences in personality? The theory invoked required that there be differences in reactions to finger pains. The authors reasoned that middle-class children would react differently to pain in children to 16 different variables. For example, one finger pain point was ‘used: acceptance of task, washing, and so on’. The reactions were significantly different on most of the variables. In a ‘control experiment,’ the same procedure was followed using crayons instead of finger pains. The two groups did not differ significantly on any of the 11 variables measured, in surprising contrast to the finger pain results. The study was nonexperimental because it was not possible to manipulate the independent variable and because the children came to the study with their reactions ready-made, as it were.


This use of a control study was ingenious and crucial. Imagine the researchers' consterna-
tion if the differences between the two groups on the crayon task had been significant!
Now consider a study by Sarnoff et al., in which it was predicted that English and
American children would differ significantly in test anxiety but not in general anxiety.19
The hypothesis was carefully delineated: If eleven-plus examinations are taken, then test
anxiety results. (The eleven-plus examinations are given to English school children at
eleven years of age to help determine their educational futures.) Since it was possible that
there might be other independent variables causing the difference between the English and
American children on test anxiety, the investigators evidently wished to rule out at least
some of the major contenders. This they accomplished by carefully matching the samples;
they probably reasoned that the difference in test anxiety might be due to a difference in
general anxiety, since the measure of test anxiety obviously must reflect some general
anxiety. If this were found to be so, the major hypothesis would not be supported. There-
fore Sarnoff and his colleagues, in addition to testing the relation between examination
and test anxiety, also tested the relation between examination and general anxiety.

The method of testing alternative hypotheses, although important in all research, is
particularly important in nonexperimental studies, because it is one of the only ways to
"control" the independent variables of such research. Lacking the possibility of randomi-
zation and manipulation, nonexperimental researchers, perhaps more so than experimen-
talists, must be very sensitive to alternative hypothesis-testing possibilities.

EVALUATION OF NONEXPERIMENTAL RESEARCH

The reader may have concluded from the preceding discussion that nonexperimental re-
search is inferior to experimental research, but this conclusion would be unwarranted. It
is easy to say that experimental research is "better" than nonexperimental research, or
that experimental research tends to be "trivial," or that nonexperimental research is
"merely correlational." Such statements, in and of themselves, are oversimplifications.
What the student of research needs is a balanced understanding of the strengths and
weaknesses of both kinds of research. To be committed unequivocally to experimentation
or to nonexperimental research may be shortsighted.

The Limitations of Nonexperimental Interpretation

Nonexperimental research has three major weaknesses, two of which have already been
discussed in detail: (1) the inability to manipulate independent variables, (2) the lack of
power to randomize, and (3) the risk of improper interpretation. In other words, compared
to experimental research, other things equal, nonexperimental research lacks control; this
lack is the basis of the third weakness: the risk of improper interpretation.

The danger of improper and erroneous interpretations in nonexperimental research
stems in part from the plausibility of many explanations of complex events. It is easy to
accept the first and most obvious interpretation of an established relation, especially if one
works without hypotheses to guide investigation. Research unguided by hypotheses, re-
search "to find out things," is most often nonexperimental. Experimental research is
more likely to be based on carefully stated hypotheses.

Hypotheses are if-then predictions. In a research experiment the prediction is from a
well-controlled x to y. If the prediction holds true, we are relatively safe in stating the


conditional, If x, then y. In a nonexperimental study under the same conditions, however,
we are considerably less safe in stating the conditional, for reasons discussed earlier.
Careful safeguards are more essential in the latter case, especially in the selection and
testing of alternative hypotheses, such as the supposed lack of relation between the ele-
ven-plus examination and general anxiety in the Sarnoff study. A predicted (or unpre-
dicted) relation in nonexperimental research may be quite spurious, but its plausibility and
conformity to preconception may make it easy to accept. This is a danger in experimental
research, but it is less of a danger than it is in nonexperimental research because an
experimental situation is so much easier to control.

Nonexperimental research that is conducted without hypotheses, without predictions,
research in which data are just collected and then interpreted, is even more dangerous in
its power to mislead. Significant differences or correlations are located if possible and
then interpreted. Assume that an educator decides to study the factors leading to under-
achievement. He selects a group of underachievers and a group of normal achievers and
administers a battery of tests to both groups. He then calculates the means of the two
groups on the tests and analyzes the differences with t tests. Among, say, twelve such
differences, three are significant. The investigator concludes, then, that underachievers
and normal achievers differ on the variables measured by these three tests. Upon analysis
of the three tests, he thinks he understands what characterizes underachievers. Since all
three of the tests seem to measure insecurity, the cause of underachievement is therefore
insecurity.

When guided by hypotheses the credibility of the results of studies like the one just
cited may be enhanced, but the results are still weak because they capitalize on chance: by
chance alone one or two results of many statistical tests may be significant. Above all,
plausibility can be misleading. A plausible explanation often seems compelling—even
though quite wrong! It seems so obvious, for example, that conservatism and liberalism are
opposites. The research evidence, however, seems to indicate that they are not op-
posites.20 Another difficulty is that plausible explanations, once found, are often hard
to test. According to Merton, post factum explanations do not lend themselves to
nullifiability because they are so flexible. Whatever the observations, he says, new
interpretations can be found to "fit the facts."21

The Value of Nonexperimental Research

Despite its weaknesses, much nonexperimental research must be done in psychology,
sociology, and education simply because many research problems do not lend themselves
to experimentation. A little reflection on some of the important variables in behav-
ioral research—intelligence, aptitude, home background, achievement, social class, ri-
gidity, ethnocentrism—will show that they are not manipulable. Controlled inquiry is
possible, of course, but true experimentation is not.

It can even be said that nonexperimental research is more important than experimental
research. This is, of course, not a methodological observation. It means, rather, that most
social scientific and educational research problems do not lend themselves to experimen-
tation, although many of them do lend themselves to controlled inquiry of the nonexperi-
mental kind. Consider Piaget's studies of children's thinking, the authoritarianism studies

20See F. Kerlinger, "Social Attitudes and Their Criterial Referents: A Structural Theory," Psychological

ADDITIONAL

Causality and Scientific Research

The study and analysis of "causal relations" in research has recently preoccupied social scientists. Economists and econometricians seem to have been the leaders in this work.22 Simon, a psychologist, has also been a pioneer.23 Sociologists have written extensively on the subject.24 The analytic and conceptual movement is productive, as we will see in the later chapter. It is perhaps unfortunate that the word "cause" and "causal relations" have been used. They imply that science can find the causes of phenomena. The position taken in this book is that the study of cause and causation is an endless maze. One of the difficulties is that the word "cause" has surplus meaning and metaphysical overtones. Perhaps more important, it is not really needed. Scientific research can be done without invoking cause and causal explanations, even though the words and other words that imply cause are almost impossible to avoid and thus will occasionally be used. Blalock points out that causal laws cannot be demonstrated empirically, but that it is helpful to think causally.25 I agree that causal laws cannot be demonstrated empirically, but am equivocal about thinking causally. There is little doubt that scientists do think causally and that when they talk of a relation between $p$ and $q$ they hope or believe that $p$ causes $q$. But no amount of evidence can demonstrate that $p$ causes $q$.

This position is not so much an objection to causal notions as it is an affirmation that they are not necessary to scientific work. Evidence can be brought to bear on the empirical validity of conditional statements of the form $p$, then $q$; alternative hypotheses can be tested, and probabilistic statements can be made about $p$ and $q$ and other $p*$ and $q*$ and conditions $r$, $s$, and $t$. Invocation of the word "cause" and the expression "causal relation" does nothing really constructive. Indeed, it can be misleading.

In expert hands and used with circumspection, path analysis and related methods can help to clarify theoretical and empirical relations.26 But when their espousal and use imply that causes are sought and found, such methods can also be misleading. In sum, the elements of deductive logic in relation to conditional statements, a probabilistic framework, and the testing of alternative hypotheses are sufficient aids to scientific nonexperimental work without the excess baggage of causal notions and methods presumably geared to strengthening causal inferences. We rest the case with some apt words of Bertrand Russell:

the word "cause" is so inextricably bound up with misleading associations as to make its complete extirpation from the philosophical vocabulary desirable . . . . the reason physics has ceased to look for causes is that, in fact, there are no such things. The law of causality . . . is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm.27

25Ibid., p. 6.
in education is mixed: experimental and noneperimental. Political science is also mixed. Anthropology is mostly noneperimental. (Think, too, of economics, astronomy, physics, chemistry, and biology.) Why are some sciences predominantly experimental and others noneperimental? Explain specifically what you mean.

8. The venturesome student may wish to take a plunge into stimulating, provocative, controversial, and important thinking. The famous Club of Rome report has outraged some observers, startled almost anyone who has read it, and disturbed everyone. Using societally important variables—natural resources, pollution, population, for example—and their complex interactions, ultimate disaster to cities and the world has been predicted. The research on which the conclusions are based is entirely noneperimental. Try reading the report and perhaps one of the works of the pioneer of the area of study, Professor Forrester of MIT. Do you think that the research's noneperimental character lowers its credibility?

28 McClelland, op. cit., p. 100, Table 3.6.


Chapter 23

Laboratory Experiments, Field Experiments, and Field Studies

Social scientific research can be divided into four major categories: laboratory experiments, field experiments, field studies, and survey research. This breakdown stems from two sources, the distinction between experimental and nonexperimental research and that between laboratory and field research.

