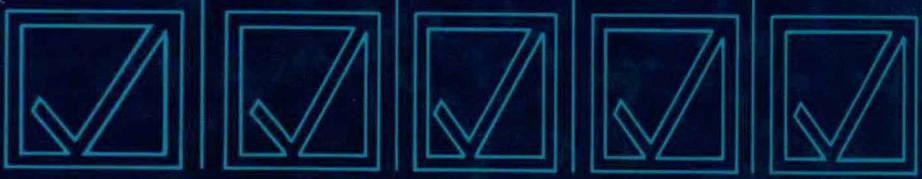
SECOND EDITION

EDUCATIONAL EVALUATION

W. JAMES POPHAM



Educational Evaluation

Second Edition

W. James Popham
University of California, Los Angeles

PRENTICE HALL, ENGLEWOOD CLIFFS, NEW JERSEY 07632

Library of Congress Cataloging-in-Publication Data
POPHAM, W. JAMES.
Educational evaluation / W. James Popham.—2nd ed. p. cm.

Includes bibliographies and index.
ISBN 0-13-240482-6 : $25.50
1. Educational evaluation. 2. Educational surveys.
3. Educational consultants. I. Title.
LB2822.75.P66 1988 87-17639
379.1'54—dc19 CIP

Editorial/production supervision
and interior design: Sue Dib
Cover design: Lundgren Graphics
Cartoons: Joan Orme
Manufacturing buyer: Carol Bystrom

Printed in the United States of America

10 9 8 7 6 5 4 3

ISBN 0-13-240482-6 01

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Prentice-Hall of Southeast Asia Pte. Ltd. *Singapore*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

Educational Evaluation

Preface

In the mid-1970s I wrote an introductory textbook about educational evaluation. Although there were several published collections of essays about educational evaluation then available, as far as I know that book was the first singly authored textbook focused on that subject. I wrote that volume chiefly for introductory courses on educational evaluation such as the one I had been teaching at UCLA since the mid-1960s. Now, over a dozen years later, I have revised that first edition of *Educational Evaluation.*

Almost any field undergoes substantial changes in a dozen years. Such has certainly been the case with respect to educational evaluation. As will be recounted in Chapter 1, the burgeoning interest of the 1970s in educational evaluation has mellowed more than a mite. The zeal of recent converts to the field has subsided. Whereas in the mid-1970s educational evaluation could be regarded as a fresh field in flux, an additional decade has allowed it to mature ever so slightly. Educational evaluators are, in general, not so brash and presumptuous as they were a decade ago. We have witnessed a galaxy of ineffectual educational evaluations during the past dozen years. Such indifferent evaluations have induced a degree of well-warranted humility in most educational evaluators, and it is hoped that such humility is reflected in the following pages.

This revision incorporates not only a more tempered view of the potency of educational evaluation, but also a treatment of relevant technical advances. Some chapters have not been dramatically modified. After all, we compute a median the same way today that we did in the 1970s. Other chapters, however, have been totally rewritten in accord with more current knowledge. A few of these major alterations are described below.

Chapter 1 deals with introductory concepts in educational evaluation and reflects a far greater awareness of the role of the educational evaluator as an *educator*—one who must typically function smack in the middle of a political maelstrom. Heightened attention is given in this chapter to the

evaluator's activities directed toward program improvement rather than program termination/retention. Chapter 2 reviews the most prevalent "models" of educational evaluation, including several of the more recent approaches to the conduct of educational evaluations.

Chapter 3, dealing with the role of educational objectives in evaluation, has been completely redone. Too many educational evaluators today are working with conceptions regarding objectives that were sired in the 1960s and should have been sent scurrying several years thereafter. Chapter 3 blasts away at some of these outmoded views of educational objectives. Similarly, Chapter 6 has been substantially rewritten to incorporate advances in the use of criterion-referenced measurement devices.

Chapter 12, dealing with cost-analysis procedures, has been totally revamped.

Chapter 13 treats the topic of teacher evaluation more comprehensively than in the first edition. These days educational evaluators are called on frequently to assist in the evaluation of classroom teachers. Greater attention is given to teacher evaluation in this edition to help ready the reader for such evaluative chores.

In a related vein, there's a brand-new Chapter 14 dealing with instructional principles. I have been astonished during the past two decades to find that many individuals who bill themselves as educational evaluators know practically nothing about the rudiments of education. Because educational evaluators are most often involved in the appraisal and massaging of educational programs, such unfamiliarity with instructional technology obviously limits the evaluator's effectiveness. Chapter 14 will not transform an instructional ignoramus into a pedagogical paragon, but it will point out the road to paragonville.

Most of the remaining chapters have been updated or reworked according to their needs. One chapter from the first edition was summarily dumped. I unplugged its life-support systems rather than engage in Herculean resuscitation efforts. All of the end-of-chapter references have been brought up to date.

