BEHAVIORAL RESEARCH

ASSESSING THE VALIDITY OF RESEARCH FINDINGS IN PSYCHOLOGY

PAUL D. CHERULNIK

Behavioral Research

ASSESSING THE VALIDITY OF RESEARCH FINDINGS IN PSYCHOLOGY

Paul D. Cherulnik

Susquehanna University

HARPER & ROW, PUBLISHERS, New York
Cambridge, Philadelphia, San Francisco,
London, Mexico City, São Paulo, Sydney

To my Mother and Father

Sponsoring Editor: George A. Middendorf
Project Editor: Pamela Landau
Production Manager: William Lane
Compositor: Com Com Division of Haddon Craftsmen
Printer and Binder: R. R. Donnelley & Sons Company
Art Studio: Fineline, Inc.
Cover Design: Diane Saxe

BEHAVIORAL RESEARCH
Assessing the Validity of Research Findings in Psychology

Library of Congress Cataloging in Publication Data

Cherulnik, Paul D., 1941–
Behavioral research.

Bibliography: p.
1. Psychological research. 2. Psychology—Methodology. I. Title. [DNLM: 1. Behavioral sciences. 2. Research—Methods. 3. Research design—Standards. BF 76.5 C523b]
BF76.5.C46 1983 150'.72 82-21354
ISBN 0-06-041258-5

Preface

When I first got the assignment to teach a course in research methods (Principles of Psychological Research it was called), I was frightened by it. It seemed such a bewildering array of complex, abstract concepts to try to get across to students who, for the most part, weren't planning to become researchers anyway. Glancing through the textbooks that were available for the course did little to calm my fears. They seemed to me to lack any real logical organization. Even worse, they mixed in with discussions of central questions about methodology apparently unrelated topics (from my point of view) such as psychophysics, statistics, and research report writing. Those books weren't going to help me to get organized to teach all those things I had learned over so long a time, from so many sources, without really knowing myself how I had done it.

In the midst of paging through texts, looking for one that would help me and my prospective students, I somehow remembered hearing about a new approach to understanding research methods that had been published several years earlier, first as a chapter in a handbook of methods for educational research (Campbell and Stanley, 1963) and then reprinted by itself as a paperback (Campbell and Stanley, 1966). The fact that I knew one of its authors to be one of our leading thinkers about methodological problems, and the rumors I'd heard about it becoming a kind of underground classic, persuaded me to search for it. When I finally read it, I was impressed more than I had been by anything I'd read about research methods up to that point. The analysis it provided of the strengths and weaknesses of various approaches to doing research was coherent, well organized, and complete. That slim volume of less than 100 pages covered more of the conceptual ground, could be applied to more fields of research, and seemed more logical than I had believed possible. Moreover, I felt I had learned more fundamental truth about research methods

in the few hours it took to read that book than I had in the ten years I had studied psychology up to that point. Most important, I believed that the truth could be explained to undergraduate students of psychology.

I finally decided that this was the approach I had to take to teaching the methods course. However, the Campbell and Stanley book was not appropriate as a text for my students. It presumed too much background. There were too few definitions of basic concepts and too few examples for people who hadn't already done considerable background reading in the behavioral sciences. So I wrote an outline of the system Campbell and Stanley had devised, reorganized it a little, added a lot of examples and definitions, and presented it to the students orally and on the classroom blackboards.

Thankfully, they were able to learn a great deal that way, and over the succeeding years my organization, definitions, and examples improved steadily. But the students always felt somewhat uncomfortable without a textbook to carry around with them, which they could consult at their convenience, without waiting for the next class to ask for clarification of a puzzling point or to see where the story went next. I assigned books from time to time that contained bits and pieces of the system I was trying to teach, more than anything else just to give my students and myself the comfort only a textbook can provide. Finally, however, I decided to turn my own notes into a new book. Together with the considerable rethinking that putting it all down formally on paper required, and the helpful suggestions of colleagues who read parts or all of it, the result was the book you are about to read.

Despite my faith in the basic structure of the system we'll be using to understand research methods, I wouldn't want to give you a false sense of security. The issues we will be working to understand are extremely complex. Persons who feel they understand all there is to know about methodology (as the old joke goes) simply fail to comprehend the seriousness of the problems they face. There are many practicing researchers, as the examples you will be reading about should convince you, who haven't got them all straight. There are important questions about behavior for which there are no ready methods of arriving at an acceptable answer. And there are some very vigorous disagreements about how to solve many methodological problems.

