EVALUATION IN EDUCATION

Foundations of Competency Assessment and Program Review

Third Edition

RICHARD M. WOLF

EVALUATION IN EDUCATION

Foundations of Competency Assessment and Program Review

Third Edition

RICHARD M. WOLF

PRAEGER

New York
Westport, Connecticut
London

Copyright Acknowledgment

The author and publisher gratefully acknowledge permission to reprint material from:

C. E. Beeby, 1938. *The Intermediate School of New Zealand.* Wellington, N.Z.: New Zealand Council for Educational Research.

Library of Congress Cataloging-in-Publication Data

Wolf, Richard M.
  Evaluation in education : foundations of competency assessment and
program review / Richard M. Wolf — 3rd ed.
    p. cm.
  Includes bibliographical references and index.
  ISBN 0-275-93616-3
  1. Educational evaluation. 2. Educational surveys.
3. Educational tests and measurements—Evaluation. I. Title.
LB2822.75.W65 1990
379.1'54—dc20   90-44606

Library of Congress Catalog Card Number: 90-44606
ISBN: 0-275-93616-3

First published in 1990

Praeger Publishers, One Madison Avenue, New York, NY 10010
An imprint of Greenwood Publishing Group, Inc.

Printed in the United States of America

The paper used in this book complies with the Permanent Paper Standard issued by the National Information Standards Organization (Z39.48-1984).

10 9 8 7 6 5 4 3 2 1

EVALUATION IN EDUCATION

To Marie

FIGURES AND TABLES

FIGURES

TABLES

PREFACE

The preparation of a new edition of a book is a test of the durability of one's ideas. When I began writing the first edition of this book almost fifteen years ago, I was trying to address issues in the field of evaluation as I saw them then. I recognized that the field was in its early stages of development and was likely to change considerably over time. Of course, I had no idea how the field would change. Accordingly, I decided to take as nondoctrinaire an approach to evaluation as I could. In my case, this meant identifying questions that might be asked when undertaking an evaluation study and the kinds of information one would need in order to answer those questions. In contrast, a number of writers at that time were staking out various ideological positions in evaluation. Some, for example, regarded true experiments as the only way to estimate the effects of educational programs while others regarded evaluation as nothing more than a fact-gathering enterprise for administrators. I felt that such strong positions were unwarranted in a fledgling field.

My sense now is that my initial decision was not only prudent but also wise. The basic questions that must be addressed when conducting an evaluation study and the classes of information needed to answer those questions have not changed in the intervening years. Furthermore, the absence of a specific doctrine has enabled me to avoid getting entangled in useless arguments. Thus, the present edition should be seen more as an extension and refinement of some basic ideas than as a wholly new work. Of course, there is new material here—even a new chapter. But this should be seen as additional layers on the same structure.

While the basic views presented in this book have not changed, the field of evaluation has. It has gone from its infancy to become an established field. This is reflected throughout this edition, but especially in chapter fourteen. There is much that one can point to with pride in the field, and some of this is described in that chapter. But there are some dark clouds on the horizon.

There is a growing debate in the social sciences about method. For most of the past half century, the social sciences have attempted to develop objective measurement procedures to be used in as scientifically rigorous a way as possible. The collection of views, approaches, and the like that have traditionally been used come under the general heading of "positivism." There is, of course, much more to positivism than this, but it can suffice for now. Various people have expressed dissatisfaction with positivism for a number of reasons and have sought other ways of gaining knowledge in the social sciences. The debate is often referred to as quantitative vs. qualitative approaches to evaluation and research. This is an overly terse summarization of the debate. Much more is involved than simply a difference in techniques and procedures. In fact, what is at stake are the basic ways of knowing in the behavioral sciences. Some aspects of this debate are presented in chapter fourteen. It's hardly a beginning, however. The interested reader is referred to the references listed at the end of that chapter. The reason for mentioning the debate in the preface is to alert the reader to an issue that will occupy the field for some time to come. However, it should not be viewed as an impediment to the conduct of evaluation studies for two reasons. First, the approach advocated throughout this book includes the use of both quantitative and qualitative techniques and procedures. Second, it is possible to plan and carry out evaluation studies without reference to "pure" paradigms. The consensus of practicing evaluation workers is that one can be effective and successful in the field without having to use ideologically pure paradigms. The eclecticism of the approach advocated here is its serviceability.