Revising a textbook is something like attempting to rekindle a long-dormant love affair. In the process of reacquainting yourself with the former object of your affection, you encounter many qualities that you admire and a score that you abhor. ("How could I have ever cared for someone like *that*?" or "How could I have ever written something so patently silly?") Yet, although a refurbished romance can never quite be the same as a first love, with sufficient effort it can be quite satisfying. During the rewriting process I found myself getting caught up again in an enthusiasm for the topic. I really do like the field of educational evaluation. I hope that affection comes through in the following pages.

W. James Popham
Los Angeles

Contents

chapter 1

Educational Evaluation: A History without Mystery

Once upon a time there was a word. And the word was *evaluation.* And the word was good.

Teachers used the word in a particular way. Later on, other people used the word in a different way. After a while, nobody knew for sure what the word meant. But they all knew it was a good word. *Evaluation* was a thing to be cherished. But what kind of a good thing was it? More important, what kind of a good thing *is* it?

LOOKING BACKWARD

Evaluation as student testing For centuries the term *evaluation* has been used by classroom teachers, who thought of it only in relation to the grading of students. For most educators, indeed, the idea of evaluation was essentially equivalent to the idea of testing. A teacher might as readily say, "At the end of the term I shall *evaluate* my pupils," as say, "At the end of the term I shall *test* my pupils." To most teachers, evaluation *meant* testing.

Testing, of course, has been around almost as long as we have had recorded history. Over 4,000 years ago, for example, it was common practice in China to examine key officials every three years to determine their fitness for office.[1] Even before these Chinese civil service examinations, there were surely teachers who tried to find out the degree to which their students were learning. Future archaeologists will unquestionably unearth an artifact or two indicating that Neanderthal teachers of dinosaur-dodging employed true-false tests (which, depending on the size of the dinosaur, may have been life-death tests).

Equating evaluation with student testing was the way that classroom *teachers* used the term *evaluation.* As we shall see, other people began to use the notion in a substantially different way.

Tyler's 1930s view of evaluation In 1932 Ralph W. Tyler, then a professor at Ohio State University, was designated to be the research director of the Eight Year Study, a formal appraisal of the college performance of students prepared in "progressive" high schools versus the college performance of students prepared in more conventional high schools. During the conduct of the Eight Year Study, Tyler came to view evaluation not as the appraisal of *students*, but rather as the appraisal of an educational *program's* quality. In the Eight Year Study, of course, that approach meant determining the quality of the educational programs represented by the progressive and conventional high schools that Tyler was studying.

Tyler's fundamental strategy was to determine the degree to which the objectives of an educational program had been attained. In essence, Tyler argued that a program should be judged positively to the extent that it promoted students' mastery of the objectives that the program's architects had established prior to the program's initiation. This objective-based (or goal-based) conception of evaluation, devised by Tyler in the 1930s, was destined to influence the view of subsequent generations of educators. (A more detailed description of the *Tylerian* approach is contained in Chapter 2.) Tyler's conception of educational evaluation stimulated a number of American educators to view evaluation as something other than testing students for the purpose of giving grades.

A beeping from above In October 1957, Americans were astonished to hear the "beep-beep" of a space satellite launched by the Soviet Union. Americans were not accustomed to being beaten to the punch in such matters. How, indeed, could another nation have exceeded the United States in scientific or technical matters? To many Americans, Sputnik, the first space satellite, was a substantial source of embarrassment. When embarrassed, of course, people often search for scapegoats. In the post-Sputnik soul searching of American policymakers, the schools became splendid scapegoat candidates. After all, if U.S. schools had been doing their jobs more effectively, then American rather than Russian scientists would have been boosting spheres into space. It was clearly time to spruce up America's public schooling!

As a consequence of such "let's catch up" analyses, substantial amounts of federal dollars were given to the development of new science and mathematics curriculum approaches. These federally funded curriculum development projects were not only supposed to result in more current conceptualizations of the subject areas under scrutiny (for example, "modern math" was born in one of these projects), but also supposed to lead to the development of instructional materials by which American students could acquire updated conceptions of science and mathematics.

The professional staffs of these curriculum development projects, dominantly subject matter experts, found little virtue in the existing methods for determining whether their instructional materials worked. Such complaints led Lee J. Cronbach to author an important essay dealing with how educational evaluation might best be used in such curriculum development ventures. In a 1963 article entitled, "Course Improvement Through

Evaluation,"[2] Cronbach argued that if educational evaluation were to be of assistance to curriculum developers, it had to be focused on the decisions faced by curriculum specialists during the process of their development efforts. Moreover, he argued that evaluation activities should deal less with comparisons between programs and more with the degree to which a given program promoted its desired consequences.