A LABORATORY EXPERIMENT: MILLER STUDIES OF THE LEARNING OF VISCERAL RESPONSES

A brilliant series of experiments by Neal Miller and his colleagues has upset another long-held and well-cherished belief: that learning occurs only with voluntary responses, and that the involuntary autonomic system is subject only to classical conditioning. This means, in effect, that responses like moving the hand and talking can be brought under control and thus taught, but that involuntary responses, like heart rate, intestinal contractions, and blood pressure, cannot be brought under instrumental control and thus not "taught." Miller and his colleagues' work has shown that, through instrumental conditioning, the heart rate can be changed, stomach contractions can be altered, and even urine formation can be increased or decreased! This discovery is of enormous theoretical and practical importance. To show the nature of laboratory experiments, we take one of Miller's interesting and creative experiments.

The idea of the experiment is simple: reward one group of rats when their heart rates go up, and reward another group when their heart rates go down. This is a straightforward example of the two-group design discussed earlier. Miller's big problem was control. There are a number of other causes of changed heart rate—for example, muscular exertion. To control such extraneous variables, Miller and a colleague (Trowill) paralyzed the rats with curare. But if the rats were paralyzed, what could be used as reward? They decided to use direct electrical stimulation of the brain. The dependent variable, heart rate, was continuously recorded with the electrocardiograph. When a small change in heart rate occurred (in the "right" way: up for one group, down for the other), an animal was given an electrical impulse to a reward center of its brain. This was continued until the animals were "trained."

The increases and decreases of heart rate were statistically reliable but small: only five percent in each direction. So Miller and another colleague (DiCaro) used the technique known as shaping, which, in this case, means rewarding first small changes and then requiring increasingly large changes to obtain the rewards. This increased the heart rate changes to an average of 20 percent in either direction. Moreover, further research, using escape from mild shock as reinforcement, showed that the animals remembered what they had learned and "differentiated" the heart responses from other responses.

Miller has been successful in "training" a number of other involuntary responses: intestinal contraction, urine formation, and blood pressure, for example. In short, visceral responses can be learned and can be shaped. But can the method be used with people? Miller says that he thinks people are as smart as rats, but that it has not yet been completely proved. Although the use of curare might present difficulty, people can be hypnotized, says Miller.

A FIELD EXPERIMENT: WALSTER, CLEARY, AND CLIFFORD'S STUDY OF BIAS IN COLLEGE ADMISSIONS

Do American colleges discriminate against women and black applicants? We used the fine field experiment of Walster, Cleary, and Clifford in Chapter 14 to illustrate factorial

To understand Miller's studies, we must define certain psychological terms. In classical conditioning a neutral stimulus, inherently unable to produce a certain response, becomes able to by being associated repeatedly with a stimulus inherently capable of producing it. The most famous example is Pavlov's dog salivating at the clicking of a metronome, which had been repeatedly associated with meat powder. In instrumental conditioning, a reinforcement given to an organism immediately after it has made a response produces an increment in the response. Pigeons, for example, will peck their beaks bloody after having been subjected to certain forms of instrumental conditioning. In short, reward a response and it will be repeated. Voluntary responses or behavior are thought to be superior, presumably because they are under the control of the individual, whereas involuntary responses are inferior because they are not. It has been believed that involuntary responses can be modified only by classical conditioning and not by instrumental conditioning. In other words, the possibilities of "teaching" the heart, the stomach, and the blood are remote, since classical conditioning conditions are difficult to come by. If the organs are subject to instrumental conditioning, however, they can be brought under experimental control, they can be "taught," and they can "learn." For authoritative accounts of both kinds of conditioning and their relation to learning, see E. Hilgard and G. Bower, Theories of Learning, 4th ed. Englewood Cliffs, N.J.: Prentice-Hall, 1975, chaps. 3, 4, 5. 7

analysis of variance. Hence it is not necessary to labor all the study details here.\textsuperscript{5} Recall that the authors randomly selected 240 colleges from a college guide and sent the colleges application letters from fictitious individuals. In the letters, they manipulated applicants’ race, sex, and ability levels (three such levels). For example, a candidate might be a \textit{black female of high ability}, or a \textit{white male of medium ability}. This is, of course, a $2 \times 2 \times 3$ factorial design. The letter to each college was from a “candidate” who represented one cell of the 12 cells of the design. The dependent variable measure was obtained by quantifying the degrees of the colleges’ acceptances of candidates. This amounted to a five-point scale: (1) rejection; (2) rejection, but qualified; \ldots ; (5) acceptance, with encouragement.

Race and sex were found not to be significant. From these results alone one might conclude that there was no bias in admissions. But recall that the interaction of race and ability was significant. Although there were no differences in admissions at the high and medium levels of ability, the mean acceptances of men and women at the low ability level were significantly different. The male mean was 3.00, and the female mean was 1.93. (See Table 14.14.) So evidently there was discrimination: women were discriminated against at the low ability level!

\section*{A FIELD STUDY: NEWCOMB’S BENNINGTON COLLEGE STUDY}

In one of the most important studies of the influence on students of a college environment, Newcomb studied the entire student body of Bennington College. About 600 young women, from 1935 to 1939\textsuperscript{6} An unusual facet of the study was Newcomb’s attempt to explain both social and personality factors in influencing attitude changes in the students. Although other hypotheses were tested, the principal hypothesis of the Bennington study was that new students would conform to the norms of the college group, and that the more the students assimilated to the college community, the greater would be the change in their social attitudes.

Newcomb used a number of paper-and-pencil attitude scales, written reports on students, and individual interviews. The study was longitudinal and nonexperimental. The independent variable, while not easy to categorize, can be said to be the social norms of Bennington College. The dependent variables were social attitudes and certain behaviors of the students.

Newcomb found significant changes in attitudes between freshmen, on the one hand, and juniors and seniors, on the other. The changes were toward less conservatism on a variety of social issues. For example, the political preferences of freshmen and seniors in the 1936 presidential election were much less conservative than those of freshmen and sophomores. Of 52 juniors and seniors, 15 percent preferred Landon (Republican), whereas of 52 freshmen, 62 percent preferred Landon. The percentages of preferences for Roosevelt (Democrat) were 54 percent and 29 percent. The mean scores of all students for four years on a scale designed to measure political and economic conservatism were: freshmen, 74.2; sophomores, 69.4; juniors, 65.9; and seniors, 62.4. Evidently the college had affected the students’ attitudes.


\textsuperscript{6} T. Newcomb. Personality and Social Change. New York: Holt, Rinehart and Winston, 1943. A number of the Bennington students were later restudied in follow-up research designed to test the permanence of the changes made by Bennington: T. Newcomb, K. Koenig, R. Flicks, and D. Warwick. \textit{Persistence and Change at Bennington College and Its Students After Twenty-Five Years}. New York: Wiley, 1967. In general, it was found that the changes had lasted; evidently Bennington’s influence was persistent over the years.

Laboratory Experiments, Field Experiments, and Field Studies \textsuperscript{367}

Newcomb asked a “control” question: Would these attitudes have changed in other colleges? To answer this question, Newcomb administered his conservatism measures to students of Williams College and Skidmore College. The comparable mean scores of Skidmore students, freshmen through seniors, were: 79.9, 78.1, 77.0, and 74.1. It seems that Skidmore (and Williams) students did not change as much and as consistently over time as did the Bennington students.

\section*{CHARACTERISTICS AND CRITERIA OF LABORATORY EXPERIMENTS, FIELD EXPERIMENTS, AND FIELD STUDIES}

A laboratory experiment is a research study in which the variance of all or nearly all of the possible influential independent variables not pertinent to the immediate problem of the investigation is kept at a minimum. This is done by isolating the research in a physical situation apart from the routine of ordinary living and by manipulating one or more independent variables under rigorously specified, operationalized, and controlled conditions.

\section*{Strengths and Weaknesses of Laboratory Experiments}

The laboratory experiment has the inherent virtue of the possibility of relatively complete control. The laboratory experimenter can, and often does, isolate the research situation from the life around the laboratory by eliminating the many extraneous influences that may affect the independent and dependent variables.

In addition to situation control, laboratory experimenters can ordinarily use random assignment and can manipulate one or more independent variables. There are other aspects to laboratory control: the experimenter in most cases can achieve a high degree of specificity in the operational definitions of his variables. The relatively crude operational definitions of field situations, such as many of those associated with the measurement of values, attitudes, aptitudes, and personality traits, do not plague the experimentalist, though the definitional problem is never simple. The Miller experiment is a good example. The operational definitions of reinforcement and heart rate change are precise and highly objective.

Closely allied to operational strength is the precision of laboratory experiments. Precision means accurate, definite, unambiguous. Precise measurements are made with precision instruments. In variance terms, the more precise an experimental procedure is, the less the error variance. The more accurate or precise a measuring instrument is, the more certain we can be that the measures obtained do not vary much from their "true" values. This is the problem of reliability, which will be discussed in a later chapter.

Precise laboratory results are achieved mainly by controlled manipulation and measurement in an environment from which possible "contaminating" conditions have been eliminated. Research reports of laboratory experiments usually specify in detail how the manipulations were done and the means taken to control the environmental conditions under which they were done. By specifying exactly the conditions of the environment, we reduce the risk that subjects may respond equivocally and thus introduce random variance into the experimental situation. Miller’s experiment is a model of laboratory experimental precision.