Some years ago I eagerly awaited the publication of a book that I had admired considerably in its first and second editions. Alas, the third edition was a terrible disappointment. While it contained fresh and interesting material, it was clearly lacking something. After some reflection, I discovered what was lacking. The author used the third edition to build on material presented in the first and second editions. This was a grave mistake since it was unrealistic to expect people to read, at the very least, the second edition before tackling the third. It is unfortunate that this was reflected in generally unfavorable reviews and poor sales.

I vowed not to commit the same error. Accordingly, this edition does not depend on having read either of the previous editions of this book. For those who have read previous editions, I apologize for a fair amount of repetition and hope that you find the new material worthwhile. For people who have not read either of the previous editions, you will find everything I have to say about evaluation at this time in this book.

CONTENTS

1

THE NATURE OF EDUCATIONAL EVALUATION

INTRODUCTION

Any work that sets out to deal with a relatively new aspect of education is obliged to furnish the reader with a definition, description, and discussion of that aspect. This is particularly true of the burgeoning field of educational evaluation where there is considerable confusion. This confusion stems partly from the fact that many of the techniques and procedures used in evaluating educational enterprises are rather technical, and educators are often not knowledgeable about such matters. A more basic reason for the confusion, however, is that different authors have different notions of what educational evaluation is or should be. These dissimilar views sometimes stem from the training and background of the writers, the particular professional concerns with different aspects of the educational process, from specific subject-matter concerns, from differences in temperament, and even from differing epistemological views. A consequence of all this is that a reader unfamiliar with the field is often exposed to writings that not only differ but are even contradictory. Such writings are not just expressions of honest differences about what evaluation is and how it should be carried out. Often they are reflections of deep philosophical conflicts about what evaluation is or should be. At other times, they reflect a confusion that often attends the development of a relatively new field of inquiry.

One goal of this book is to reduce the confusion about what evaluation is and is not, how it should be organized and carried out, how the results of evaluation studies should be reported, and how they can be used. There is no intent, however, to shield the reader from honest differences that exist within the field. These can and should be exposed and discussed. However, it is not necessary or even desirable to deal with a number of highly idiosyncratic views regarding educational evaluation. Rather, the emphasis here is on the presentation of a conceptualization of educational evaluation that attempts to be comprehensive, coherent, sensible, and practical. It combines features emphasized by a number of writers in the field but attempts to weld them into a unified view of educational evaluation. It sometimes sacrifices the private concerns of writers when these might interfere with the basic ideas of evaluation. The critical reader can deal with the subtleties, complexities, and differences that exist at the frontiers of evaluation once the basic ideas are learned.

TOWARD A DEFINITION OF EVALUATION

There are a number of definitions of educational evaluation. They differ in level of abstraction and often reflect the specific concerns of the person who formulated them. At the most general level, evaluation has been defined as "a formal appraisal of the quality of educational phenomena" (Popham, 1988). This definition, unfortunately, does not help very much since it is left to the reader to determine what the terms "formal appraisal" and "quality" mean. A somewhat more elaborate definition was provided by L. J. Cronbach who defined evaluation as the "collection and use of information to make decisions about an educational program" (Cronbach, 1963). By "educational program" Cronbach meant anything ranging from a set of instructional materials and activities, distributed on a national level, to the educational experiences of a single learner. The context of Cronbach's remarks, however, were the curricula that were developed in the late 1950s and early 1960s that were intended to upgrade the quality of instruction and learning in various subject-matter areas. Cronbach's concern was with the testing and modification of the new courses that emerged from various study groups and educational laboratories. It was his belief that only by extensive information-gathering activities in actual classroom situations would it be possible to determine where and how programs were succeeding and failing so that modifications could be made on as sound a basis as possible. Cronbach's article suggested various kinds of information that could be sought in an evaluation enterprise and how these could be analyzed and used in decision making for the purposes of course improvement. His article went on to discuss a number of other issues in evaluation, some of which will be taken up later.

It is Cronbach's definition of evaluation that is of interest here. It is composed of two elements. The first, "the collection and use of information," puts equal emphasis on collection and use of information. The idea is that decisions are to be made on the basis of information, not on impressions or beliefs about how an educational program is supposed to function. In his article, Cronbach clearly stated that the kind of information he was primarily interested in was information relating to learner performance. Specifically, Cronbach wanted to find out what changes a course or curriculum produced in learners, what kinds of questions learners could answer after having studied a particular subject for a period of time, what kinds of problems they could solve, and what kinds of issues they could deal with. Cronbach asserted that this kind of information could provide the basis for sound decision making. The second element of Cronbach's definition, "to make decisions,'" denotes an action orientation. Evaluation should lead to action, as opposed to conclusions not acted on. While he does not say so specifically, he implies that evaluation activity that does not contribute to the decision-making process is a waste of time and money. Evaluation, according to Cronbach, must contribute to the decision-making process, notably to course improvement, if it is to have any justification in education.

