EVALUATION MODELS

Viewpoints on Educational and Human Services Evaluation

Edited by

George F. Madaus

Michael Scriven

Daniel L. Stufflebeam

Kluwer · Nijhoff Publishing

Boston/The Hague/Dordrecht/Lancaster

Evaluation Models

Viewpoints on Educational and Human Services Evaluation

George F. Madaus
Michael S. Scriven
Daniel L. Stufflebeam

Kluwer-Nijhoff Publishing
Boston The Hague Dordrecht Lancaster

Distributors for North America:

Kluwer Academic Publishers
190 Old Derby Street
Hingham, MA 02043 U.S.A.

Distributors outside North America:

Kluwer Academic Publishers Group
Distribution Centre
P.O. Box 322
3300AH Dordrecht, The Netherlands

Library of Congress Cataloging in Publication Data

Main entry under title:

Evaluation models.

(Evaluation in education and human services)
Includes index.
1. Educational surveys — Addresses, essays, lectures. 2. School management and organization — Evaluation — Addresses, essays, lectures. 3. Curriculum evaluation — Addresses, essays, lectures. I. Madaus, George F. II. Scriven, Michael. III. Stufflebeam, Daniel L. IV. Series.
LB2823,E87 1983 379.1'54 82-22562
ISBN 0-89838-123-1

Printed in the United States of America

Preface

Attempting formally to evaluate something involves the evaluator coming to grips with a number of abstract concepts such as value, merit, worth, growth, criteria, standards, objectives, needs, norms, client, audience, validity, reliability, objectivity, practical significance, accountability, improvement, process, product, formative, summative, costs, impact, information, credibility, and — of course — with the term *evaluation* itself. To communicate with colleagues and clients, evaluators need to clarify what they mean when they use such terms to denote important concepts central to their work. Moreover, evaluators need to integrate these concepts and their meanings into a coherent framework that guides all aspects of their work. If evaluation is to lay claim to the mantle of a profession, then these conceptualizations of evaluation must lead to the conduct of defensible evaluations.

The conceptualization of evaluation can never be a one-time activity nor can any conceptualization be static. Conceptualizations that guide evaluation work must keep pace with the growth of theory and practice in the field. Further, the design and conduct of any particular study involves a good deal of localized conceptualization. In any specific situation, the evaluator needs to define and clarify for others the following: the audiences and information requirements, the particular object to be evaluated, the purposes of the study, the inquiry approach to be employed, the concerns and issues to be examined, the variables to be assessed, the bases for interpreting findings, the communication mode to be used, the anticipated uses of the findings, and the standards to be invoked in assessing the quality of the work.

It is a small wonder, then, that attempts to conceptualize evaluation have been among the most influential works in the fast-growing literature of evaluation. The contents of this anthology attest to the fact that there has been a rich array of theoretical perspectives on evaluation. Given the complexity of evaluation work, the wide range of evaluative situations, the political contexts within which studies

occur, the service orientation of evaluations, and the varied backgrounds and beliefs of those who write about evaluation, it is easy to understand why the various generalized conceptualizations of evaluation found in the literature differ in many important respects. The ways that evaluation is conceptualized differ over the role of objectives in the process, the desirability of presenting convergent or divergent findings, the use or absence of experimental controls, and the place of hard or soft data in arriving at conclusions. It is also understandable that given evaluators sometimes follow one general approach in one kind of evaluation assignment and a quite different approach in another setting. Since the contexts in which evaluations take place are so variable, it is fortunate that evaluators can look to the literature for optional ways to conceptualize the evaluation process in order to find one which best suits a particular context.

Amidst this diversity of conceptual approaches to evaluation, however, a consensus has begun to emerge regarding the principles that should undergird all evaluations. This consensus is embodied in the two major sets of standards for evaluations that have been issued recently by the Joint Committee on Standards for Educational Evaluation and by the Evaluation Research Society. The appearance of these standards is one sign — but a sure one — that evaluation has begun to mature as a profession.

