Performance Evaluation

Proven Approaches for Improving Program and Organizational Performance

INGRID J. GUERRA-LÓPEZ

PERFORMANCE EVALUATION

PERFORMANCE EVALUATION

Proven Approaches for Improving Program and Organizational Performance

INGRID GUERRA-LÓPEZ

JOSSEY-BASS
A Wiley Imprint
www.josseybass.com

Published by Jossey-Bass
A Wiley Imprint
989 Market Street, San Francisco, CA 94103-1741—www.josseybass.com

Library of Congress Cataloging-in-Publication Data

Guerra-López, Ingrid.
Performance evaluation : proven approaches for improving program and organizational performance / Ingrid Guerra-López. — 1st ed.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-7879-8883-8 (pbk.)
1. Employees—Rating of. 2. Performance standards.
3. Organizational effectiveness. I. Title.
HF5549.5.R3G84 2008
658.3'125—dc22

2007050973

Printed in the United States of America
FIRST EDITION

HB Printing 10 9 8 7 6 5 4 3 2 1
PB Printing 10 9 8 7 6 5 4 3 2 1

CONTENTS

ACKNOWLEDGMENTS

Recognition for any achievement is usually owed to a group of people rather than any one individual alone. First, I thank Jossey-Bass and, in particular, Andrew Pasternack for his interest in my potential contributions, which was the major catalyst for this work. I also thank Seth Schwartz for his graciousness and flexibility as I encountered a glitch or two along the way. Much gratitude is owed to Beverly Miller and Barbara Armentrout for their outstanding editing and proofing work.

I thank my mentor, Roger Kaufman, who continues to push me to think critically and support my ideas with evidence. A special recognition is owed to Hillary Andrei for her invaluable support in various aspects of my evaluation work. My gratitude is also owed to Rahat Sharma for her formatting support. I am grateful as well to my students, colleagues, and clients who are constant sources of ideas and reflections. The work in this book is the product of such ideas and reflections.

Finally, I am thankful to my husband, Jorge, for being once again excited, and patient, as I embarked on yet another project.

PREFACE

There are excellent books on evaluation already in print. Some focus on the history and theory of evaluation, some provide a comprehensive analysis of evaluation models and approaches, and some primarily offer insights into methodology. Although this book addresses these concepts, it is not intended to replicate or refute any other work. Rather, it is organized in such a way as to illustrate evaluation in the context of performance improvement. This work is directed at the following audiences:

- Performance improvement practitioners who seek to do evaluation well, wish to become sufficiently versed in evaluation so as to work well with evaluators, or want to integrate an evaluative perspective in all phases of performance improvement, from needs assessment to implementation, to evaluation itself
- Evaluators who seek to do systemic, performance-focused evaluations with the ultimate end of improving not only programs or solutions but also the organizations and the clients, including our shared society, whose needs they are meant to meet
- Students who want a solid conceptual grounding in evaluation and a guide for applying such concepts in courses and their own work and who will generate the future evaluation and performance improvement models, in part based on concepts presented here
- Instructors who are looking for a text that addresses important foundations of evaluation and presents the models and approaches in the context of the performance improvement field
- Clients, including those external to the organization stakeholders, who wish to become better-informed consumers of and partners in evaluation efforts

If you want to more clearly understand the impact this book has had on your thinking about evaluation, then I suggest that before reading it, you respond to the questions listed below. Once you have completed reading the book, I recommend answering each of the questions again, paying close attention to how your views have changed, if at all, as a result

of reading this book. This exercise is meant to support deeper reflection and insight about fundamental assumptions associated with evaluation and implications for practice.

- What is your current model or approach to evaluation? Use these questions to guide your response (if you don't have a current model or approach, think about what your basic assumptions about evaluation are):

  What aspects of evaluation are addressed by your theory?

  What are the main purposes, processes, or mechanisms associated with your model?

  Does your model target a specific setting? If so, what is it?

  With what other models (if you know of any) is your model most compatible? Explain how they are compatible.

  With what other models is your theory least incompatible? Explain what the incompatibilities are.