The greatest weakness of the laboratory experiment is probably the lack of strength of independent variables. Since laboratory situations are, after all, situations that are created for special purposes, it can be said that the effects of experimental manipulations are
usually weak. Increases and decreases in heart rate by electrical brain reinforcement, while striking, were relatively small. Compare this to the relatively large effects of independent variables in realistic situations. In the Bennington study, for example, the college community apparently had a massive effect. In laboratory research on conformity, only small effects are usually produced by group pressure on individuals. Compare this to the relatively strong effect of a large group majority on an individual group member in a real-life situation. The board of education member, who knows that an action he wants carried goes against the wishes of the majority of his colleagues, understands that the majority of the community is under heavy pressure to converge on the norm.

One reason for the preoccupation with laboratory precision and refined statistics is the weakness of laboratory effects. To detect a significant difference in the laboratory requires situations and measures with a minimum of random noise and accurate and sensitive statistical tests that will show relations and significant differences when they exist.

Another weakness is a product of the first: the artificiality of the experimental research situation. Actually, it is difficult to know if artificiality is a weakness or simply a neutral characteristic of laboratory experimental situations. When a research situation is deliberately contrived to exclude the many distractions of the environment, it is perhaps illogical to label the situation with a term that expresses in part the result being sought. The criticism of artificiality does not come from experimenters, who know that experimental situations are artificial; it comes from individuals lacking an understanding of the purposes of laboratory experiments.

The temptation to interpret the results of laboratory experiments incorrectly is great. While Miller's results are believed by social scientists to be highly significant, they can only tentatively be extrapolated beyond the laboratory. Similar results may be obtained in real-life situations, and there is evidence that they do in some cases. But this is not necessarily so. The relations must always be tested anew under nonlaboratory conditions. Miller's research, for instance, will have to be carefully and cautiously done with human beings in hospitals and even in schools.

Although laboratory experiments have relatively high internal validity, then, they lack external validity. Earlier we asked the question: Did X, the experimental manipulation, really make a significant difference? The stronger our confidence in the "truth" of the relations discovered in a research study, the greater the internal validity of the study. When a relation is discovered in a well-executed laboratory experiment, we generally can have considerable confidence in it, since we have exercised the maximum possible control of the independent variable and other possible extraneous independent variables. When Miller "discovered" that visceral responses could be learned and shaped, he could be relatively sure of the "truth" of the relation between reinforcement and visceral response—in the laboratory. He had achieved a high degree of control and of internal validity.

One can say: If I study this problem using field experiments, maybe I will find the same relation. This is an empirical, not a speculative, matter; we must put the relation to test in the situation to which we wish to generalize. If a researcher finds that individuals converge on group norms in the laboratory, as Sherif did, does the same or similar phenomenon occur in community groups, faculties, legislative bodies? This lack of external validity is the basis of the objections of many educators to the animal studies of learning theory. Their objections are only valid if an experimenter generalizes from the behavior and learning of laboratory animals to the behavior and learning of children.


Capable experimentalists, however, rarely blunder in this fashion—they know that the laboratory is a contrived environment.

**Purposes of the Laboratory Experiment**

Laboratory experiments have three related purposes. One, they are a means of studying relations under "pure" and uncontaminated conditions. Experimenters ask: Is x related to y? How is it related to y? How strong is the relation? Under what conditions does the relation change? They seek to write equations of the form y = f(x), make predictions on the basis of the function, and see how well and under what conditions the function performs.

A second purpose should be mentioned in conjunction with the first purpose: the testing of predictions derived from theory, primarily, and other research, secondarily. For instance, on the basis of Sherif's norm-convergence finding, one might predict to a number of other laboratory and field experimental situations, as Sherif did in his later studies of boys in camp situations. Asch, though, argued that Sherif's stimulus was ambiguous in the sense that different people would "interpret" it differently. He wondered whether the convergence phenomenon would work with clear stimuli in a more realistic setting. A series of experiments showed that it did.

A third purpose of laboratory experiments is to refine theories and hypotheses, to formulate hypotheses related to other experimentally or nonexperimentally tested hypotheses, and, perhaps most important, to help build theoretical systems. This was one of Miller's and Sherif's major purposes. Although some laboratory experiments are conducted without this purpose, of course, most laboratory experiments are theory-oriented.

The aim of laboratory experiments, then, is to test hypotheses derived from theory, to study the precise interrelations of variables and their operation, and to control variance under research conditions that are uncontaminated by the operation of extraneous variables. As such, the laboratory experiment is one of the great inventions of all time. Although weaknesses exist, they are weaknesses only in a sense that is really irrelevant. Concealing the lack of representativeness (external validity) the well-done laboratory experiment still has the fundamental prerequisite of any research: internal validity.

**THE FIELD EXPERIMENT**

A field experiment is a research study in a realistic situation in which one or more independent variables are manipulated by the experimenter under as carefully controlled conditions as the situation will permit. The contrast between the laboratory experiment and the field experiment is not sharp: the differences are mostly matters of degree. Sometimes it is hard to label a particular study "laboratory experiment" or "field experiment." Where the laboratory experiment has a maximum of control, most field experiments must operate with less control, a factor that is often a severe handicap.

---


Strengths and Weaknesses of Field Experiments

Field experiments have values that especially recommend them to social psychologists, sociologists, and educators because they are admirably suited to many of the social and educational problems of interest to social psychology, sociology, and education. Because independent variables are manipulated and randomization used, the criterion of control can be satisfied—at least theoretically.

The control of the experimental field situation, however, is rarely as tight as that of the laboratory. We have here both a strength and a weakness. The investigator in a field experiment, though he has the power of manipulation, is always faced with the unpleasant possibility that his independent variables are contaminated by uncontrolled environmental variables. We stress this point because the necessity of controlling extraneous independent variables is particularly urgent in field experiments. The laboratory experiment is conducted in a tightly controlled situation, whereas the field experiment takes place in a natural, often loose, situation. One of the main preoccupations of the field experimenter, then, is to try to make the research situation more closely approximate the conditions of the laboratory experiment. Of course this is often a difficult goal to reach, but if the research situation can be kept tight, the field experiment is powerful because one can in general have greater confidence that relations are indeed what one says they are.

As compensation for dilution of control, the field experiment has two or three unique virtues. The variables in a field experiment usually have a stronger effect than those of laboratory experiments. The effects of field experiments are often strong enough to penetrate the distractions of experimental situations. The principle is: The more realistic the research situation, the stronger the variables. This is one advantage of doing research in educational settings. For the most part, research in school settings is simpler than routine educational activities, and thus need not necessarily be viewed as something special and apart from school life. Despite the pleas of many educators for more realistic educational settings, field experimenters must necessarily have to be trained, not necessarily as particularly skilled operators. They should be able to work with people, talk to them, and convince them of the importance and necessity of their research. They should be prepared to spend many hours, even days and weeks, of patient discussion with people responsible for the institutional or community situation in which they are to work. For instance, if they are to work in a rural school system, they should have knowledge of rural as well as general educational problems, and of the particular rural system they wish to study. School researchers become impatient with these preliminaries, because they are anxious to get the research job done. They find it difficult to spend the time and effort necessary in most practical situations. Others enjoy the inevitable socializing that accompanies field research.

An important obstacle to good design, an obstacle that seems ordinarily to be overlooked, is the attitude of the researcher. For instance, the planning of educational research often seems to be characterized by a negative attitude epitomized by such statements as, "That can't be done in schools," "The administrators and teachers won't allow that," and "Experiments can't be done on this problem in that situation." Starting with attitudes like this compromises any good research design before the research even begins. If a research design calls for the random assignment of teachers to classes, and if the lack of such assignment seriously jeopardizes the internal validity of the proposed study, every effort should be made to assign teachers at random. Educators planning research seem to

assume that the administrators or the teachers will not permit random assignment. This assumption is not necessarily correct, however.

The consent and cooperation of teachers and administrators can often be obtained if a proper approach, with adequate and accurate orientation, is used, and if explanations of the reasons for the use of specific experimental methods are given. The points being emphasized are these: Design research to obtain valid answers to the research questions. Then, if it is necessary to make the experiment possible, and only then, modify the "ideal" design. With imagination, patience, and courtesy, many of the practical problems of implementation of research design can be satisfactorily solved.

One other weakness inherent in field experimental situations is lack of precision. In the laboratory experiment it is possible to achieve a high degree of precision or accuracy, so that laboratory measurement and control problems are usually simpler than those in field experiments. In realistic situations, there is always a great deal of systematic and random noise. In order to measure the effect of an independent variable on a dependent variable in a field experiment, it is not only necessary to maximize the variance of the manipulated variable and any assigned variables, but also to measure the dependent variable as precisely as possible. But in realistic situations, such as in schools and community groups, extraneous independent variables abound. And measures of dependent variables, unfortunately, are sometimes not sensitive enough to pick up the messages of our independent variables. In other words, the dependent variable measures are often so inadequate they cannot pick up all the variance that has been engendered by the independent variables.

FIELD STUDIES

Field studies are nonexperimental scientific inquiries aimed at discovering the relations and interactions among sociological, psychological, and educational variables in real social structures. In this book, any scientific studies, large or small, that systematically pursue relations and test hypotheses, that are nonexperimental, and that are done in life situations like communities, schools, factories, organizations, and institutions will be considered field studies.

The investigator in a field study first looks at a social or institutional situation and then studies the relations among the attitudes, values, perceptions, and behaviors of individuals and groups in the situation. He ordinarily manipulates no independent variables. Before we discuss and appraise the various types of field studies, it will be helpful to consider examples. We have already examined field studies in Chapter 22 and in this chapter: the Authoritarian Personality study, the Newcomb Bennington study, and others. We now briefly examine two smaller field studies.