This book is an up-to-date reflection of the conceptual development of evaluation, particularly program evaluation, and is divided into three major sections. The first includes a historical perspective on the growth of evaluation theory and practice and two comparative analyses of the various alternative perspectives on evaluation. The second part contains articles that represent the current major schools of thought about evaluation, written by leading authors in the field of evaluation, including, articles by Tyler, Scriven, Stake, Eisner, Floden, Airasian, Guba and Lincoln, Stufflebeam, Cronbach, Steinmetz (on Provus's work), Weiss and Rein, Madaus, and Koppelman. These articles cover objectives-oriented evaluation, responsive evaluation, consumer-oriented evaluation, decision and improvement-oriented evaluation, naturalistic evaluation, discrepancy evaluation, adversarial evaluation, connoisseur evaluation, accreditation, accountability, and social experimentation. This section concludes with a forecast on the future of evaluation by Nick Smith. The final section describes and discusses the recently released *Standards for Evaluations of Educational Programs, Projects, and Materials* and summarizes the 95 theses recently issued by Cronbach and Associates in calling for a reformation of program evaluation.

In one sense, the core of this book presents a set of alternative evaluation models. These are not models in the sense of mathematical models used to test given theories, but they are models in the sense that each one characterizes its author's view of the main concepts involved in evaluation work and provides guidelines for using these concepts to arrive at defensible descriptions, judgments,

and recommendations. We are aware that some writers in the field have urged against according alternative perspectives on evaluation the status of models; but we think the alternative suggestion that these alternatives be called something else, such as persuasions or beliefs, would do little more than puzzle the readers. We are comfortable in presenting the alternative conceptualizations of evaluation that appear in the second part of the book, not as models *of* evaluation as it does occur, but as models *for* conducting studies according to the beliefs about evaluation that are held by the various authors. In this sense, they are idealized or "model" views of how to sort out and address the problems encountered in conducting evaluations.

We owe an enormous debt to the authors of the articles that appear in this book. We would like also to thank the various journals that gave us permission to reprint key pieces. We especially wish to thank Ralph Tyler and Peter Airasian for writing articles specifically for this book, as well as Egon Guba and Yvonna Lincoln, who adapted their article to fit within our space limitations. We also are grateful to Phil Jones, our publisher, who consistently supported our developmental effort. Thanks is extended also to Carol Marie Spoelman, Caroline Pike, and Mary Campbell for their competent clerical assistance. Special thanks to Rita Comtois for her administrative assistance throughout the project. Bernie Richey's editorial help throughout is appreciated.

We believe this book should be of interest and assistance to the full range of persons who are part of any evaluation effort, including, especially, the clients who commission evaluation studies and use their results, evaluators, and administrators and staff in the programs that are evaluated. We believe the book should be useful as a text for courses in program evaluation and for workshops as well. Further, it should prove to be an invaluable reference book for those who participate in any aspect of formal evaluation work. We hope that this book will assist significantly those involved in program evaluation to increase their awareness of the complexity of evaluation; to increase their appreciation of alternative points of view; to improve their ability to use theoretical suggestions that appear in the literature; to increase their testing and critical appraisal of the various approaches; and, ultimately, to improve the quality and utility of their evaluations.

Contents

List of Figures and Tables

List of Figures

List of Tables

I AN OVERVIEW OF MODELS AND CONCEPTUALIZATIONS

1 PROGRAM EVALUATION: A Historical Overview

George F. Madaus, Daniel Stufflebeam, and Michael S. Scriven

Program evaluation is often mistakenly, viewed as a recent phenomenon. People date its beginning from the late 1960s with the infusion by the federal government of large sums of money into a wide range of human service programs, including education. However, program evaluation has an interesting history that predates by at least 150 years the explosion of evaluation during the era of President Johnson's Great Society and the emergence of evaluation as a maturing profession since the sixties. A definitive history of program evaluation has yet to be written and in the space available to us we can do little more than offer a modest outline, broad brush strokes of the landscape that constitutes that history. It is important that people interested in the conceptualization of evaluation are aware of the field's roots and origins. Such an awareness of the history of program evaluation should lead to a better understanding of how and why this maturing field has developed as it did. As Boulding (1980) has observed, one of the factors that distinguishes a mature and secure profession from one that is immature and insecure is that only the former systematically records and analyzes its history. Therefore since program evaluation continues to mature as a profession, its origins and roots need to be documented.