- In what ways has your model of evaluation been applied? Was it successful? What, if anything, would you have done differently? What were the biggest lessons you learned? Provide specific situations and examples.

The book is divided into four parts, beginning with an introduction to the foundations of evaluation, then proceeding to a collection of models chosen specifically for the reputation and applicability in the performance improvement field. Part Three looks at the tools and techniques that are common in various evaluation perspectives, and Part Four concludes with a look at continual improvement and the future of evaluation in performance improvement.

THE AUTHOR

Ingrid Guerra-López is an associate professor at Wayne State University in Detroit, associate research professor at the Sonora Institute of Technology in Mexico, and principal of Intelligence Gathering Systems. She publishes, teaches, consults, and conducts research in the areas of performance measurement and evaluation, assessment and analysis, and strategic alignment. She has published four other books, including *Evaluating Impact: Evaluation and Continual Improvement for Performance Improvement Practitioners,* and is coauthor of *Practical Evaluation for Educators: Finding What Works and What Doesn't.* She has published numerous chapters in books in the performance improvement field, including the 2006 *Handbook for Human Performance Technology.* Her articles have appeared as well in *Performance Improvement, Performance Improvement Quarterly, Human Resource Development Quarterly, Educational Technology, Quarterly Review of Distance Education, International Public Management Review,* and the *International Journal for Educational Reform,* among others.

PART

1

INTRODUCTION TO EVALUATION

CHAPTER

1

FOUNDATIONS OF EVALUATION

This chapter defines and describes evaluation and sets the frame for this book within the principles of performance improvement. Various kinds of evaluation, as well as some closely related processes, are differentiated from each other. The basic challenges that evaluators face are laid out, and the reason that stakeholder commitment is so important is examined. The benefits of evaluation to an organization are listed. Finally, definitions are provided for some key terms used throughout the book and in the evaluation field.

In our daily lives, we encounter decision points on an almost continuous basis: Should I do this, or should I do that? Should I go right or left? Should I take the highway or the back streets? Should I buy now or later? Should I take my umbrella today or not? Life in an organizational setting is no different: We face decisions about which programs to

sustain, which to change, and which to abandon, to name but a few organizational dilemmas. How do members of an organization go about making sound decisions? With the use of relevant, reliable, and valid data, gathered through a sound evaluation process aligned with desired long-term outcomes.

Unfortunately, these data are not always available, and if they are, many decision makers do not know they exist, or do not have access to them, or do not know how to interpret and use them to make sound decisions that lead to improved program and organizational performance. In fact, Lee Cronbach (1980) and others have argued that decisions often emerge rather than being logically and methodically made.

Effective leaders are capable of making sound decisions based on sound data, and evaluators can do much to influence the leadership decision-making process. Evaluation can provide a systematic framework that aligns stakeholders, evaluation purposes, desired results and consequences, and all evaluation activities, so that the evaluation product is a responsive and clear recipe for improving performance. This in essence allows the decision-making process to become clearer and more straightforward. Evaluation is the mechanism that provides decision makers with feedback, whether through interim reports and meetings or a final report and debriefing.

A BRIEF OVERVIEW OF EVALUATION HISTORY

Michael Scriven (1991) describes evaluation as a practice that dates back to samurai sword evaluation. Another type of evaluation was in evidence as early as 2000 B.C.: Chinese officials held civil service examinations to measure the ability of individuals applying for government positions. And Socrates included verbal evaluations as part of his instructional approach (Fitzpatrick, Sanders, & Worthen, 2004).

In response to dissatisfaction with educational and social programs, a more formal educational evaluation can be traced back to Great Britain during the 1800s, when royal commissions were sent by the government to hear testimony from the various institutions. In the 1930s, Ralph Tyler issued a call to measure goal attainment with standardized criteria (Fitzpatrick et al., 2004). And during the 1960s, Scriven and Cronbach introduced formative (used to guide developmental activities) and summative (used to determine the overall value of a program or solution) evaluation, and Stufflebeam stressed outcomes (program results) over process (program activities and resources) (Liston, 1999).

