Robert C. Calfee

Experimental Methods in Psychology

EXPERIMENTAL METHODS IN PSYCHOLOGY

ROBERT C. CALFEE
Stanford University

HOLT, RINEHART AND WINSTON
New York Chicago San Francisco Philadelphia
Montreal Toronto London Sydney
Tokyo Mexico City Rio de Janeiro Madrid

Acknowledgment for permission to reproduce excerpts from previously published material is made to:

American Psychological Association. From Biederman et al. (1973) and Collingwood et al. (1978), papers published in the *Journal of Experimental Psychology* and the *Journal of Educational Psychology*, respectively. From Ethical principles in the conduct of research with human participants (1973). Copyright © 1973, 1978 by American Psychological Association, Washington, D.C.

Psychonomic Society Inc., Austin, Texas. From Klatsky et al. (1982), paper published in *Memory & Cognition*.

V. H. Winston & Sons, Silver Spring, Maryland. From Kulka and Kessler (1978), paper published in *Applied Social Psychology*.

Library of Congress Cataloging in Publication Data

Calfee, Robert C.
Experimental methods in psychology.

Includes index.
1. Psychology, Experimental—Methodology.
2. Psychology—Research. I. Title.

BF181.C245 1985 150'.72 84-9100

ISBN 0-03-018851-2

Address correspondence to
383 Madison Avenue
New York, NY 10017

Printed in the United States of America
Published simultaneously in Canada

5 6 7 8 9 039 9 8 7 6 5 4 3 2 1

CBS COLLEGE PUBLISHING
Holt, Rinehart and Winston
The Dryden Press
Saunders College Publishing

Cover illustration: Fred Otnes / Bill Erlacher Artists Associates

Composition by York Graphic Services, Inc., York, Pennsylvania
Printing and binding by R. R. Donnelley & Sons Co., Chicago

To the memory of W. G. Chase

Preface

Ten years ago I wrote *Human Experimental Psychology*. That text combined methodology with content, and broke new ground by focusing on studies of human cognition as the basis for the undergraduate course in experimental psychology.

The tendency today is to concentrate once more on experimental methods: it is almost taken for granted in many places that the examples will be largely from studies of human cognition.

The goal of this book is to introduce the student to the work of the experimental psychologist, to the reasoning that underlies the development of a research question, the planning of an investigation, and the analysis and interpretation of the findings. Theory, measurement, design—these are central themes in this text. In addition, and consistent with the concentration on human cognition, there is detailed consideration of the "human" side of the experiment. Finally, separate chapters deal with the tasks of reviewing the literature, reading a research study, and writing an investigation.

This is not a text on statistics. Some students will have had an introduction to psychological statistics prior to the course; for them, the appendix should provide adequate review of basic statistical concepts and procedures. For the student who has not had statistics, the appendix will need instructional support.

After Chapter 1, most of the chapters are designed to stand alone. The order of presentation suits my purposes and style, but a different sequence should be quite feasible. The text is relatively lean, but has sufficient examples to flesh out the concepts. It is not a compendium; I have been relatively sparing of references. These decisions are consistent with the emphasis on method rather than content.

Several people have from time to time labored with me on this work; reviewers and editors contributed substantially to a clearer and more accurate final version of the book. I wish particularly to acknowledge James Anderson, Troy State University; Edwin Brainerd, Clemson University; Thomas Mehle, University of Nebraska/Lincoln; Terry Pettijohn, Ohio State University/Marion; Lanna Ruddy, SUNY/Geneseo; and Philip Young, Towson State University. The ideas and the techniques communicated spring from many other sources as well—instruction, guidance, and examples have come from many teachers, colleagues, students, and friends over the past twenty years. Finally, Jay Thorp has been the constant companion who kept all the pieces together, corrected my stylistic faux pas, asked good questions, and handled other tasks too numerous to list. I am especially grateful for her help.