This definition of evaluation, emphasizing the collection and use of information about learner performance for purposes of making sound decisions about educational programs, is a distinct improvement on the "formal appraisal of the quality of educational phenomena" definition, but it still does not go far enough in saying what evaluation is. A more extended definition, supplied by C. E. Beeby, describes evaluation as "the systematic collection and interpretation of evidence, leading, as part of the process, to a judgment of value with a view to action" (Beeby, 1978). This definition has four key elements. First, the use of the term "systematic" implies that what information is needed will be defined with some degree of precision and that efforts to secure such information will be planful. This does not mean that only information that can be gathered through the use of standard tests and other related measures will be obtained. Information gathered by means of observational procedures, questionnaires, and interviews can also contribute to an evaluation enterprise. The important point is that however information is gathered it should be acquired in a systematic way. This does not exclude, a priori, any kind of information. The second element in Beeby's definition, "interpretation of evidence," introduces a critical consideration sometimes overlooked in evaluation. The mere collection of evidence does not, by itself, constitute evaluation work. Yet uninterpreted evidence is often presented to indicate the presence (or absence) of quality in an educational venture. High dropout rates, for example, are frequently cited as indications of the failure of educational programs. Doubtless, high dropout rates are indicators of failure in some cases, but not all. There may be very good reasons why people drop out of educational programs. Personal problems, acceptance into educational programs, and landing a good job are reasons for dropping out that may in no way reflect on a program being studied. In some cases, dropping out of an educational program may indicate that the program has been quite successful. For example, a few years ago, the director of a community college program that was training people for positions in the computer field observed that almost two-thirds of each entering class failed to complete the two-year program. Closer examination revealed that the great majority of "dropouts" had left the program at the end of the first year to take well paying jobs in the computer department of various companies (usually ones they had worked in while receiving their training). The personnel officers and supervisors of these companies felt that the one year of training was not only more than adequate for entry- and second-level positions but provided the foundation on which to acquire the additional specialized knowledge and skill required for further advancement. Under such circumstances, a two-thirds dropout rate before program completion was no indication of a program failure or deficiency. In was, in fact, a strong indicator of success.

Clearly, information gathered in connection with the evaluation of an educational program must be interpreted with great care. If the evaluation worker[1] cannot make such interpretations himself, he or she must enlist the aid of others who can, otherwise, the information might be seriously misleading. In the above example, the problem of interpretation was relatively simple. Dropout statistics are easily gathered, and one can usually have confidence in the numbers. More complex situations arise when one uses various tests, scales, or observational and self-report devices such as questionnaires. In these situations the interpretation of evaluation information can be extremely difficult. Unfortunately, the interpretation of information has too often been neglected. Specific mention of it in a definition is welcome since it focuses attention on this critical aspect of the evaluation process.

The third element of Beeby's definition, "judgment of value," takes evaluation far beyond the level of mere description of what is happening in an educational enterprise. It casts the evaluation worker, or the group of persons responsible for conducting the evaluation, in a role that not only permits but requires that judgments about the worth of an educational endeavor be made. Evaluation not only involves gathering and interpreting information about how well an educational program is succeeding in reaching its goals, but judgments about the goals themselves. It involves questions about how well a program is helping to meet larger educational and social goals. Given Beeby's definition, an evaluation worker who does not make a judgment of value, or who, for political or other reasons avoids making a judgment, is not an evaluation worker in the full sense of the term. Whoever does make such a judgment after the systematic groundwork has been laid is completing an evaluation.

Lest the reader get the mistaken impression that the evaluation worker has great power in education, a distinction needs to be made between two types of judgments. The first is the judgment of value of the enterprise being evaluated. This is the type described above and is clearly within the scope of the evaluation worker's professional function. The second type of judgment is taken in light of the first and, along with other relevant factors, is the decision on future policy and action. This is clearly the domain of administrators, governing boards, and other policymakers. If these decision makers make both kinds of judgments, they are taking over an essential part of the professional evaluation function. This is to be avoided.