Jones and Cook tested the socially and politically important hypothesis that preferences for social policies to advance racial equality are influenced by racial attitudes. More specifically, individuals with positive attitudes toward blacks will favor societal change policies and individuals with negative attitudes toward blacks will favor self-improvement policies. They measured attitudes toward blacks with a well-constructed and validated scale (we examine it in a later chapter) and independently by membership in four groups, such membership assumed and shown to be correlated with social attitudes (e.g., the group assumed to have the most favorable attitudes toward blacks had participated in civil rights and prointegration activities). The dependent variable, preference for social policy, was measured with a set of 30 policy items, each of which had two alternatives, one favoring social change and the other self-improvement. The hypothesis was supported: attitudes toward blacks evidentially influence judgments of effective social policies.

The second smaller field study, part of a larger study by McClelland on the relation between Protestantism and capitalistic growth, was summarized in Chapter 21. Recall that several variables—Protestant-Catholic, need for achievement, and electricity consumption (an index of capitalistic growth or development), among others—were measured in 25 countries in 1925 and 1950. It was found, as predicted, that Protestantism was positively related to capitalistic growth.

Note that the problems of both field studies were attacked nonexperimentally: neither randomization nor experimental manipulation was possible. Note, too, an important difference in the data-gathering methods of the two studies. In the Jones and Cook study, data were collected directly from students at two universities. In the McClelland study, however, data on the variables in the 25 countries were "indirectly" collected from published sources, mainly world population statistics. While some might argue that the Jones and Cook study data are stronger than the McClelland data because the former were collected "directly," there is really no difference in principle: both studies are nonexperimental field studies and both are creative and important contributions.

Types of Field Studies

Katz has divided field studies into two broad types: exploratory and hypothesis-testing. The exploratory type, says Katz, seeks what is rather than predicts relations to be found. The massive Equality of Educational Opportunity, cited in Chapter 22, exemplifies this type of field study. Exploratory studies have three purposes: to discover significant variables in the field situation, to discover relations among variables, and to lay the groundwork for later, more systematic and rigorous testing of hypotheses.

Throughout this book to this point, the use and testing of hypotheses have been emphasized. It is well to recognize, though, that there are activities preliminary to hypothesis-testing in scientific research. In order to achieve the desirable aim of hypothesis-testing, preliminary methodological and measurement investigation must often be done. Some of the finest work of the twentieth century has been in this area. An example is that done by the factor analyst, who is preoccupied with the discovery, isolation, specification, and measurement of underlying dimensions of achievement, intelligence, aptitudes, attitudes, situations, and personality traits.

The second subtype of exploratory field studies, research aimed at discovering or uncovering relations, is indispensable to scientific advance in the social sciences. It is necessary to know, for instance, the correlates of variables. Indeed, the scientific meaning of a construct springs from the relations it has with other constructs. Assume that we have no scientific knowledge of the construct "intelligence": we know nothing of its causes or concomitants. For example, suppose that we know nothing whatever about the relation of intelligence to achievement. It is conceivable that we might do a field study in school situations. We might carefully observe a number of boys and girls who are said to be intelligent or nonintelligent by teachers (though here we introduce contamination, because teachers must obviously judge intelligence, in part at least, by achievement). We may notice that a larger number of "more intelligent" children come from homes of higher socioeconomic levels; they solve problems in class more quickly than other children; they
have a wider vocabulary, and so on. We now have some clues to the nature of intelligence, so that we can attempt to construct a simple measure of intelligence. Note that our "definition" of intelligence springs from what presumably intelligent and nonintelligent children do. A similar procedure can be followed with the variable "achievement."

**Strengths and Weaknesses of Field Studies**

Field studies are strongly in realism, significance, strength of variables, theory orientation, and heuristic quality. The variance of many variables in actual field settings is large, especially when compared to the variance of the variables of laboratory experiments. Consider the contrast between the impact of social norms in a laboratory experiment like Sherif's and the impact of these norms in a community where, say, certain actions of teachers are frowned upon and others approved. Consider also the difference between studying cohesiveness in the laboratory where subjects are asked, for example, whether they would like to remain in a group (measure of cohesiveness) and studying the cohesiveness of a school faculty where remaining in the group is an essential part of one's professional future. Compare the group atmosphere in the Bennington College Study and that in a field experiment where different atmospheres are engendered by college instructors playing different roles. Variables such as social class, prejudice, conservatism, cohesiveness, and social climate can have strong effects in these studies. The strength of variables is not an unaided blessing, however. In a field situation there is usually so much noise in the channel that even though the effects may be strong and the variance great, it is not easy for the experimenter to separate the variables.

The realism of field studies is obvious. Of all types of studies, they are closest to real life. There can be no complaint of artificiality here. (The remarks about realism in field experiments apply, a fortiori, to the realism of field studies.)

Field studies are highly heuristic. One of the research difficulties of a field study is to keep the study contained within the limits of the problem. Hypotheses frequently fling themselves at one. The field is rich in discovery potential. For example, one may wish to test the hypothesis that the social attitudes of board of education members is a determinant of board of education policy decisions. Starting to gather data, however, many interesting notions that can deflect the course of the investigation can arise. To complicate matters, the field situation almost always has a plethora of variables and variance. Think of the many possible independent variables that we can choose as determinants of delinquency or of school achievement. In an experimental study, these variables can be controlled to a large extent, but in a field study they must somehow be controlled by more indirect and less satisfactory means.

Another methodological weakness is the lack of precision in the measurement of field variables. In field studies the problem of precision is more acute, naturally, than in field experiments. The difficulty encountered by Atin (and others) in measuring college environment is one of many similar examples. Administrative environment, for example,


18 Studies of organizations, for example, are mostly field studies, and the measurement of organizational variables well illustrates the difficulties. "Organizational Effectiveness" appears to be as complex as "Teacher Effectiveness." For a thorough and enlightening discussion, see: D. Katz and R. Kahn, The Social Psychology of Organizations, 2d ed. New York: Wiley, 1978, chap. 8, especially pp. 214-226. This superb book well repays careful reading and study.

19 For details, see Katz, op. cit., especially pp. 65ff.

Laboratory Experiments, Field Experiments, and Field Studies • 375

was measured by students' perceptions of aspects of the environment. Much of the lack of precision is due to the greater complexity of field situations. Other weaknesses of field studies are practical problems: feasibility, cost, sampling, and time. These difficulties are really potential weaknesses—none of them need be a real weakness. The most obvious questions that can be asked are: Can the study be done with the facilities at the investigator's disposal? Can the variables be measured? Will it cost too much? Will it take too much time and effort? Will the subjects be cooperative? Is random sampling possible? Anyone contemplating a field study has to ask and answer such questions. In designing research it is important not to underestimate the large amounts of time, energy, and skill necessary for the successful completion of most field studies. The field researcher needs to be a salesman, administrator, and entrepreneur, as well as investigator.
is correct? If you do not agree, say why you don’t. Before making snap judgments, read and ponder the references given in Study Suggestion 9, below.

9. Unfortunately, there has been much uninformed criticism of experiments. Before pronouncing rational judgments on any complex phenomenon we should first know what we’re talking about, and, second, we should know the nature and purpose of the phenomenon we criticize. To help you reach rational conclusions about the experiment and experimentation, the following references are offered as background reading.


**Kaplan, A.** *The Conduct of Inquiry.* San Francisco: Chandler, 1964, chap. IV. This chapter called “Experiment” seems to include most controlled observation.


**Fisher, R.** *The Design of Experiments,* 6th ed. New York: Hafner, 1951. A justly famous book on the statistical basis of inference from experiments. Chapter II is particularly enlightening, including as it does Fisher’s classic description of determining whether a lady can, by tasting a cup of tea (with milk), tell whether the milk or the tea was first added to the cup. Pay particular attention to Sections 9 and 10 on randomization.

---

**Chapter 24**

**Survey Research**

Survey research studies large and small populations (or universes) by selecting and studying samples chosen from the populations to discover the relative incidence, distribution, and interrelations of sociological and psychological variables. Survey research is often called *sample surveys,* probably because survey research developed as a separate research activity, along with the development and improvement of sampling procedures. Surveys, as such, are not new. Social welfare studies were done in England as long ago as the eighteenth century. Survey research in the social scientific sense, however, is quite new—it is a development of the twentieth century.

Survey research is considered to be a branch of social scientific research, which immediately distinguishes it from the status survey. Its procedures and methods have been developed mostly by psychologists, sociologists, economists, political scientists, and statisticians. These individuals have put a rigorous scientific stamp on survey research and, in the process, have profoundly influenced the social sciences.

The definition also links populations and samples. Survey researchers are interested in the accurate assessment of the characteristics of whole populations of people. They want to know, for example, how many persons in the United States vote for a Republican candidate and the relationship between such voting and variables like sex, race, religious

---

1 This chapter concentrates on the use of survey research in scientific research and neglects so-called status surveys, the aim of which is to learn the status quo rather than to study the relations among variables. There is no intention of derogating status surveys; they are useful, even indispensable. The intention is to emphasize the importance and usefulness of survey research in the scientific study of socially and educationally significant problems. The work of public opinion pollsters, such as Gallup and Roper, is also neglected. For a good account of polls and other surveys, see M. Parten, *Surveys, Polls, and Samples.* New York: Harper & Row, 1950, chap. 1. Though old, this book is still valuable. The current standard text is D. Warwick and C. Festinger, *The Sample Survey: Theory and Practice.* New York: McGraw-Hill, 1973. This book has the advantage of having been guided by the thinking and practice of the Survey Research Center, University of Michigan. It also has the advantage of having a cross-cultural emphasis.

preference, and the like. They want to know the relation between attitudes toward education and public support of school budgets.