R. C. C.

CONTENTS

1

The Structure of Psychological Experiments

This book is about experimental psychology. More particularly, it is about the *methods* of experimental psychology, because it is these methods that distinguish the field. While we might define experimental psychology as "whatever experimental psychologists do," many researchers besides experimental psychologists use the same techniques: social and clinical psychologists, educational and developmental psychologists, linguists, anthropologists, sociologists, and even a few computer scientists now employ experimental methods in behavioral research.

In studying the various aspects of the experimental method of research you will recognize some overlap with other research methods. For instance, description is the foundation for virtually all scientific research; the naturalist depends on careful description as much as the anthropologist does. Hence, this book covers the descriptive techniques needed for experimental investigations. Issues of measurement—what it means to measure something, how one establishes the validity and reliability of a measure, and so on—are common to all science, and so this topic is also addressed.

Our primary aim is to explore the experimental method; to set forth the principles underlying it; and to describe the manner in which experiments are planned, observations taken, and data analyzed and interpreted. Several experimental techniques will be described, followed by a couple of examples, one from research in the laboratory and a second from reports of a more practical character. You will see how the behavioral researcher plans an investigation and carries it out, and how he or she makes sense of the findings and communicates them to others. You will learn about the importance of theory in formulating a research question and how theory aids in interpreting the results of an investigation.

This book emphasizes the study of cognition or thought processes in human beings. Such an emphasis does not mean that the experimental method is limited to these domains; a good deal of research in experimental psychology uses animal subjects, and many psychologists restrict themselves to the investigation of regularities in observable behavior without any particular concern for the underlying processes. The decision to concentrate on human psychology is based largely on the need for simplicity; the concepts and techniques for the study of animal behavior differ significantly from those for research on human beings. The focus on cognitive processes reflects the growing importance of this area within the field of experimental psychology.

This introductory chapter presents an overview of the reasoning behind the experimental method in psychology. The task of formulating a research question is discussed, as is the job of reaching a meaningful answer to the question. Research is by its nature a risky enterprise, with a lot of stumbling around in the dark; if the researcher already understood the phenomenon under investigation, experiments would be unnecessary. The challenge is to formulate the right question, after which follow-up studies may be relatively straightforward. Even in the best of circumstances, however, breakthroughs are less common than confusion and failure. In doing research it helps to have a certain amount of persistence and drive, as well as a dash of optimism and humor.

In this chapter you will also find a list of the major "building blocks" used in experimental research. As you have probably discovered in other science courses, one of the hurdles in any field of study is the vocabulary. Many of the terms in this chapter will be new to you; others will be familiar words with new meanings. These words will be used again and again in later chapters, and this first pass is intended to alert you to the importance of certain concepts and their labels.

WHAT'S THE QUESTION?

Psychological research serves two primary purposes. First, it helps us discover the variables that determine human behavior and the circumstances under which these influences hold sway. Second, it leads to the creation of theories that help us understand how people think and act. In achieving these ends the researcher's first step is the formulation of a scientific question, one that is clear and precise and that can be answered by objective evidence.

"Does noise slow a person's reactions?" This question is more complex than might appear at first glance but, suitably phrased, it can provide the starting point for a scientific investigation. Certain points require further elaboration. How is noise to be defined? Noise is not the same as loudness; it is characterized by a quality of obnoxiousness. Kryter (1966), noting that "there is a basic 'unwantedness' or 'noisiness' to sound beyond that due solely to its measurable loudness" (p. 1346), identified three features that appear critical to noise: high pitch, complexity, and length (more than a quarter-second). This analysis of basic issues is essential to formulating a research question. Notice that the definition offered by Kryter begins to identify the variables of potential relevance. The researcher would also have to answer a number of other questions: What reactions? What kinds of people? What kinds of situations? But all these questions are answerable. The question that begins this paragraph thus provides a reasonable basis for a scientific inquiry.

Not all questions are scientifically answerable. "Is noise bad for you?" cannot be answered scientifically because the answer depends on a value-laden judgment; one must resort to something other than the scientific method to decide the meaning of *bad*. Nor is "Do loud noises frighten away ghosts?" a valid scientific question, because it is impossible to obtain objective evidence about ghostly behavior. Until we develop more trustworthy methods for recording the activity of spirits, this question and others like it must remain outside the sphere of experimental inquiry.