An illustration may be given. Several years ago I was involved in the evaluation of a program for disadvantaged junior high-school students held on Saturday mornings at a local school. Expectations with regard to student performance were more than fulfilled. There was a high degree of enthusiasm among students and teachers. Also, parents were pleased that their children were constructively occupied. The program, while voluntary, was consistently well attended. All in all, the program was highly successful, and the evaluation report clearly communicated this. The most difficult task in preparing the evaluation report was identifying suggestions for ways to improve the program.

Unfortunately, the program was terminated at the end of the year. The reason given was that it was too expensive. Opening a public school on Saturday morning required additional outlays for heating and light, custodial salaries, and insurance as well as for an administrator, required for legal reasons. When these additional costs and, one suspects, inconveniences were taken into account, it was decided that the program, although highly successful, could no longer be justified. The situation, however, was one in which the evaluation workers fulfilled their professional function and the administrators fulfilled theirs. The fact that the decision about future policy was inconsistent with the judgment of the program's value must be accepted as one of those unhappy situations in which other institutional factors had a determining influence on future action.

It is also possible that a decision might be made to retain a marginally effective program. It may be that the political or public value of a program is deemed important enough to continue it, despite a low level of effectiveness. It is also possible that funds may be available to operate a program of marginal quality that might not be available for other more worthwhile endeavors. It is the decision maker's job to determine whether to fund it or not. The point remains: the evaluation workers, or those charged with the evaluation of a program, should render a value judgment; it is the responsibility of decision makers to decide on future policy and action. Each has an area of responsibility, and each must be respected within their domain. This must be understood at the outset. If it is not, there is danger that evaluation workers may become frustrated or cynical when they learn that policy decisions have been made contrary to what results of their evaluation suggest.

The last element of Beeby's definition, "with a view to action," introduces the distinction between an undertaking that results in a judgment of value with no specific reference to action and one that is deliberately undertaken for the sake of future action. The same distinction is made by Cronbach and Suppes although the terms "conclusion-oriented" and "decision-oriented" were used (1969). Educational evaluation is clearly decision-oriented. It is intended to lead to better policies and practices in education. If this intention is in any way lacking, an evaluation enterprise should probably be dropped, since evaluation workers should be able to use their time to better advantage.

So far no mention has been made about what kinds of action might be undertaken as the result of an evaluation study. The range is considerable. A conscious decision to make no changes could result from a carefully conducted evaluation study, or a decision to abolish a program altogether, although the latter case is not very likely. In fact, I do not know of a single instance where a decision to terminate a program was based solely on the results of an evaluation study. Between these extremes, modifications in content, organization, and time allocation could occur, as well decisions about additions, deletions, and revisions in instructional materials, learning activities, and criteria for staff selection. Such decisions come under the general heading of course improvement and are discussed in some detail by Cronbach (1963). M. Scriven used the term "formative evaluation" to characterize many of these kinds of decisions (1967). In contrast, decisions about which of several alternative programs to select for adoption or whether to retain or eliminate a particular program are "summative"' in nature, to use Scriven's terminology. Scriven's distinction between formative and summative evaluation has achieved a fair measure of popular acceptance although the number of clearly summative studies is small. The basic idea is that evaluation studies are undertaken with the intention that some action will be taken as a result.

DIFFERENCES BETWEEN EVALUATION, MEASUREMENT, RESEARCH, AND LEARNER APPRAISAL

Beeby's definition of evaluation goes some distance toward specifying what evaluation is. However, in order to function effectively, a definition must not only say what something is, it should also say what it is not. This is particularly important with regard to evaluation. Three activities that are related to evaluation are measurement, research, and learner appraisal. Evaluation shares some similarities with each. The differences, however, are considerable and need to be examined so that evaluation can be brought more sharply into focus.

Evaluation and Measurement

Measurement is the act or process of measuring. It is essentially an amoral process in that there is no value placed on what is being measured. Measurements of physical properties of objects such as length and mass do not imply that they have value; they are simply attributes of interest. Similarly, in the behavioral sciences, the measurement of psychological characteristics such as word knowledge, neuroticism, attitudes toward various phenomena, problem solving, and mechanical reasoning does not in itself confer value on these characteristics.

In evaluation, quite the opposite is the case. The major attributes studied are chosen precisely because they represent educational values. Objectives are