Only rarely, however, do survey researchers study whole populations: they study samples drawn from populations. From these samples they infer the characteristics of the defined population or universe. The study of samples from which inferences about populations can be drawn is needed because of the difficulties of studying whole populations. Random samples can often furnish the same information as a census (an enumeration and study of an entire population) at much less cost, with greater efficiency, and sometimes greater accuracy.

Sample surveys attempt to determine the incidence, distribution, and interrelations among sociological and psychological variables, and, in so doing, usually focus on people, the vital facts of people, and their beliefs, opinions, attitudes, motivations, and behavior. The social scientific nature of survey research is revealed by the nature of its variables, which can be classified as sociological facts and opinions and attitudes. Sociological facts are attributes of individuals that spring from their membership in social groups: sex, income, political and religious affiliation, socioeconomic status, education, age, living expenses, occupation, race, and so on.3

The second type of variable is psychological and includes opinions and attitudes, on the one hand, and behavior, on the other. Survey researchers are interested not only in relations among sociological variables; they are more likely to be interested in what people think and do and the relations between sociological and psychological variables. The study of the quality of American life done by the Survey Research Center of the University of Michigan, for instance, reports depressing data on the relation between race and feelings of trust in people, a sociological and a psychological variable.4 Data are given in Table 24.1. The relation is substantial. Evidently black people feel less trusting of people than whites. As Campbell et al. say (p. 455), "those people who have been least successful in their encounters with society have the least reason to feel trustful of it."

Survey researchers, of course, also study the relations among psychological variables (see, for example, Table 10.3, Chapter 10). But most relations of survey research are those between sociological and psychological variables: between education and tolerance (see Table 10.9), between race and self-esteem (Table 10.11), and between education and sense of political efficacy (Study Suggestions, Chapter 10), for example.

**TYPES OF SURVEYS**

Surveys can be conveniently classified by the following methods of obtaining information: personal interview, mail questionnaire, panel, and telephone. Of these, the personal interview far overshadows the others as perhaps the most powerful and useful tool of social scientific survey research. These survey types will be briefly described here; in later chapters, when studying methods of data collection, we will study the personal interview in depth.

**Interviews and Schedules**

The best survey research uses the personal interview as the principal method of gathering information. This is accomplished in part by the careful and laborious construction of a schedule or questionnaire.5 Schedule information includes factual information, opinions and attitudes, and reasons for behavior, opinions, and attitudes. Interview schedules are difficult to construct; they are time-consuming and relatively costly; but there is no other method that yields the information they contain.

The factual information gathered in surveys includes the so-called sociological data mentioned previously: sex, marital status, education, income, political preference, religious preference, and the like. Such information is indispensable, since it is used in studying the relations among variables and in checking the adequacy of samples. These data, which are entered on a "face sheet," are called "face sheet information." Face sheet information, at least part of it, is ordinarily obtained at the beginning of the interview.

Much of it is neutral in character and helps the interviewer establish rapport with the respondent. Questions of a more personal nature, such as those about income and personal habits, and questions that are more difficult to answer such as the extent of the knowledge or ability of the respondent, can be reserved for later questioning, perhaps at the end of the schedule. The timing must necessarily be a matter of judgment and experience.6

Other kinds of factual information include what respondents know about the subject under investigation, what respondents did in the past, what they are doing now, and what they intend to do in the future. After all, unless we observe behavior directly, all data about respondents' behavior must come from others or from other people. In this special sense, past, present, and future behavior can all be classified under the "facts" of behavior, even if the behavior is only an intention. A major point of such factual questions is that the respondent presumably knows a good deal about his own actions and behavior. If he says he voted for a school bond issue, we can believe him—unless there is compelling evidence to the contrary. Similarly, we can believe him, perhaps with more reservation (since the event has already happened), if he says he is going to vote for a school bond issue.

Just as important, maybe even more important from a social scientific standpoint, are the beliefs, opinions, attitudes, and feelings that respondents have about cognitive objects.7 Many of the cognitive objects of survey research may not be of interest to the researcher: investments, certain commercial products, political candidates, and the like. Other cognitive objects are more interesting: the United Nations, the Supreme Court, educational practices, integration, Federal aid to education, college students, Jews, and the women's liberation movement.

The personal interview can be helpful in learning respondents' reasons for doing or believing something. When asked reasons for actions, intentions, or beliefs, people may say they have done something, intend to do something, or feel certain ways about some-

---

3For a complete description of such personal and social facts, see Parkin, op. cit., pp. 169-174.

4A. Campbell, P. Converse, and W. Rodgers, *The Quality of American Life. New York: Russell Sage Foundation*, 1976. I calculated the percentages of Table 24.1 from the reported percentages and frequencies of Campbell et al.'s Table 13-11, p. 455, to show the relation clearly.

5The term "schedule" will be used. It has a clear meaning: the instrument used to gather survey information through personal interview. "Questionnaire" has been used to label personal interview instruments and attitudinal or personality instruments. The latter are called "scales" in this book.

6See Warwick and Lintner, op. cit., pp. 150-151.

7Cognitive object is an expression used to indicate the object of an attitude. Almost anything can be the object of an attitude, but the term is ordinarily reserved for important social objects: for example, groups (religious, racial, educational) and institutions (education, marriage, political parties). A more general and probably better term, though one not in general use, is referent.
thing. They may say that group affiliations or loyalties or certain events have influenced them. Or they may have heard about issues under investigation via public media of communication. For example, a respondent may say that, while he was originally opposed to Federal aid to education because he and his political party have always opposed government interference, he now supports Federal aid because he has read a great deal about the problem in newspapers and magazines and has come to the conclusion that Federal aid will benefit American education.

A respondent’s desires, values, and needs may influence his attitudes and actions. When saying how he favors Federal aid to education the respondent may indicate that his own educational aspirations were thwarted and that he has always yearned for more education. Or he may indicate that his religious group has, as a part of its value structure, a deep commitment to the education of children. If the individual under study has accurately sounded his own desires, values, and needs—and can express them verbally—the personal interview can be very valuable.

**Other Types of Survey Research**

The next important type of survey research is the panel. A sample of respondents is selected and interviewed, and then reinterviewed and studied at later times. The panel technique enables the researcher to study changes in behaviors and attitudes.

Telephone surveys have little to recommend them beyond speed and low cost. Especially when the interviewer is unknown to the respondent they are limited by possible nonresponse, uncooperativeness, and by reluctance to answer more than simple, superficial questions. Yet telephoning can sometimes be useful in obtaining information essential to a study. Its principal defect, obviously, is the inability to obtain detailed information.

The mail questionnaire, another type of survey, has serious drawbacks unless it is used in conjunction with other techniques. Two of these defects are possible lack of response and the inability to check the responses given. These defects, especially the first, are serious enough to make the mail questionnaire worse than useless, except in highly sophisticated hands. Responses to mail questionnaires are generally poor. Returns of less than 40 or 50 percent are common. Higher percentages are rare. At best, the researcher must content himself with returns as low as 50 or 60 percent.

As a result of low returns in mail questionnaires, valid generalizations cannot be made. Although there are means of securing larger returns and reducing deficiencies—follow-up questionnaires, enclosing money, interviewing a random sample of nonrespondents and analyzing nonresponder data—these methods are costly, time-consuming, and often ineffective. As Parten says, “Most mail questionnaires bring so few returns, and these from such a highly selected population, that the findings of such surveys are almost invariably open to question.”

The best advice would seem to be not to use mail questionnaires if a better method can possibly be used. If they are used, every effort should be made to obtain returns of at least 80 to 90 percent or more, and lacking such returns, to learn something of the characteristics of the nonrespondents.

---

**THE METHODOLOGY OF SURVEY RESEARCH**

Survey research has contributed much to the methodology of the social sciences. Its most important contributions, perhaps, have been to rigorous sampling procedures, the overall design and the implementation of the design of studies, the unambiguous definition and specification of the research problem, and the analysis and interpretation of data.

In the limited space of a section of one chapter, it is obviously impossible to discuss adequately the methodology of survey research. Only those parts of the methodology germane to the purposes of this book, therefore, will be outlined: the survey or study design, the so-called flow plan or chart of survey researchers, and the check of the reliability and validity of the sample and the data-gathering methods. (Sampling was discussed in Part Three, analysis in Part Four.)

Survey researchers use a "flow plan" or chart to outline the design and subsequent implementation of a survey. The flow plan starts with the objectives of the survey, lists each step to be taken, and ends with the final report. First, the general and specific problems that are to be solved are carefully and as completely as possible. Since, in principle, there is nothing very different here from the discussion for problems and hypotheses of Chapter 2, we can omit detailed discussion and give one simple hypothetical example. An educational investigator has been commissioned by a board of education to study the attitudes of community members toward the school system. On discussing the general problem with the board of education and the administrators of the school system, the investigator notes a number of more specific problems such as: Is the attitude of the members of the community affected by their having children in school? Are their attitudes affected by their educational level?

One of the investigator's most important jobs is to specify and clarify the problem. To do this well, he should not expect just to ask people what they think of the schools, although this may be a good way to begin if one does not know much about the subject. He should also have specific questions to ask that are aimed at various facets of the problem. Each of these questions should be built into the interview schedule. Some survey researchers even design tables for the analysis of the data at this point in order to clarify the research problem and to guide the construction of interview questions. Since this procedure is recommended, let us design a table to show how it can be used to specify survey objectives and questions.