The complexity of these problems makes it necessary to be patient in your approach to them. Some of the concepts that have to be understood can only be developed slowly, through repeated explanations and examples. There are no simple definitions that can be extracted. Although there is a glossary of terms at the back of the

book the definitions found there can only help to jog your memory, reminding you of examples or bits of the discussions of the concepts. True understanding of the issues can be earned only by reading and thinking about them through the entire course of the book. The greatest strength of the approach we'll be taking is that it is organized as a system. Knowing one part of it increases one's understanding of every other part. At the same time, this makes it impossible to convey a full understanding of a single issue by a concise definition. There are no shortcuts.

Rest assured, however, that your patience will be rewarded. The importance of understanding research methods goes far beyond your studies in psychology or any other behavioral science. I think you'll find, as I have, that it may be even more important as you read newspapers and magazines, or watch the news or other public affairs programs on television. In a very real sense, an understanding of research methods is a guide to thinking, not only about science but also about important questions of many other kinds.

There are many people to whom I owe thanks for helping me in writing this book. I owe Donald Campbell and Julian Stanley a great intellectual debt. They taught me a great deal of what I know about research methods and started me on the way to learning the rest. Much of what I have learned through that inspiration I owe to so many others who have devoted themselves to the analysis of methodological questions. Some of them have been credited for their ideas in the text, but those who have not will have to settle for this anonymous mention. There have also been many collegues who have read sections of drafts of the manuscript. They found many of my errors, but only I can be blamed for the rest.

On a more personal level, there are also many people to whom I am grateful. My wife Beverly has put up with my absences and my moods and has collaborated on the work as well. She has provided love and ideas and clerical help throughout the entire project, and it never would have been finished without her. My parents encouraged me to follow a career of scholarship and I hope that this book, dedicated to them, resolves some of their doubts about the wisdom of that action. Many teachers have inspired me and I would especially like to thank one of them, Ed Hollander, for the confidence he has shown in me over the years. Jack Bevan, who was my boss at the College of Charleston, helped considerably during the period when I did most of the writing there. George Middendorf, my editor at Harper & Row, showed great patience and provided necessary encouragement. His personal wit and charm helped me to open those dreaded envelopes containing criticisms and suggestions from reviewers. And there have been too many others—colleagues, stu-

dents, and other friends—to thank them all personally. If this book makes a contribution, all these people deserve a share of the credit. I only hope that some of those who read it will someday want to include us in a list like this one.

Paul Cherulnik

Contents

Chapter 1
Toward an Understanding of Validity

INTRODUCTION

Those who pursue the study of psychology beyond the introductory level may often be surprised to learn that many of the research findings once presented to them as facts about behavior have been challenged. Advanced courses are often given over, in large part, to the consideration of conflicting theories about various aspects of behavior. And it is rarely possible to choose among these theories on the basis of research findings. Those on each side of an issue seem to be able to muster sufficient supportive evidence to create a standoff.

This condition prevails in all the various subfields of psychology, whether social or developmental or learning. It is also widespread in other behavioral and social sciences, such as sociology and political science, and in such diverse fields as history and philosophy, and even physics and mathematics. It seems to be a rather general rule that the further one progresses in the study of any field, the more one is exposed to uncertainty regarding the true nature of the phenomena with which that field is concerned. Students in an introductory psychology course may learn that reinforcement increases the rate

of response, and they may confidently believe that fact for some time. Then, in an advanced course in social psychology, an instructor may present research evidence which seems to prove that some people's behavior is not modified to any substantial degree by reinforcement.

In point of fact, the literature of psychology, and every other scientific discipline, consists of many such inconsistencies. How, then, can one decide among the varying opinions of psychologists, and other scientists, which seem to cloud every issue of importance? One key, it turns out, is a very complex set of principles and techniques that is often referred to as "the scientific method" or "methodology." The purpose of this book is to introduce the reader to some basic principles that make it possible to evaluate the contradictory findings and assertions that plague every scientific discipline.