Although Cronbach's 1963 essay would later be viewed as a perceptive rethinking of matters evaluational, his views failed to attract much immediate interest beyond that of the individuals most directly involved in curriculum development projects. By and large, there was little interest on the part of American educators in evaluation per se.

ESEA of 1965 In 1965 the U.S. Congress enacted a piece of precedent-setting legislation, the Elementary and Secondary Education Act (ESEA), which for the first time dished out federal dollars aplenty to local educational systems. As a consequence of ESEA, America's educational game was changed in a major way. Let's see why.

Education in the United States has historically been the responsibility of the individual states. Whereas many other nations have a federally directed educational system, educational enterprise in the United States is predicated on greater local autonomy. The financial support of schools, like their governance, has been drawn from state and local taxes rather than federal revenues. In the past few decades, however, that financial-support pattern has been shifted. Acting in part from a growing concern about the quality of the schools, federal lawmakers began to enact legislation in the 1950s and 1960s that provided financial support for schools throughout the nation. Such legislation emerged, in part, because of civil rights groups' concerns as to how the schools were serving minority students. The impact of these federal initiatives on education and, in particular, on educational evaluation has been enormous.

Many of the earliest federal education laws of the late 1950s provided modest funds for the support of research activities, particularly for specialized learner populations, such as mentally retarded or otherwise disadvantaged learners. But when Congress enacted the Elementary and Secondary Act of 1965, a comprehensive and heavily funded law providing for thousands of federal grants to local education agencies, the national government was clearly jumping into local education with both financial feet. In the debate preceding the passage of ESEA it was apparent that many federal legislators recognized the possibility that this considerable financial investment in the nation's educational system might be less effective than some were claiming. Senator Robert F. Kennedy and others contended that the new law must contain provisions for the *mandatory* evaluation of whether local agencies had used their federal grants properly. In the final version of the bill, two of its five titles (Title I and Title III) stipulated that each project conducted under support from those titles be evaluated and that such evaluations be reported to the federal government. Because of the magnitude of the funding, this stipulation meant that thousands of locally operated educational endeavors were now obligated to be

evaluated. More precisely, local educators had to evaluate a given year's ESEA Title I and Title III projects if they wished to continue to receive the subsequent year's Title I and Title III money. And money, as we know, represents an exceptionally powerful incentive for getting people to modify their behavior. With the passage of ESEA, educators suddenly shifted their evaluation activities from the realm of rhetoric to that of reality.

The federal requirements to conduct local education evaluations quite naturally resulted in the production of tons of evaluation reports that, for the most part, would have better served a school's paper drive than a nation's need to know whether its education laws were working. Educators untrained in educational evaluation simply foundered under the pressure to produce useful evaluation reports. Because the only substantial methodological training they had, if any, was in hypothesis-testing research, they tried to treat evaluation problems with a researcher's tools. More often, there was not even a misdirected effort to apply research techniques; there was merely a gathering of data, any data, so a number-laden report could be relayed to the federals. Appraisals of the resulting evaluation efforts, not surprisingly, were highly negative. Federal officials directing ESEA Title I and Title III programs as well as external reviewers[3] all concluded that the pool of evaluation expertise among the nation's educators resembled a puddle instead of an ocean.

However, this very situation, a scene in which educational evaluation was required but the would-be evaluators were nonexistent, provided the chief stimulus for what soon became a rapidly expanding field. Several outstanding educational scholars, trained in other specializations, turned their attention to the process of educational evaluation. Although there had been occasional references to evaluation in the literature of education, until the 1967 essays by Scriven[4] and Stake[5] few writers, other than Tyler and Cronbach, had addressed themselves seriously to the overall conceptual nature of educational evaluation. Prior to the 1967 publication of the papers by Scriven, a philosopher, and Stake, a psychometrician, barely legible copies of prepublication manuscripts were feverishly passed around the educational community. Interest in educational evaluation was intense. Any kind of writing on the topic was a treasured commodity.

Thus, during the last half of the 1960s, American educators in large numbers were caught up with the conduct of educational evaluations. This attention to educational evaluation was not spurred by the unwavering professionalism of educators, nor by their latent desire to avoid the waste of taxpayers' dollars. Rather, educators were *forced* to evaluate their programs in a systematic manner in order to obtain governmental funds. The federal evaluation requirements of ESEA soon were emulated by state legislatures so that even state dollars for education were accompanied by evaluation requirements. American educators evaluated their programs in the post-ESEA period not because they *wanted* to, but because they *had* to.

An era of evaluative optimism Even though they had been initially forced to engage in educational evaluation, a good many educators became thoroughgoing converts to the virtues of systematic educational evaluation. In addition, state and national legislators began to look with favor on

evaluation, for it appeared capable of indicating whether a particular legislatively supported program was worth the money the program cost. The late 1960s and early 1970s could be characterized as a period of profound optimism regarding the potential contributions of educational evaluation.