Take the question: Is attitude related to educational level? The question requires that "attitude" and "educational level" be operationally defined. Positive and negative attitudes will be inferred from responses to schedule questions and items: If, in response to a broad question like, "In general, what do you think of the school system here?" a respondent says, "It is one of the best in this area," it can be inferred that he has a positive attitude toward the schools. Naturally, one question will not be enough. Related questions should be used, too. A definition of "educational level" is quite easy to obtain. It is decided to use three levels: (1) Some College, (2) High School Graduate, and (3) Non-High School Graduate. The analysis paradigm might look like Figure 24.1.

<table>
<thead>
<tr>
<th>Positive Attitude</th>
<th>Negative Attitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some College</td>
<td></td>
</tr>
<tr>
<td>High School Graduate</td>
<td></td>
</tr>
<tr>
<td>Non-High School Graduate</td>
<td></td>
</tr>
</tbody>
</table>

---

8See Warwick and Linneringer, op. cit., pp. 62ff.
9See Parten, op. cit., pp. 391-402, for a discussion of the inadequacies of mail questionnaires and remedies for remediable deficiencies. Also, Warwick and Linneringer, op. cit., pp. 131-132.
10Parten, op. cit., p. 400.

---

The virtue of paradigms like this is that the researcher can immediately tell whether he has stated a specific problem clearly and whether the specific problem is related to the general problem. It also gives him some notion as to how many respondents he will need to fill the table cells adequately, as well as provide him guidelines for coding and analysis. In addition, as Katz says, “By actually going through the mechanics of setting out such tables, the investigators are bound to discover complexities of a variable that need more detailed measurement and qualifications of hypotheses in relation to special conditions.”

The next step in the flow plan is the sample and the sampling plan. Because sampling is much too complex to be discussed here,12 we outline only the main ideas. First, the universe to be sampled and studied must be defined. Is all citizens living in the community included? Those citizens paying school taxes? Those with children of school age? Once the universe is defined, a decision is made as to how the sample is to be drawn and how many cases will be drawn. In the best survey research, random samples are used. Because of their high cost and greater difficulty of execution, random samples are often bypassed for quota samples. In a quota (or quota control) sample, “representativeness” is presumably achieved by assigning quotas to interviewers—so many men and women, so many whites and blacks, and so on. Quota sampling should be avoided: while it may achieve representativeness, it lacks the virtues of random sampling.

The next large step in a survey is the construction of the interview schedule and other measuring instruments to be used. This is a laborious and difficult business bearing virtually no resemblance to the questionnaires often hastily put together by neophytes. The main task is to translate the research question into an interview instrument and into any other instruments constructed for the survey. One of the problems of the study, in fact, may be: How are permissive and restrictive attitudes toward the discipline of children related to perceptions of the local school system? Among the questions to be written are permissive and restrictive attitudinal ones: How do you feel children should be disciplined? After drafts of the interview schedule and other instruments are completed, they are pretested on a small representative sample of the universe. They are then revised and put in final form.

The steps outlined above constitute the first large part of any survey. Data collection is the second large part. Interviewers are oriented, trained, and sent out with complete instructions as to whom to interview and how the interview is to be handled. In the best surveys, interviewers are allowed no latitude as to whom to interview. They must interview those individuals and only those individuals designated, generally by random devices. Some latitude may be allowed in the actual interviewing and use of the schedule, but not much. The work of interviewers is also systematically checked in some manner. For example, every tenth interview may be checked by sending another interviewer to the same respondent. Interview schedules are also studied for signs of spurious answering and reporting.

The third large part of the flow plan is analytical. The responses to questions are coded and tabulated. Coding is the term used to describe the translation of question responses and respondent information to specific categories for purposes of analysis.13 Take the example of Figure 24.1. All respondents must be assigned to one of the three educational-level categories and a number (or other symbol) assigned to each level. Then each person must also be assigned to a “positive attitude” or “negative attitude” category. To aid in the coding, content analysis may be used. Content analysis is an objective and quantitative method for assigning types of verbal and other data to categories (see Chapter 30, below). Coding can mean the analysis of factual response data and then the assignment of individuals to classes or categories, or the assigning of categories to individuals, especially if one is preparing cards for computer analysis. Such cards consist of a large number of columns with a number of cells in each column. The fifth column may be assigned, say, to sex, and the first two cells of the column, or the numbers 0 and 1, used to designate female and male.

Tabulation is the recording of the numbers of types of responses in the appropriate categories, after which statistical analysis follows: percentages, averages, relational indices, and appropriate tests of significance. The analyses of the data are studied, collated, assimilated, and interpreted. Finally, the results of this interpretative process are reported.

Checking Survey Data

Survey research has a unique advantage among social scientific methods: it is often possible to check the validity of survey data. Some of the respondents can be interviewed again, the results of both interviews checked against each other. It has been found that the reliability of personal factual items, like age and income, is high.14 The reliability of attitude responses is harder to determine because a changed response can mean a changed attitude. The reliability of average responses is higher than the reliability of individual responses. Fortunately, the researcher is usually more interested in averages, or group measures, than in individual responses.

One way of checking the validity of measuring instruments is to use an outside criterion. One compares one’s results to some outside, presumably valid, criterion. For instance, a respondent tells us he voted in the last election of school board members. We can find out whether he did or not by checking the registration and voting records. Ordinarily, individual behavior is not checked, because information about individuals is hard to obtain, but group information is often available. This information can be used to test to some extent the validity of the survey sample and the responses of the respondents.

A good example of an outside check on survey data is the use of information from the last census. This is particularly useful in large-scale surveys, but it may also help in smaller ones. Proportions of men and women, races, educational levels, age, and so on in the sample and in the U.S. census are compared. In the Verba and Nie study of political participation, for example, the authors report a number of such comparisons.15 Table 24.2 reports some of them. It is obvious that the sample estimates are accurate: only one of them, Age 20-34, deviates from the Census estimates by more than 2 percent, which is

12Simple coding is discussed in Warner and Lininger, op. cit., chap. 9. For detailed discussions of coding and coding problems, instructional materials are available from the Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48104. The Institute also issues bibliographies on survey research and related matters.

13Simple coding is discussed in Warner and Lininger, op. cit., chap. 9. For detailed discussions of coding and coding problems, instructional materials are available from the Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48104. The Institute also issues bibliographies on survey research and related matters.


reassuring evidence of the adequacy of the sample. To be sure, the sample was large (>2500), but smaller samples have also been found to be quite accurate.17

TWO STUDIES

Many surveys have been done, both good and bad. Most of them would probably not interest the student because they are little more than refined attempts to obtain simple information: studies of presidential voting, of industrial plants, and so on. There are, however, surveys of considerable, even great, interest and significance to behavioral scientists. Two of these are summarized below. They were chosen for their intrinsic interest and because they studied problems that are theoretically and socially important. They also seem to represent a healthy trend toward using research study to study and understand complex psychological and sociological phenomena. One of them was a large national study (more than 2,500 interviews) and the other a much smaller study done in two suburban districts of Los Angeles (two samples of about 200 each). The latter study also included the "rare luxury of a complete replication."

Verba and Nie: Political Participation in America18

Recall that Verba and Nie asked, among other things, how the political participation of citizens of a democracy influences governmental processes. They interviewed more than 2,500 residents of the United States in 200 locales in 1967, selected by an area probability sampling procedure. (Their census-sample comparisons showed generally high agreement.19) The main finding was mentioned in Chapter 22: citizen participation does indeed influence political leaders, but it is the more affluent, better-educated, and generally

Higher status citizens whose participation is influential. The authors point out that although Americans are not noted for class-based ideology, social status does relate to participation.20 The study was especially characterized by sophisticated measurement and analytic methodology, and by a major disconcerting finding. We will return to it in its methodology in later chapters.

Kinder and Sears: Prejudice and Politics21

This study explicitly stated and tested alternative theories of racial prejudice. Indeed, the authors deplored the lack of empirical confrontation between alternative theories in the prejudice literature (p. 414). The study is also unusual because its circumstances "permitted . . . the rare luxury of a complete replication" (albeit with small samples).

The theories of prejudice tested were called symbolic racism and racial threat. Symbolic racism, say the authors, is a blend of antiblack sentiment and traditional moral values of the Protestant Ethic. It is resistance to change in the racial status quo based on moral feelings that blacks violate traditional American values of self-reliance, individualism, the work ethic, and discipline. This syndrome of determinants of racial prejudice is the descendant of an older social-cultural learning theory, which emphasized that prejudice was learned by children along with other normative values and attitudes. But, say Kinder and Sears, white America has become, at least in principle, racially egalitarian, and another explanation is therefore necessary. This explanation they call symbolic racism.

The alternative explanation, group conflict theory, emphasizes the threats that blacks pose to the private lives of whites: moving into white neighborhoods, moving into better jobs displacing whites, insistence on integrated schools and enforced busing and racial mixing of young children, and the threat of rising violence perceived as due to black criminality. The implications of both theories were used to construct measures of personal racial threats and symbolic racism, which were related to the dependent variable, intended votes for two candidates for mayor of Los Angeles in 1969 and 1973, Yorty, a white conservative, and Bradley, a black liberal.