In 1963, Cronbach published an important work, "Course Improvement Through Evaluation," challenging educators to measure real learning rather than the passive mastery of facts. Moreover, he proposed the use of qualitative instruments, such as interviews and observations, to study outcomes. In the latter part of the 1960s, well-known evaluation figures such as Edward Suchman, Michael Scriven, Carol Weiss, Blaine Worthen, and James Sanders wrote the earliest texts on program evaluation.

In 1971, Daniel Stufflebeam proposed the CIPP model of evaluation, which he said would be more responsive to the needs of decision makers than earlier approaches to evaluation were. In that same year, Malcolm Provus proposed the discrepancy model of evaluation. In 1972, Scriven proposed goal-free evaluation in an effort to encourage evaluators to find unintended consequences. In 1975, Robert Stake provided responsive evaluation. In 1981, Egon Guba and Yvonna Lincoln proposed naturalistic evaluation on the basis of Stake's work, feeding the debate between qualitative and quantitative methods (Fitzpatrick et al., 2004).

All of this was occurring in the context of a movement to account for the billions of dollars the U.S. government was spending on social, health, and educational programs (Fitzpatrick et al., 2004; Patton, 1997). In order to address a demand for accountability, those responsible for programs soon began to ask evaluators for advice on program improvement. Thus, the initial purpose of program evaluation was to judge the worthiness of programs for continued funding.

When Sputnik became the catalyst for improving the U.S. position in education, which was lagging compared to other countries, educational entities in particular began to commission evaluations, partly in order to document their achievements. The need for evaluators soon grew, and government responded by funding university programs in educational research and evaluation. In the 1970s and 1980s, evaluation grew as a field, with its applications expanding beyond government and educational settings to management and other areas. Evaluations are now conducted in many different settings using a variety of perspectives and methods.

EVALUATION: PURPOSE AND DEFINITION

While some rightly say that the fundamental purpose of evaluation is the determination of the worth or merit of a program or solution (Scriven, 1967), the ultimate purpose, and value, of determining this worth is in

providing the information for making data-driven decisions that lead to improved performance of programs and organizations (Guerra-López, 2007a). The notion that evaluation's most important purpose is not to prove but to improve was originally put forward by Egon Guba when he served on the Phi Delta Kappa National Study Committee on Evaluation around 1971 (Stufflebeam, 2003). This should be the foundation for all evaluation efforts, now and in the future. Every component of an evaluation must be aligned with the organization's objectives and expectations and the decisions that will have to be made as a result of the evaluation findings. These decisions are essentially concerned with how to improve performance at all levels of the organization: internal deliverables, organizational gains, and public impact. At its core, evaluation is a simple concept:

- It compares results with expectations.
- It finds drivers and barriers to expected performance.
- It produces action plans for improving the programs and solutions being evaluated so that expected performance is achieved or maintained and organizational objectives and contributions can be realized (Guerra-López, 2007a).

Some approaches to evaluation do not focus on predetermined results or objectives, but the approach taken in this book is based on the premise of performance improvement. The underlying assumption is that organizations, whether they fully articulate this or not, expect specific results and contributions from programs and other solutions. As discussed in later chapters, this does not prevent the evaluator or performance improvement professional from employing means to help identify unanticipated results and consequences. The worth or merit of programs and solutions is then determined by whether they delivered the desired results, whether these results are worth having in the first place, and whether the benefits of these results outweigh their costs and unintended consequences.

An evaluation that asks and answers the right questions can be used not only to determine results but also to understand those results and to modify the evaluation so that it can better meet the intended objectives within the required criteria. This is useful not only to identify what went wrong or what could be better but also to identify what should be maintained. Through appreciative inquiry (Cooperrider & Srivastva, 1987), evaluation can help organizations identify what is

going right. Appreciative inquiry is a process that searches for the best in organizations in order to find opportunities for performance improvement. Here too the efforts are but a means to the end of improving performance. Although the intentions of most evaluators are just that, the language and approach used are charged with assumptions that things are going wrong. For instance, the term *problem solving* implies from the start that something is wrong. Even if this assumption is not explicit in the general evaluation questions, it makes its way into data collection efforts. Naturally the parameters of what is asked will shape the information evaluators get back and, in turn, their findings and conclusions. If we ask what is wrong, the respondents will tell us. If we ask what went right, again they will tell us. The key point is that evaluation should be as unbiased as possible. Evaluators should ask and answer the right questions, so that the data they get are indeed representative of reality.