Although personal experiences, thoughts, and feelings cannot be studied directly, people's reports of their experiences can serve as data for scientific inquiry. "Does noise really make people feel more irritable?" The word *really* is the problem here. We can certainly ask people to describe their feelings and experiences: "Tell me how you feel when you are exposed to various kinds of noise." The responses to such a question comprise objective data, which may demonstrate that people experience increased irritation in noisy environments. We define certain types of behavior as indicators of irritation and then show that these behavioral indicators increase with an intensification of noise. For instance, the subject may shout "I can't take this racket any more!" and leave the experimental room. Even this dramatic response is not a sure sign of what is *really* happening in the subject's mind. Nor is it essential to wrestle with this issue; for practical purposes we can go a long way by relying on observable behavior and verbal reports.

WHAT'S THE ANSWER?

An experiment is typically designed to test one or more hypotheses. The researcher may suspect that noise slows down a person's reactions and plans a study to investigate this proposition. The primary focus is on the effect of *noise* versus *no noise* on some performance measure. Other variables, many of which are not of immediate interest, may also influence performance. Noise may slow down the reactions of naturally calm people but speed up those of nervous people. The sound of jet planes near an airport may slow the work rate in nearby factories while rock music may increase it.

Suppose that, cognizant of such complexities, the researcher decides nonetheless to ignore all factors except the one or two of immediate concern. The "answer" yielded by such an experiment, whether or not it supports the investigator's hypothesis, will almost certainly be of limited applicability and may in fact be misleading. The findings of any given study depend on the context in which the study is carried out. In the present example, the effect of noise on people's reactions is likely to depend on the type of noise, its duration and intensity, the subject's ability to tolerate (or appreciate) the noise, the reactions being measured, the length of the study, and so on. For the results to be meaningful the investigator must identify these secondary sources of variation and must find a way to *control* them.

Research on behavioral and social questions is a challenge. Many factors may affect the outcome of any given study, and we lack trustworthy theories to guide our thinking about these complexities. As a consequence the researcher faces some tough decisions in planning an experiment. Which variables are most relevant to the outcome of the study? What kinds of subjects should be tested, and how many? A multifaceted experiment with many variables built into the design constitutes a rich source of information, but it is also difficult to plan and to manage. A simple design with one or two variables is easier to carry out, but the results are likely to be ambiguous and untrustworthy.

In a study of the effects of noise, for instance, should the subjects be male, or female, or both? If only one person is tested, the effects may depend in some unknown way on the characteristics of the particular individual.

	Ability?
Subject 1	Motivation?
	Gender?
	Age?
	Et cetera?

If two people are tested, then the investigator has at least a peek at individual differences. For instance, if gender is thought to be related to the effects of noise, then the experimenter can test one male and one female.

	Ability?
Subject 1	Motivation?
Male	Age?
	Et cetera?
	Ability?
Subject 2	Motivation?
Female	Age?
	Et cetera?

If resources permit the testing of four subjects, two subject variables can then be included in the design. For instance, suppose the researcher has a hunch that both gender and age are important determinants of the influence of noise on performance. With four subjects in the design, two college students and two middle-aged subjects—a male and a female from each age level—can be selected. This plan allows the experimenter to separate the effects of gender and age.

Subject 1	Ability?
Male	Motivation?
College	Et cetera?

Subject 2	Ability?
Male	Motivation?
Middle-aged	Et cetera?
Subject 3	Ability?
Female	Motivation?
College	Et cetera?
Subject 4	Ability?
Female	Motivation?
Middle-aged	Et cetera?

If the two middle-aged subjects both find noise objectionable compared with quiet while the college students report no difference, age may be an important factor in considering the effects of noise. If males and females both find noise equally objectionable (or rewarding), then gender can be eliminated as a control factor.