White residents of two suburban communities in Los Angeles were selected in what appears to have been quota sampling. Two samples of adults were interviewed in 1969 and in 1973. The questions asked focused mainly on issues relating to personal threat and symbolic racism. The researchers' analysis indicated that symbolic racism was an important form of racial prejudice than racial threat. That is, symbolic racism accounted for more of the variance of candidate preference than racial threat. Various control analyses further supported this finding. The authors concluded that "racial attitudes were major determinants of voting in both mayoral elections" (ibid., p. 427), and racial threats to whites' lives were largely irrelevant. These findings and conclusions are highly important both theoretically and practically.

These two studies clearly show a trend toward using survey research as a tool to test theory and hypotheses in contrast to older use in which the emphasis was on finding "what is there." We can expect this trend to continue and to grow, especially in sociology, political science, and education.

17In one study of Detroit done by the University of Michigan, the sample was only 735, but the sample estimates were close to those of the Census of the 1950 census. Detroit Area Study, University of Michigan. A Social Profile of Detroit. Ann Arbor: University of Michigan, 1952, p. 36. Campbell and Katona, op. cit., pp. 41-48, discuss methods of checking sample validity and reliability. Warwick and Linnenger, op. cit., pp. 311-314, give tables of sampling errors, with an explanation of their statistical meaning and use. We learn, for instance, that reported percentages between 20 and 80 from a sample of 700 have a standard error of 4. To reduce this to 2 requires a sample of 3,000!
18Verba and Nie. op. cit. The survey was done by the National Opinion Research Center.
19Ibid., p. 349, Table A-1.
20Ibid., p. 339.
APPLICATIONS OF SURVEY RESEARCH TO EDUCATION

The Verba and Nie and the Kinder and Sears studies clearly show the applicability of survey research and its methodology in sociology, social psychology, and political science. Survey research's strong emphasis on representativeness, overall design and plan of research, and expert interviewing using carefully and competently constructed interview schedules have had and will continue to have beneficial influence on behavioral research. Despite its evident potential value in all behavioral research fields, survey research has not been used to any great extent where it would seem to have large theoretical and practical value: in education. Its distinctive usefulness in education and educational research seems not to have been realized. This section is therefore devoted to application of survey research in educational and educational problems.

Obviously, survey research is a useful tool for educational fact-finding. An administrator, a board of education, or a staff of teachers can learn a great deal about a school system or a community without contacting every child, every teacher, and every citizen. In short, the sampling methods developed in survey research can be very useful. It is unsatisfactory to depend upon relatively hit-or-miss, so-called representative samples based on "expert" judgments. Nor is it necessary to gather data on whole populations; samples are sufficient for many purposes.

Most research in education is done with relatively small nonrandom samples. If hypotheses are supported, they can later be tested with random samples of populations, and if again supported, the results can be generalized to populations of schools, children, and laymen. In other words, survey research can be used to test hypotheses already tested in more limited situations, with the result that external validity is increased.

Survey research seems ideally suited to some of the large controversial issues of education. For example, its ability to handle "difficult" problems like integration and school closings through careful and circumspect interviewing puts it high on the list of research approaches to such problems. Researchers and educators can study the impact of integration and of school closings on communities and their school systems. Interviews of random samples of citizens and teachers of school districts just starting integration or the closing of certain elementary schools because of declining enrollment can provide valuable information on the concerns and fears of citizens so that appropriate measures to inform them and lessen their fears can be taken. The effect of these measures can be studied.

Survey research is probably best adapted to obtaining personal and social facts, beliefs, and attitudes. It is significant that, although hundreds of thousands of words are spoken and written about education and about what people presumably think about education, there is little dependable information on the subject. We simply do not know what people's attitudes toward education are. We have to depend on feature writers and so-called experts for this information. Boards of education frequently depend on administrators and local leaders to tell them what the people think. Will they support an expanded budget next year? What will they think about a merger of adjoining school districts? How will they react to busing white and black children to achieve desegregation?

ADVANTAGES AND DISADVANTAGES OF SURVEY RESEARCH

Survey research has the advantage of wide scope: a great deal of information can be obtained from a large population. A large population or a large school system can be studied with much less expense than that incurred by a census. While surveys tend to be more expensive than laboratory and field experiments and field studies, for the amount and quality of information they yield they are economical. Furthermore, existing educational facilities and personnel can be used to reduce the costs of the research.

Survey research information is accurate—within sampling error, of course. The accuracy of properly drawn samples is frequently surprising, even to experts in the field. A sample of 600 to 700 individuals or families can give a remarkably accurate portrait of a community—its values, attitudes, and beliefs.

With these advantages go inevitable disadvantages. First, survey information ordinarily does not penetrate very deeply below the surface. The scope of the information sought is usually emphasized at the expense of depth. This seems to be a weakness, however, that is not necessarily inherent in the method. The Verba and Nie and other studies show that it is possible to go considerably below surface opinions. Yet the survey seems best adapted to extensive rather than intensive research. Other types of research are perhaps better adapted to deeper exploration of relations.

A second weakness is a practical one. Survey research is demanding of time and money. In a large survey, it may be months before a single hypothesis can be tested. Sampling and the development of good schedules are major operations. Interviews require skill, time, and money. Surveys on a smaller scale can avoid these problems to some extent, even though it is generally true that survey research demands large investments of time, energy, and money.

Any research that uses sampling is naturally subject to sampling error. While it is true that survey information has been found to be relatively accurate, there is always the one chance in twenty or a hundred that an error more serious than might be caused by minor fluctuations of chance may occur. The probability of such an error can be diminished by building safety checks into a study—by comparing census data or other outside information and by sampling the same population independently.

A potential rather than an actual weakness of this method is that the survey interview can temporarily lift the respondent out of his own social context, which may make the results of the survey invalid. The interview is a special event in the ordinary life of the respondent. This apartness may affect the respondent so that he talks to, and interacts with, the interviewer in an unnatural manner. For example, a mother, when queried about her child-rearing practices, may give answers that reveal methods that she would like to use rather than those she does use. It is possible for interviewers to limit the effects of lifting respondents out of social context by skilled handling, especially by one's manner and by careful phrasing and asking of questions.

Survey research also requires a good deal of research knowledge and sophistication. The competent survey investigator must know sampling, question and schedule construction, interviewing, the analysis of data, and other technical aspects of the survey. Such knowledge is hard to come by. Few investigators get this kind and amount of experience. As the value of survey research, both large- and small-scale, becomes appreciated, it can be anticipated that such knowledge and experience will be considered, at least in a minimal way, to be necessary for researchers.


Study Suggestions

1. Here are several good examples of survey research. Choose one of them and read the first chapter (or chapters) to learn the problem of the study. Then go to the technical methodological appendix to see how the sampling and interviewing were done. (Most published survey research studies have such appendices.) Try to determine the main variables and their relations.

Studies cited in footnotes 4, 16, 17, 21, 22.


2. This study suggestion is for students of education. Read as much of the Gross, Mason, and McEachern survey research study (footnote 22) as you can. It is a methodological model for large-scale educational research. And, of course, it reports a number of interesting findings about superintendents and boards of education and their views. Perhaps most important, it shows that a scientific approach and practical concern for educational practice can be combined.

3. Rensis Likert, an outstanding social scientist, a methodological pioneer of survey research, and the founder of the Institute for Social Research of the University of Michigan (of which the Survey Research Center is a part), recently died. Two of his colleagues wrote an obituary in which they described Likert's contributions. It is suggested that students read the obituary, which is virtually an account of the birth and growth of important methodological aspects of survey research, as well as an interesting description of the contributions of this creative and competent individual.

Appendix D

The Research Report

THIS APPENDIX has two purposes: to outline some of the main points of report writing and to cite appropriate references to guide the reader.

THE PURPOSE

The purpose of the research report is to tell readers the problem investigated, the methods used to solve the problem, the results of the investigation, and the conclusions inferred from the results. It is not the function of the investigator to convince the reader of the virtue of the research. Rather, it is to report, as expeditiously and clearly as possible, what was done, why it was done, the outcome of the doing, and the investigator's conclusions. The report should be so written that readers can reach their own conclusions as to the adequacy of the research and the validity of the reported results and conclusions.

To achieve this purpose is not easy. The writer must strive for the right blend of detail and brevity, for objectivity, and for clarity in presentation. Perhaps the best criterion question is: Can another investigator replicate the research by following the research report? If not, due to incomplete or inadequate reporting of methodology or to lack of clarity in presentation, then the report is inadequate.1

1The realities of publishing and its costs limit the above statement. Book publishers and journal editors do not have the space available to make it possible to publish enough details of research studies so that they can be replicated. Indeed, the constraints on editors are such that they can hardly publish sufficient details of research for readers to make informed judgments of the methodological adequacy of the studies. Nevertheless, the criterion question should always be kept in mind.
THE STRUCTURE

The structure of the research report is simple. It is almost the same as the structure of the research itself: the problem, the methodology, the results. Here is a general outline:

I. Problem
   1. Theory, hypotheses, definitions
   2. Previous research; the literature
II. Methodology-Data Collection
   1. Sample and sampling method
   2. How hypotheses were tested (methodology), experimental procedures, instrumentation
   3. Measurement of variables
   4. Methods of analysis, statistics
   5. Pretesting and pilot studies
III. Results, Interpretation, and Conclusions

THE PROBLEM

The problem section differs greatly in different reports. In theses and books, it is usually long and detailed. In published research reports, it is kept to a minimum (see footnote 1). The basic precept, though seemingly obvious, is not easy to follow: Tell readers what the problem research is. Tell it to them in question form. For example, What effect does egalitarianism have on the mental status of schoolchildren? Does past experience with materials have a negative effect on problem-solving involving the materials? How do social attitudes influence judgments of the effectiveness of social policies? The statement of the general problem is usually not precise and operational. Rather, it sets the general stage for the reader. The subproblems, however, should be more precise. They should have implications for testing. For example: Can a person conversing with others manipulate conversation by agreeing or disagreeing with the others, or by paraphrasing what they have said? The Jones and Cook statement given in the preceding paragraph is made more operational by specifying the social attitudes and the judgments of social policies affecting blacks. Do attitudes toward blacks affect recommendations of social policies for improving black welfare? Do individuals with positive attitudes toward blacks recommend societal change, and do individuals with negative attitudes recommend that blacks improve themselves?