In specific terms, before evaluators start to plan, and certainly before they collect data, they must determine why they are conducting an evaluation. Is this their initiative, or were they directed to do this work? What is the motivation for the study? What are they looking to accomplish and contribute as a result of this evaluation? Here are some general reasons for conducting an evaluation:

- To see if a solution to a problem is working, that is, delivering valued ends
- To provide feedback as part of a continual monitoring, revision, and improvement process
- To provide feedback for future funding of initiatives
- To confirm compliance with a mandate
- To satisfy legal requirements
- To determine if value was added for all stakeholders
- To hold power over resources
- To justify decisions that have already been made

Although the last two in this list are particularly driven by political agendas, in reality most reasons can be politicized; thus, it takes an insightful evaluator to recognize the feasibility of conducting an honest evaluation. An experienced evaluator will recognize, most of the

time, whether evaluation stakeholders are truly interested in using evaluation findings to improve performance or are more concerned with advancing their political interests. With careful attention to detailed planning, either goal can be made to fit a data-driven and results-oriented action approach to evaluation. But if taken too narrowly—in isolation and without proper context—each has its own narrow set of problems, blind spots, and special data generation and collection issues. Perception of the purpose of the evaluation can shape and limit the data that are observed (or not observed), collected (or not collected), and interpreted (or ignored). Thus, evaluators and stakeholders must begin the planning process with a clear articulation of what decisions must be made with the results of their findings, decisions that are linked to the overall purpose for conducting the evaluation.

PERFORMANCE IMPROVEMENT: A CONCEPTUAL FRAMEWORK

The field of performance improvement is one of continuous transition and development. It has evolved through the experience, reflection, and conceptualization of professional practitioners seeking to improve human performance in the workplace. Its immediate roots stem from instructional design and programmed instruction. Most fundamentally, it stems from B. F. Skinner and his colleagues, whose work centered on the behavior of individuals and their environment (Pershing, 2006).

The outgrowth of performance improvement (also called *human performance technology*) from programmed instruction and instructional systems design was illustrated in part by Thomas Gilbert's behavioral engineering model, which presented various categories of factors that bear on human performance: clear performance expectations, feedback, incentives, instruments, knowledge, capabilities, and internal motives, for example. This landmark model was published in Gilbert's 1978 book, *Human Competence: Engineering Worthy Performance,* and was based in large part on the work Gilbert conducted with Geary Rummler and Dale Brethower at the time. Pershing (2006) declares that Joe Harless's 1970 book, *An Ounce of Analysis Is Worth a Pound of Objectives,* also had a significant impact on the field and was well complemented by Gilbert's work. Together these works served as the basis for many researchers who have contributed to and continue to help develop the performance improvement field.

Currently the International Society for Performance Improvement, the leading professional association in the field, defines *performance improvement* as a systematic approach to improving productivity and competence, using a set of methods and procedures—and a strategy for solving problems—for realizing opportunities related to the performance of people. More specifically, it is a process of selection, analysis, design, development, implementation, and evaluation of programs to most cost-effectively influence human behavior and accomplishment. This series of steps, commonly known as the ADDIE model, is the basic model from which many proposed performance improvement evaluation models stem. Pershing (2006) summarized performance improvement as a systematic combination of three fundamental processes: performance analysis (or needs assessment), cause analysis (the process that identifies the root causes of gaps in performance), and intervention selection (selecting appropriate solutions based on the root causes of the performance gaps). These three processes can be applied to individuals, small groups, and large organizations. The proposition that evaluation of such interventions should also be at the core of these fundamental processes is presented in the final chapter of this book.