Now we can put together the "question" and the "answer." The researcher formulates a psychological question, in part by identifying a set of variables that are hypothesized to contribute substantially to variation in a performance measure. The choice of variables represents the investigator's "best guess" about the state of the world. The "answer"—the results of the experiment—depends greatly on the soundness of these initial decisions. The findings will be inconclusive if the primary variables are weak or if the researcher ignores secondary variables that produce large and unpredictable fluctuations in performance. Sometimes the answer is clear-cut but quite different from the experimenter's original hypothesis. If the original decisions are sound, the experimenter can learn something even from unexpected findings, which will help in planning a follow-up study. Surprises are not uncommon in behavioral research, and virtually never is a question answered with any certainty from the results of a single study.

PSYCHOLOGICAL THEORIES

A considerable amount of psychological research is generated by the psychologist's curiosity about the effects of particular variables on observable behavior. Empirically based experiments answer straightforward questions of the form "What will happen to X if I vary Y?"

- What is the effect on worker productivity when noise level is varied?
- What is the recall rate for strings of numbers ranging in length from six to twelve?

- If the goal is to ensure that people wear seat belts when driving , what is the relative effectiveness of (a) buzzers, (b) interlocks (the car will not start until the belts are fastened), (c) laws (the police arrest anyone whose belt appears not to be fastened), and so on?
- What is the relation between the number of people in a small group and the time taken to solve various types of problems?

Theoretical questions are generated by a different kind of curiosity, in which the psychologist puzzles about the events that underlie observable behavior. The researcher faces the same decisions in planning a theoretical study as in designing an empirical investigation. One must still select a set of factors and plan a design that will provide trustworthy and interpretable data. The model aids in selecting the variables and in interpreting the outcome of the research.

The most active theoretical arena in psychology at the present time falls under the heading of **information-processing theory**. The focus of this theoretical approach is on mental processes. The mind is seen as a kind of processor of information, which takes various inputs from the "outside world," transforms and translates them, and then produces behavioral responses as outputs.

Suppose you are a researcher interested in how a student adds two numbers. You select several addition problems and present them to a sample of students. You record the answer given by each student to each problem and the length of time taken to produce the answer. The observable measures are correctness and speed, both of which might be expected to vary with the difficulty of the problem (50 + 20 = ? is likely to be easier than 47 + 26 = ?). The observed relations between performance and difficulty do not tell you how the student is thinking, however. For this you must construct a model describing the thought processes in which difficulty can be related to performance in some clear-cut way. The predictions of the model can then be compared with the experimental findings.

To illustrate, consider two theoretical models for describing the mental processes that underlie the addition of a pair of single-digit numbers. One possibility is that the student computes the sum by mentally "counting on the fingers." In this case the larger the sum, the longer it should take to count up to the answer (7 + 9 = ? should take longer than 3 + 2 = ?). A second possibility is that the student looks up the numbers in a table in memory. Since the student must search the table for the answer no matter what the value of the two digits, the amount of time it takes to find the answer should be the same for any pair of digits.

How might we evaluate the two models? Both predict that the student will give the correct answer, and so data on accuracy are irrelevant to the comparison. The two models do differ in their predictions about the time required to add a pair of numbers. The counting model predicts that the response time will increase as the sum of the two numbers goes up. We therefore need to arrange all the sums in order of increasing value: 0 + 0 = 0 is lowest; 0 + 1 = 1 and 1 + 0 = 1 are next; then 0 + 2 = 2, 1 + 1 = 2, and 2 + 0 = 2; and so on. The counting model predicts

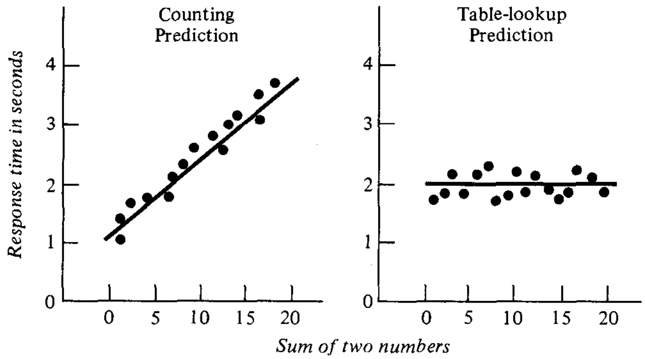

Figure 1–1 Hypothetical results based on counting models and table-lookup models of mental arithmetic

a steady increase in response time as the sum increases (Figure 1–1). The table-lookup model predicts that response time should be the same for all sums, as illustrated by the flat line in the right-hand panel of the figure. The scattering of dots in the two panels shows the kind of data that would be consistent with each of the models; we expect some variability or "noise" in the observations, but the general pattern of results is quite different for each.