Where to begin? For convenience we shall describe six periods in the life of program evaluation. The first is the period prior to 1900, which we call the *Age of Reform;* the second, from 1900 until 1930, we call the *Age of Efficiency and*

Testing; the third, from 1930 to 1945, may be called the *Tylerian Age;* the fourth, from 1946 to about 1957, we call the *Age of Innocence;* the fifth, from 1958 to 1972, is the *Age of Expansion* and finally the sixth, from 1973 to the present, the *Age of Professionalization.*

The Age of Reform 1800–1900

This period in the history of program evaluation encompasses the nineteenth century. It was the Industrial Revolution with all of its attendant economic and technological changes, which transformed the very structure of society. It was a period of major social changes, of cautious revisionism and reform (Pinker, 1971). It was a time of drastic change in mental health and outlook, in social life and social conscience, and in the structures of social agencies. It was when the laissez-faire philosophy of Bentham and the humanitarian philosophy of the philanthropists was heard (Thompson, 1950). It was a period marked by continued but often drawn out attempts to reform educational and social programs and agencies in both Great Britain and the United States.

In Great Britain throughout the nineteenth century there were continuing attempts to reform education, the poor laws, hospitals, orphanages, and public health. Evaluations of these social agencies and functions were informal and impressionistic in nature. Often they took the form of government-appointed commissions set up to investigate aspects of the area under consideration. For example, the Royal Commission of Inquiry into Primary Education in Ireland under the Earl of Powis, after receiving testimony and examining evidence, concluded that "the progress of the children in the national schools of Ireland is very much less than it ought to be."[1] As a remedy, the Powis Commission then recommended the adoption of a scheme known as "payment by results" already being used in England, whereby teachers' salaries would be dependent in part on the results of annual examinations in reading, spelling, writing, and arithmetic (Kellaghan & Madaus, 1982). Another example of this approach to evaluation was the 1882 Royal Commission on Small Pox and Fever Hospitals which recommended after study that infectious-disease hospitals ought to be open and free to all citizens (Pinker, 1971).

Royal commissions are still used today in Great Britain to evaluate areas of concern. A rough counterpart in the United States to these commissions are presidential commissions (for example, the President's Commission on School Finance), White House panels (e.g., the White House Panel on Non Public Education), and congressional hearings. Throughout their history royal commissions, presidential commissions and congressional hearings have served as a means of evaluating human services programs of various kinds through the examination of evidence either gathered by the Commission or presented to it in

testimony by concerned parties. However, this approach to evaluation was sometimes merely emblematic or symbolic in nature. N.J. Crisp (1982) captures the pseudo nature of such evaluations in a work of fiction when one of his characters discusses a royal commission this way: "Appoint it, feel that you've accomplished something, and forget about it, in the hope that by the time it's reported, the problem will have disappeared or been overtaken by events."[2]

In Great Britain during this period when reform programs were put in place, it was not unusual to demand yearly evaluations through a system of annual reports submitted by an inspectorate. For example, in education there were schools inspectors that visited each school annually and submitted reports on their condition and on pupil attainments (Kellaghan & Madaus, 1982). Similarly the Poor Law commissioners had a small, paid inspectorate to oversee compliance with the Poor Law Amendment Act of 1834 (Pinker, 1971). The system of maintaining an external inspectorate to examine and evaluate the work of the schools exists today in Great Britain and Ireland. In the United States, external inspectors are employed by some state and federal agencies. For example, the Occupational Safety and Health Administation (OSHA) employs inspectors to monitor health hazards in the workplace. Interestingly, the system of external inspectors as a model for evaluation has received scant attention in the evaluation literature. The educational evaluation field could benefit from a closer look at the system of formal inspectorates.

Two other developments in Great Britain during this period are worthy of note in the history of evaluation. First, during the middle of the nineteenth century a number of associations dedicated to social inquiry came into existence. These societies conducted and publicized findings on a number of social problems which were very influential in stimulating discussion (for example, Chadwick's Report on the Sanitary Condition of the Laboring Population of Great Britain in 1842 [Pinker, 1971]). Second, often in response to these private reports, bureaucracies that were established to manage social programs sometimes set up committees of enquiry. These were official, government-sponsored investigations of various social programs, such as provincial workhouses (Pinker, 1971). Both these examples are important in that they constitute the beginnings of an empirical approach to the evaluation of programs.

In the United States perhaps the earliest formal attempt to evaluate the performance of schools took place in Boston in 1845. This event is important in the history of evaluation because it began a long tradition of using pupil test scores as a principal source of data to evaluate the effectiveness of a school or instructional program. Then, at the urging of Samuel Gridley Howe, written essay examinations were introduced into the Boston grammar schools by Horace Mann and the Board of Education. Ostensibly the essay exam, modeled after those used in Europe at the time, was introduced to replace the *viva voce* or oral examinations. The latter mode of examination had become administratively awkward with