This is the context in which evaluation is seen and described in this book—not as an isolated process but rather as one of a series of processes and procedures that, when well aligned, can ensure that programs and organizations efficiently and effectively deliver valuable results.

MAKING EVALUATION HAPPEN: ENSURING STAKEHOLDERS' BUY-IN

One of the most important elements of any evaluation is its stakeholders. Before we define the stakeholders, it is worthwhile to define the term *stake.* A *stake* is essentially a claim, an interest, or a share in some endeavor and how that claim or interest might be affected by anything that is used, done, produced, or delivered. The traditional view of a stake used to be limited to the financial realm (for example, stockholders), but in fact a claim or interest can be financial, legal, or moral (Carroll, 2000). Thus, a stakeholder is any individual or group with a stake in an endeavor and can either affect or be affected by the decisions and actions of the organization.

Stakeholders can be broadly categorized as internal (owners, employees, and management) and external (customers, customers' customers, the community, suppliers, competitors, the government, and the

media, to name a few), and both categories can then be subdivided into various groups.

Not every individual within each stakeholder group has to participate directly in an evaluation; what is important is that those who participate are seen as representative by their group members. The greater the sense of stakeholder involvement and influence there is, the less likely it is that the evaluator will encounter resistance to the evaluation process and the findings.

While ideally evaluators will select stakeholders who will help define useful evaluation expectations, questions, and criteria, in fact, they realistically will be faced with stakeholders who have their own special interests or represent a powerful lobby. Although it is not particularly unusual for human beings to have their own special interests, evaluators should neutralize as much as possible the risk that the evaluation will become a manipulation tool for the special interests of one—or some—at the expense of others.

A vital challenge in working with stakeholders to help all be successful is to keep them focused on results and consequences rather than on politics of means. Single-issue politics from both within and outside organizations have a tremendous impact on defining objectives and selecting means. It is essential that evaluators learn enough about the specific political climate of a given evaluation to understand how it will affect the evaluation and the implementation of its recommendations. If evaluation recommendations are not implemented or are implemented improperly, performance probably will not improve, and the evaluation may have been conducted in vain.

THE EVALUATOR: A JOB OR A ROLE?

The term *evaluator* describes not only one profession or occupation, but also a given role at a particular time. Individuals conducting evaluation often wear many hats. They may be internal employees, members of the management team, faculty members, or consultants who have acquired interest and expertise in measurement and evaluation through education, training, or experience. In some cases, individuals arrive at this point by default and face an unexpected request to conduct an evaluation. They could be trainers who are charged with demonstrating the value of their training programs and departments. They may even be individuals who because of their status as a subject matter expert in some solution or program are also faced with demonstrating the value of their efforts.

Their common function is, or should be, an interim goal to document the results and impact achieved by a given solution: a program, a project, a tool, or the use of a resource. The final goal should be to use this information to make sound decisions and help the organization take appropriate action to improve performance at all levels.

Evaluators should be competent in some basic areas. Sanders (1979) proposed that at a minimum, evaluators should be able to

- Accurately describe the object (the evaluand) and context of that which is being evaluated
- Conceptualize the purpose and framework of the evaluation
- Derive useful evaluation questions, data requirements, and appropriate data sources
- Select the means for collecting and analyzing data
- Determine the value of the evaluand
- Effectively communicate results and recommendations to the audience
- Manage the evaluation project
- Maintain ethical standards
- Adjust to external factors influencing the evaluation
- Evaluate the evaluation

THE RELATIONSHIP TO OTHER INVESTIGATIVE PROCESSES

The results and consequences we want to accomplish are the primary drivers for deriving the useful questions of an organizational study. Another driver is the types of decisions that have to be made; in large part, they will determine what data have to be gathered and for what purpose. For instance, if decisions have to be made about what programs, interventions, and solutions should be continued, revised, or discontinued, then the data collection approach may take an evaluative perspective. That is, the data collected will be used to compare predetermined objectives with what was actually achieved. If the need is to make decisions about what results the organization should be targeting and, in turn, what types of programs, interventions, and solutions will help it get there, the data collection approach will take on a needs assessment

perspective. Notice that in both cases, results—and gaps in results—are the primary drivers.