Unfortunately, it cannot be claimed that a complete resolution to this problem will be effected. One reason is that the approach to be taken here is directed primarily toward research in psychology. That is not to say that the principles which we will consider do not apply to other disciplines as well. But, although they do, the examples that we will use and the issues with which we will deal most fully will be drawn from the literature of psychology. Therefore, they will, inevitably, be most useful in and most directly applicable to understanding the subject matter of that field. As we shall see, one of the complexities of the scientific method is that each field of study presents its own peculiar difficulties of evaluation, in addition to its share of the general problems which cut across disciplinary boundaries. We will see that even within the single discipline of psychology, different research problems may require different research methods in order to cope with their special difficulties. For example, psychological research with human subjects can present problems that are not encountered by psychologists who study the behavior of rats or pigeons.

The study of the scientific method is also complicated by the number of skills that are required to evaluate the findings of research. Many of these lie outside the scope of this book. They include statistics, a substantive knowledge of the phenomenon under investigation, including its literature of past research findings, and general knowledge that has nothing to do per se with either psychology or the scientific method. In addition, the abilities to conceptualize complex problems, and to observe behavior with sensitivity and objectivity are very important, as are such mysterious abilities as creativity and the ability to synthesize.

However, we will be able to make a good start toward being able to understand and evaluate the methods of psychological research. We will stress understanding and evaluating research, rather than

doing research, because of its wider importance. Few students of psychology will ever do research themselves. Even those psychologists who are active researchers come into contact with research findings more often as consumers of research findings from professional journals or the popular media. And every consumer, including the consumer of the products of science, needs to be able to evaluate and choose among competing products, which are all offered as the latest and best on the market. It is no less important for a psychologist to buy a good understanding of human behavior than for a driver to buy a good automobile. And science is important not only for the researcher, whose own work depends on an understanding of what has gone before, but also for the average citizen, whose participation in the democratic process depends on an understanding of the world he or she lives in.

The plan of this book has been adapted from a pattern that was laid several years ago by Donald T. Campbell and Julian C. Stanley (1963). The purpose of that plan is to make the basic principles of the scientific method understandable. Its greatest strengths are, paradoxically, its simplicity of structure and the diversity of issues which that simple structure can accommodate. The first four chapters of the book describe the plan and structure and introduce the basic concepts and a framework for organizing them. The remaining chapters evaluate a sample of each of the variety of research methods that can be used to test hypotheses in psychology—and in other disciplines as well. We will also consider some of the methodological issues which are currently being debated most vigorously in psychology.

A Sample of Psychological Research for Analysis

We will begin by considering an example of contemporary psychological research. This study was not chosen to be representative of the wide range of psychological research. No single study could be. It does, however, introduce some basic concepts, which we will refer to again and again throughout the book. And it demonstrates early on how difficult it can be to decide whether the findings of a particular piece of research can be believed. Most important, this study will help us to define the concept of *validity,* the technical term we will use to represent the complex judgment about the truth of research findings.

Before we go any further, a word about the word "study." There are many ways to categorize the research methods used in psychology. You are probably familiar with many terms already, such as experiment, investigation, study. In Chapter 4, we will look at one scheme for classifying research methods which we feel has certain

advantages that recommend its use. Until then, we will use the term "study" to refer to any research, regardless of the methods it employs.

The study in question was performed to test the hypothesis that one's success or failure in performing a task can be affected to a measurable degree by one's initial expectations. That is, it was directed to this question: Is it true that, other things being equal, people who think they will do well on an assigned task will actually outperform those who undertake the same task thinking they are likely to perform poorly? This unpublished study of mine was carried out by a student, Max Zachau, who acted as *experimenter,* the person who actually does the work of performing a study, meeting *subjects* (those whose behavior is studied), giving them instructions, distributing and collecting materials, etc.

To investigate our hypothesis, 48 of approximately 500 male students in various sections of an Introductory Psychology course were selected, simply because they were available during the times that we had chosen to do our study. All students taking the course were required to participate as subjects in research—a common practice about which we will say more later on. Their participation had actually begun several weeks earlier, when they and all their classmates spent the class period answering a number of personality questionnaires. Several faculty members had assembled that set of questionnaires for a variety of research purposes.

The 48 men came to our laboratory, an ordinary room furnished with several tables and chairs, in groups from two to five. The experimenter read a set of instructions which was designed to convince the subjects, first of all, that we could predict their performance on an upcoming task. The experimenter referred to the questionnaires that the subjects had filled out in their classes weeks before, suggesting that some of their answers had been designed to permit us to predict their performance on a task in which we were currently very interested. We gave them no specific reason for our interest.

The task turned out to be solving anagrams—unscrambling scrambled words (e.g., "IFSH" is a simple anagram whose solution is "FISH"). The subjects were told that their participation in our study would help us to find out exactly how well our questions predicted people's performance.