As for any other kind of research, a theoretically oriented experiment must include attention to the fundamental question: What variables besides those of primary interest might influence performance? In the present case these variables might affect not only the trends in the data but the appropriateness of one or the other model. For instance, we might expect that the way the student has been taught to add might be important. Some mathematics programs emphasize rote memorization while others stress the concept of counting. Students in the first program might be expected to operate according to the table-lookup model, while students in a concept-oriented program would be more likely to perform in the manner of the counting model. The amount of practice a student has had may also matter. Younger students often find fingers a convenient device in calculating sums, but with practice the older student becomes quite fast with all sums. The researcher's instructions can make a difference, too. An emphasis on speed may cause a student to resort to a table-lookup strategy, while stress on accuracy may lead a youngster to cross-check by counting.

ONE GOOD QUESTION LEADS TO ANOTHER

Experimentation is a dynamic process. In planning an experiment the researcher makes many decisions on which the outcome of the study depends. The results may confirm some predictions, but there are likely to be surprises. These can lead to additional analyses, an altered set of decisions, and even a new experiment.

For instance, suppose that an experiment shows that exposure to noise slows the reactions of middle-aged people but has no effect on, or actually reduces, the reaction time of younger subjects. For a follow-up experiment the researcher might decide to investigate the underlying processes, the previous experiences, or the attitudes and preferences that could have produced the variation due to age.

As a second example, suppose that the experimenter has evaluated the two models of mental addition proposed earlier and that for second graders *neither* model describes the data adequately. Instead, the response time increases with the value of the smaller of the two numbers, so that $3 + 2 = ?$, $2 + 9 = ?$, and $6 + 2 = ?$ all take about the same time. Based on this pattern, the experimenter might then propose a model in which the student starts with the larger of the two numbers and then adds the second number by counting. This model predicts that the size of the larger number is not relevant, so that $13 + 2 = ?$, $2 + 17 = ?$, and $21 + 2 = ?$ should take the same response time, as should any other sum involving 2. This analysis can help the researcher plan the follow-up experiment.

A point made earlier deserves reemphasis here. Suppose that an experiment has been well planned and carefully implemented, but the findings show that the data contradict the hypothesis under test or that the critical variable has no effect. Such outcomes do not mean that the experiment is a failure. Psychological research is not a matter of setting up a batch of hypotheses like so many bowling pins and then trying to knock down as many as possible. Rather, the process is one of continual exploration. The researcher formulates a question based on practical need, intellectual curiosity, or theoretical analysis. The first study reveals something about the nature of the answer but is also likely to be incomplete or uninformative in some respects and to raise some new questions. The next study sheds additional light on the topic but also turns up new issues and puzzles. Sometimes the researcher makes a major breakthrough, and fireworks light the sky (or at least the pages of a journal). More often, research is a slow and painstaking activity, a matter of tracing a path through a complex maze, with many false turns and dead ends—and without any real assurance that the maze even has a solution.

FACTORS, LEVELS, AND RESPONSE MEASURES

The set of arrangements for carrying out an experiment is called the **experimental design**. This section of the chapter defines and illustrates the basic elements for constructing a design: factors, levels, and response measures.

Factors

A *factor* is a variable that the experimenter defines and controls in such a fashion that its effects can be evaluated. A factor is also referred to as an *independent* *variable*. In the experiment on noise, the factors might include the presence or absence of noise, the type of noise, its level and intensity, and the age and gender of the subjects. In the experiment on mental addition, the factors might be the sum of the two numbers, the method used to teach addition, the amount of practice, the experience of the students, and the instructions for the task.