Some research writers, rather than state the problems, state the general and specific hypotheses. A good practice would seem to be to state the broader general problem and then to state the hypotheses, both general and specific. The reader is referred to Chapter 2 for examples. Whatever way is used, bear in mind the main purpose of informing the reader of the main area of investigation and the specific propositions that were tested.

At some point in the problem discussion the variables should be defined, or at least mentioned or generally characterized, with more specific definitions given later. Variable definition was discussed at length in Chapter 3 and need not be repeated here, except for the admonition: Inform the reader not only of the variables but also what you mean by them. Define in general and operational terms, giving justification for your definitions.

There are two main reasons for discussing the general and research literature related to the research problem. The first of these is the more important: to explain and clarify the theoretical rationale of the problem. Suppose, like Haslerud and Meyers, one was interested in investigating the relative effectiveness of transfer of self-discovery of principles by lectures and systematic enunciation of the principles to learners. Since the problem is in part a transfer of training problem, one would have to discuss transfer and some of the literature on transfer, but especially that part of the literature pertinent to this problem. One may well want to discuss to some extent philosophical and pedagogical writings on the theory of formal discipline, for instance. In this manner the investigator provides a general picture of the research topic and fits his problem into the general picture.

A second reason for discussing the literature is to tell the reader what research has and has not been done on the problem. Obviously, the investigator must show that his particular investigation has not been done before. The underlying purpose, of course, is to locate the present research in the existing body of research on the subject and to point out what it contributes to the subject.

Methodology-Data Collection

The function of the methodology-data collection section of the research report, of course, is to tell the reader what was done to solve the problem. Meticulous care must be exercised to report that the criterion of replicability is satisfied. That is, it should be possible for another investigator to reproduce the research, to reanalyze the data, or to arrive at unambiguous conclusions as to the adequacy of the methods and data collection. In books and theses there can be little question of the applicability of the criterion. In research journal reports, unfortunately, the criterion is difficult, sometimes even impossible, to satisfy. Owing to lack of journal space, investigators are forced to condense reports in such a way that it is sometimes difficult to reconstruct and test for the method that has been used. (See footnote 1.) Yet the criterion remains a good one and should be kept in mind when tackling the methodology section.

The first part of the methodology-data collection section should tell what sample or samples were used, how they were selected, and why they were selected. If eighth-grade pupils were used, the reason for using them should be stated. If the samples were randomly selected, this should be said. The method of random sampling should also be specified. If pupils were assigned at random to experimental groups, this should be reported: if they were not, this, too, should be reported with reasons for the lack of such assignment.

The method of testing the hypotheses should be reported in detail. If the study has been experimental, the manner in which the independent variable(s) has been manipulated is described. This description includes instruments used, instructions to the subjects, control procedures, and the like. If the study has been nonexperimental, the procedures used to gather data are outlined.

The report of any empirical study must include an account of the measurement of the variables of the study. This may be accomplished in few sentences in some studies. For example, in an experiment with one independent variable and a dependent variable whose measurement is simple, all that may be necessary is a brief description of the measurement of the dependent variable. Such measurement may entail only the counting of responses. In other studies, the description of the measurement of the variables may take up most of the methodology section. A factor analytic study, for instance, may require lengthy descriptions of measurement instruments and how they were used. Such descriptions will, of course, include justification of the instruments used, as well as evidence of their reliability and validity.

An account of the data analysis methods used is sometimes put into the methodology.

---

Footnotes:

problem or the methodology, should be pointed out. This is done, not to show humility or one's technical competence, but rather to enable the reader to judge the validity of the conclusions drawn from the data and the general worth of the study.

Limitations of social scientific and educational research generally come from sampling and subject assignment inadequacies, methodological weaknesses, and statistical deficiencies. Inadequacies of random sampling and subject assignment are often seen as the particular sample used. Lack of random assignment casts doubt on the adequacy of the control of independent variables and thus on the conclusions. Statistical deficiencies, similarly, can lead to incorrect conclusions. Deficiencies in measurement always affect conclusions, too. If a measurement instrument, perhaps through no fault of the writer, is only moderately reliable, a finding may be ambiguous and inconclusive. More important, the questionable validity of an instrument may seriously affect conclusions.

These matters have been discussed in the text and need no further elaboration here. It may be added, however, that the writing of the conclusions is naturally affected by the recognized and acknowledged limitations and weaknesses. Moreover, readers can hardly be expected to judge the validity of research conclusions without knowing both the positive and the negative aspects of whatever was done. It is the professional responsibility of the researcher, then, to inform readers of both the strengths and the weaknesses of the research.

THE WRITING

It is not easy to write simply and clearly. One has to work at it. One should realize that almost no writer can escape the necessity of constant revision by rereading and paring—deleting circumlocutions, redundancies, and other verbal fat. Suggestions for better research report writing follow.

Although research reports should be fairly detailed, there is no need to waste words. State the problem, the methodology, and the results as clearly, simply, and briefly as possible. Avoid hackneyed expressions like "in terms of," "with respect to," "in reference to," "give consideration to," and the like. Delete unnecessary words and expressions when revising. For example, sentences with expressions like "the fact of the matter is" and "owing to the fact that," and "as to whether" can always be revised to remove such clumsy inelegancies. For good advice on simplicity and clarity, study Strunk and White's little classic, *The Elements of Style*.

Writing scholarly papers and research reports requires a certain amount of routine drudgery that few of us like. Bibliographies, footnotes, tables, figures, and other mechanical details, however, cannot be escaped. Yet a little systematic thought can help solve most problems. That is, do not wait until you sit down to write and then find out how to handle footnotes and other mechanical details. Get a good reference book or two and study and lay out footnote and bibliographical forms, tables, figures, and headings. Put three or four types of footnote entries on 3-by-5 cards. Similarly, learn two or three methods of laying out tables. Lay out skeleton tables. Then use these samples when writing. In short, put much of the drudgery and doubt behind you by mastering the elements of the methods, instead of impeding your writing by constant interruptions to check on how to do things.

Presentation of statistical results and analyses gives students considerable trouble. Fit the problem head-on. Perhaps the best way to do this is to study statistical presentation in two or three good journals, like the *Journal of Personality and Social Psychology*, the *Journal of Educational Psychology*, the *American Educational Research Journal*, and the *American Psychologist*. The style manual of the American Psychological Association (see References) has been adopted by all psychology journals and a number of education journals. Although a bit fuzzy, it is an excellent guide, especially to statistical and tabular presentation. Turabian's manual is another good guide.

The purpose of statistical, tabular, and other condensed presentation should be kept in
mind. A statistical table, for instance, should clearly tell the reader what the data essentially say. This does not mean, of course, that a statistical table can stand by itself. Its purpose is to illuminate and clarify the textual discussion. The text carries the story; the table helps make the text clear and gives the statistical evidence for assertions made in the text. The text may say, for example, "The three experimental groups differed significantly in achievement," and the tables will report the statistical data—means, standard deviations, F ratios, levels of significance—to support the assertion. There is often no need for a table. If a hypothesis has been tested by calculating one, two, or three coefficients of correlation, these can simply be reported in the text without tabular presentation.

A fairly safe generalization to guide one in writing research reports is: first drafts are not adequate. In other words, almost any writing, as said earlier, improves upon revision. It is almost always possible to simplify first-draft language and to delete unnecessary words, phrases, and even sentences and paragraphs. A first rule, then, is to go over any report with a ruthless pencil toward the end of greater simplicity, clarity, and brevity. With experience this not only becomes possible; it becomes easier.

If an adequate outline has been used, there should be little problem with the organization of a research paper. Yet sometimes it is necessary to reorganize a report. One may find, for example, that one has discussed something at the end of the report that was not anticipated in the beginning. Reorganization is required. In any case, the possibility of improvement in communication through reorganization should always be kept in mind.

Anyone’s research writing can be improved in two ways: by letting something one has written sit for a few weeks, and by having someone else read and criticize one’s work. It is remarkable what a little time will do for one’s objectivity and critical capacity. One sees obvious blemishes that somehow one could not see before. Time helps save the ego, too. Our precious inventions do not seem so precious after a few weeks or months. We can be much more objective about them.

The second problem is harder. It is hard to take criticism, but the researcher must learn to take it. Scientific research is one of the most complex of human activities. Writing research reports is not easy, and no one can be expected to be perfect. It should be accepted and routine procedure, therefore, to have colleagues read our reports. It should be accepted routine, too, to accept our readers’ criticisms in the spirit in which we should have asked for them. There is of course no obligation to change a manuscript in line with criticism. But there is an obligation to give each criticism the serious, careful, and objective attention it deserves. Doctoral students have to consider seriously the criticisms of their sponsors—whether or not they like them or agree with them. All scholarly and scientific writers, however, should voluntarily learn the discipline of subjecting their work to their peers. They should learn that the complex business of communicating scholarly and scientific work is difficult and demanding, and that in the long run they can only profit from competent criticism and careful revision.

SOME USEFUL REFERENCES