Twenty-four subjects were told that we predicted they would perform the task successfully. The other 24 were told that we predicted they would have great difficulty with the task. The experimenter chose which instructions to give each group of subjects by flipping a coin just before they arrived (until the very end, when the two groups had to be evened out at 24).

These instructions, and the staging that went with them, including the subjects' vague memories of having filled out the questionnaires in class, were our way of creating the feelings about the quality of an upcoming performance, or expectations, to which our hypothesis referred. They were also, of course, total misrepresentations of the truth. To the best of our knowledge, the questionnaire that our subjects had filled out had nothing to do with an ability to solve anagrams. Besides, at that point we had absolutely no idea what their responses to the questions had been.

The next step in our procedure was presented as a warm-up. The subjects were told that because some of them might have solved anagrams before, whereas others had not, we wanted them all to try a sample of ten anagrams before doing the actual experimental task. Then they could all begin the task with at least some experience with anagrams. This was another lie. The real purpose of the warm-up was to strengthen the effect we started to create in the first part of our instructions to them—that is, to produce two groups of subjects who had very different expectations for their performance on the anagrams task.

Months before, a large number of students in a prior semester's Introductory Psychology sections had fulfilled their requirement to participate as subjects in research by solving a large number of anagrams. On the basis of their solutions, we were able to estimate the difficulty of each of those anagrams for the average student. They varied from some that most students could solve quickly to some that almost none of them could solve within a reasonable length of time, and through every point in between those extremes. Using those estimates, we were able to create two lists of ten anagrams that differed greatly in difficulty. These lists were then used in our warm-up, each with the identical 3-minute time limit. Those subjects who had been led to believe by our earlier instructions that they would be successful anagrams solvers received a list of easy anagrams on which to get warmed up. In fact, their average score on that list of ten was 8.63. On the other hand, the subjects who had been told at the outset that we expected them to have difficulty with our anagrams task were given a much more difficult practice list. Their average score turned out to be only 0.55 out of ten. In neither case did we actually tell the subjects what their scores were. One useful feature of anagrams, at least for our purposes in doing this study, is that they are virtually self-scoring. A subject almost always knows when the word he or she has written in the blank space on the answer sheet is a correct word and when it is not. We could be fairly certain that the subjects in one group felt confident that they had done very well in the warm-up, whereas those in the other group knew they had done very poorly.

From that point on, all the subjects were treated alike. First, they were given a list of 25 anagrams to solve—the experimental task we had been building up to. These were chosen from the same pretested pool as the two warm-up lists, but from those of moderate difficulty. Each group of subjects was given 6 minutes to complete the task. After their time was up, the experimenter collected their answer sheets and began to "debrief" them. In most studies involving deception, where subjects are lied to about the procedure, the experimenter tries to follow up the subjects' experiences in two ways. First, the subjects are questioned about their perceptions of the information that was given to them. What did they understand it to mean? Did they believe it? Or did they suspect that the true purpose of the study was being hidden from them? If so, what did they think the true purpose was? Our subjects did not admit to being suspicious about the deceptive information they had been given. In the second part of a debriefing, the experimenter tells the whole truth about the study and tries to make certain that the subjects leave without any misconceptions about their abilities or personalities as the result of the deception that was employed. In the case of the present study, our subjects were told that the information given them about their ability to solve anagrams was false, that the warm-up list was rigged to make that false information more plausible to them, and that anagram performance, in any case, is not known to be an indicator of anything important about people's intellects or personalities. They were also told the reason for the deception, the true purpose of the study. Finally, they were asked to keep all this to themselves for a few days until the study was completed.

Now, to the results. Subjects' scores on the anagrams task supported the hypothesis we were testing. Those who had been told, and "shown," that they would do well solved an average of 6.87 of the total of 25 correctly. The group who had been led to believe that they would perform poorly solved only 5.22 out of 25 correctly. This difference, although not that large in absolute terms, is statistically significant. For those unfamiliar with statistics, it means that it is large and consistent enough to satisfy most psychologists that the two groups really performed differently.

THE LANGUAGE OF ANALYSIS—FACT VERSUS ARTIFACT

Now that we have a concrete example of psychological research upon which to base our discussion, it is time to introduce the basic question with which this book is concerned. Did the difference in performance between our two groups of subjects result from the conditions that we specified in our hypothesis and then tried to create