It is helpful to distinguish three categories of factor: treatment, subject-classification, and nuisance factors. A **treatment factor** is an aspect of the environment that is directly manipulated by the experimenter; we generally think of these factors as variations in "experimental" treatments. For instance, suppose the researcher is interested in the effect of the amount of reward on performance in a learning task. Variation in the reward factor might entail paying one group of subjects ten cents for every correct answer while paying subjects in another group one dollar.

A **subject-classification factor** is a characteristic of the subject that is intrinsic and hence not easily modified but one that can be taken into account by the experimenter when selecting the sample of subjects. If the researcher includes both nine- and twelve-year-old students in a study, then age is a subject-classification factor. Gender, ability, and family background are other examples of subject-classification factors. Variables in this category allow the researcher to assess the extent of differences between individuals.

Nuisance factors are variables the experimenter builds into the design, not because of intrinsic interest in the effects but because the results are likely to be difficult to interpret if these factors are not included. Such factors are defined by their function in the design of the study and may be either treatment or subject-classification factors. While nuisance factors may be of no particular interest, the researcher cannot afford to ignore them. For instance, if the subject is tested several times during an experiment, then time should be included as a nuisance factor in the design. A person's performance may change over test trials, either because of learning or boredom or both. If the experimenter disregards the time factor, then changes in performance from one treatment condition to another may reflect variation due to the treatment conditions, the passage of time, or any combination of these two factors.

One point in the preceding paragraph deserves emphasis because of its importance: the designation of a variable in the study. A variable may be a nuisance factor in one design and a treatment or subject-classification factor in another. For instance, the researcher interested in the effects of payoff on learning rate is advised to include equal numbers of male and female subjects in each of the reward conditions. Even though the investigator may not be especially interested in the influence of the gender factor, only by including this variable in the design will there by any assurance that the findings are applicable to both sexes. A second researcher, with primary interest in the investigation of individual differences due to gender, may consider the male–female factor of primary interest as a subject-classification factor and variation in payoff a secondary or nuisance variable.

When deciding which nuisance factors to build into a design, the researcher is wrestling with the issue of **control.** The concept of control is dealt with more fully in Chapter 6; for the present you may simply imagine that an experiment is poorly controlled to the extent that one or more of its facets are not well designed. Control techniques consist of a set of procedures for designing, carrying out, and analyzing an experiment so as to ensure that the data provide a clear and convincing answer to the experimental question.

Levels

Each specific variation in a factor is called a **level** of the factor. In this book I will also on occasion use **condition** to refer to a level of factor, especially for treatment factors.

For some factors the number of levels is small. Male and female, for instance, are the only two levels for gender. Eye color and blood type are other examples of subject-classification factors with a small number of levels. Levels are often defined in a binary, "yes–no" fashion: *noise* versus *no noise* or *reward* versus *no reward*, for instance. Whenever the number of levels is restricted, the researcher faces a relatively simple decision; if the factor is to be incorporated in the design, then all levels will be included.

More often, the number of potential levels of a factor is very large, and so the choice of levels for the design is a major consideration in planning it. For instance, a factor may have many levels because it is quantitative; the subject's age, the duration or volume of a tone, or the amount of practice on a task can all be measured on a numerical scale, and so the experimenter has a virtually infinite number of choices. To be sure, some choices make more sense than others. Suppose the researcher is interested in the effects of the duration of a noisy tone on perceived discomfort. The experimenter might select 2.837 and 5.369 seconds as the levels of the tone-duration factor, but these levels would strike most people as rather bizarre.

A factor may also have many levels because it is inherently complex. For instance, the set of all brief verbal reinforcements is rather large: "That's fine!"; "Right!"; "Gee, I couldn't agree more!"; "That's the best idea I've heard yet!"; among others. While the experimenter's first inclination might be to consider every variation as a separate level, it would be more reasonable to attempt to categorize the various responses into a relatively small number of levels.

Response Measures

The researcher evaluates a subject's performance by means of one or more **response measures,** which are also referred to as *dependent variables*. Behavior in any given situation can generally be measured in a wide variety of ways, some