Table 1.1 illustrates some sample questions from both perspectives that could apply to any organization in any sector. Both approaches to data collection should be systematic and designed to answer specific questions that can be used to improve performance.

Assessors and evaluators may share data collection techniques, but the types of questions they seek to answer differ. In this sense, the roles of assessor and evaluator differ in purpose or function rather than in importance and methods.

Needs assessors help create the future by providing hard and soft data for identification of performance-based, vision-aligned missions

TABLE 1.1. Unique Perspectives of Needs Assessment and Evaluation

Needs Assessment Questions	Evaluation Questions
What value-added results should we be targeting?	How much closer did we get to reaching our vision and mission?
What value-added results are we now getting?	Did we add to or subtract value from our external clients and our shared society?
Who or what is the primary client of the results and their consequences?	Which objectives in our mission did we achieve?
How do we get from current results and consequences to desired ones?	How are we doing in comparison to last quarter? Last year?
What interim results must be accomplished and when?	Which internal results targets were reached? Not reached?
What are our options?	Which implemented programs, projects, or solutions were effective?

and building-block objectives, as well as the gaps between current and desired results. In addition, they help identify the best solutions for closing these gaps and thereby ultimately reaching the organizational vision. It should be noted that asking people what they need is not a needs assessment; this simply creates a "wants list" or "wish list" without rigorous applicability (Kaufman, 2000). *Evaluators* help to determine whether they are heading toward reaching the future they set out to create during the needs assessment process. One of the primary ways they do this is by determining the effectiveness and efficiency of the implemented programs and solutions, as well as the causal factors associated with any gaps between expected and accomplished results. Measurably improving organizational and individual performance depends heavily on these two roles and processes.

Needs Assessment Questions	Evaluation Questions
What are the most effective and efficient ways for reaching our desired or required results?	How efficient are these implemented programs, projects, or solutions?
What will it cost us to reach those results?	In which of these should we continue to invest?
What will it cost us to ignore those results?	What results do we have to justify our continued programs?
How far do we have to go to reach those results?	What should we discontinue?
Which results take priority over others?	Which projects, programs, or solutions could be successful with some modifications? Is it worth it?
Where do we have the most—and least—leverage?	Did we add or subtract value from our internal clients and employees?

Source: Guerra (2003b).

Although both assessors and evaluators collect data with regard to the results of a process or activity, evaluators collect data to determine whether results match the results expected from solutions that have already been implemented—for example, new programs, new technologies, new processes, or any other means selected to help achieve objectives. Assessors, in contrast, seek to anticipate the expected return on investment of potential interventions before they are implemented by collecting data about both current results (what is) and potential results (what should be). With these data in hand, decision makers are able to choose among competing alternatives.

So how does scientific research come into the picture? Before answering this question, let us first explore the meaning of *science.* Science is based on a series of assumptions about the world—assumptions that can be true today but false tomorrow. Scientists are always testing these assumptions, ready to change them when the findings support such a change. To this end, scientists collect data about reality and consult with other sources to ensure the reliability of the data. Results are considered basic data, later subject to repeatable observations in order to confirm findings and scientific reports. Thus, we want to make decisions and take action based on what is currently known through scientific inquiry.

Research is essentially another systematic process of inquiry, with the purpose of finding, interpreting, and updating facts, events, behavior, and theories. In this sense, research skills are a basic requirement in today's world and can be applied in just about any context, whether needs assessment, evaluation, or scientific inquiry. In fact, the heart of the data collection plan is very much the same for all of these. Following are the common elements among these three inquiry processes. These are stated generically but can be made specific to investigative contexts.

1. Important decisions that must be taken by stakeholders are identified. They lead to element 2:
2. Guiding questions, purposes, or hypotheses that the inquiry process must answer or test, which are related to element 3:
3. Key variables or results that are the central focus of the questions or hypotheses.
4. When results are not directly observable, measurable and observable indicators must be identified.