Not only was that ten-year span one of optimism, it was also an era of intense intellectual excitement regarding the nature and conduct of educational evaluation. A galaxy of conferences and workshops on evaluation were offered to educators. A spate of evaluation articles were raced into print. Many of these essays contained step-by-step "how-to" prescriptions for the conduct of educational evaluation. As will be seen in Chapter 2, whereas the enactment of 1965's ESEA found us with a dearth of formal educational evaluation models, by the early 1970s we were almost inundated by such models. Educational evaluation was in the air. It was viewed by many as the vehicle to transform shabby educational programs into shining ones. It was seen by more than a few as the long-awaited source of educational salvation.

To oversimplify a mite, many policymakers viewed evaluation as a definitive procedure for determining whether a particular program was worth its salt or, in the case of several competing programs, which program was saltiest. High-level policymakers, in particular, began to believe that formal evaluations could provide the information needed in order to determine whether a particular program should be scrapped or saved. Some newly legislated educational programs were accompanied by requirements that the programs be formally evaluated.

The educational evaluation community itself, although still fairly small in the early 1970s, began to take itself more than a little seriously. In retrospect, of course, we can dismiss the widespread enthusiasm for evaluation as the optimism engendered by recent converts. In the early 1970s, however, there was a pervasive belief that well-conducted educational evaluations could, *and should*, constitute the single most important factor in the rendering of educational decisions. Educational evaluators initiating a major project dreamed about that moment when educational policymakers, after diligently consulting the evaluator's report, would make decisions in essential accord with the report's findings. Such are the dreams of the inexperienced.

Disillusion drops by Pessimism often takes over the path trod by optimists. Such was the case with educational evaluation. The enthusiastic slogans of the early 1970s such as "Show us a decision that needs to be made, and we'll supply the evidence to make it" were replaced by more wary mottoes such as "Once in a while, if we're lucky, we might shed a bit of light on a decision."

Educational evaluators had learned a powerful lesson—the hard way—that the bulk of educational decisions are not made in a rational decision-arena in which the option best supported by credible data wins out. On the contrary, most educational decisions of importance are made in a patently political, interpersonal milieu wherein evidence plays a markedly minor role.

Educational decision makers, rather than breathlessly awaiting the evaluator's "definitive" report, typically made their choices without regard to the report's evidence of program effectiveness. Moreover, rarely did an educational evaluator's report unequivocally settle the issue of whether Program X was truly superior to Program Y. It is said that educational policymakers of the late 1970s yearned in earnest for one-armed educational evaluators, that is, evaluators not compelled to say ". . . on the other hand."

Reasonable aspirations Thus, because it has been increasingly recognized that powerful interpersonal and political factors ordinarily account for most of the action in the rendering of educational decisions, and because educational evaluations rarely contain indisputable evidence, one might reasonably wonder why educational evaluators should even *try* to enhance educational decisions by studying the quality of educational programs. Educational evaluation may appear to be a fool's mission—a mission with powerful promise but paltry payoff. And yet it is precisely that modest amount of impact on decisions to which the evaluator must aspire.

If, as a consequence of an evaluator's efforts, an enterprise can be improved in effectiveness even a few percentage points, those modest and hard-won percentages represent benefits that otherwise would have been withheld from learners. Because most educational programs are aimed at children's well-being, is it not worth the effort to carry out activities that boost, ever so slightly, the quality of learning that children achieve?

In addition, because today's evaluators increasingly recognize the political nature of the decision-making process, they can carry out their activities with more political savvy and thus will be more apt to have impact on the political community that ultimately makes the final educational decisions. Cronbach[6] argues that educational evaluators rarely should fashion their efforts to satisfy a single decision maker. Rather, he contends, evaluators should focus their efforts on informing the "relevant political community." By and large, Cronbach believes, educational evaluations are not used by individual decision makers or unitary decision-making groups. "Evaluation," he asserts, "ordinarily speaks to diverse audiences through various channels, supplying each with political communication and with food for thought."

To the extent that today's educational evaluator accepts an obligation to illuminate the thinking of a more widespread political community, it is likely that future evaluation efforts will have more impact than the evaluations of a decade ago, when it was thought that all the evaluator had to do was present a final report and then sit back while decision makers rendered decisively improved decisions.

Educational evaluators now possess more than two decades worth of trial by fire. We have learned that a score of things we thought would work—won't. We have, however, discovered a modest collection of things that do work. Today's evaluator must learn from the more than 20 years of intensive thinking and experience that followed 1965 ESEA's evaluation requirements. There is a large literature now available to the beginning