

National Cooperative Highway Research Program

NCHRP Synthesis 248

**Evaluating and Measuring the
Effectiveness of Training**

A Synthesis of Highway Practice

**Transportation Research Board
National Research Council**

TRANSPORTATION RESEARCH BOARD EXECUTIVE COMMITTEE 1997

Officers

Chair

DAVID N. WORMLEY, *Dean of Engineering, Pennsylvania State University*

Vice Chair

SHARON D. BANKS, *General Manager, Alameda-Contra Costa Transit District, Oakland, California*

Executive Director

ROBERT E. SKINNER, JR., *Transportation Research Board, National Research Council*

Members

BRIAN J. L. BERRY, *Lloyd Viel Berkner Regental Professor, Bruton Center for Development Studies, University of Texas at Dallas*
LILLIAN C. BORRONE, *Director, Port Commerce Department, The Port Authority of New York and New Jersey (Past Chair, 1995)*
DAVID G. BURWELL, *President, Rails-to-Trails Conservancy*
E. DEAN CARLSON, *Secretary, Kansas Department of Transportation*
JAMES N. DENN, *Commissioner, Minnesota Department of Transportation*
JOHN W. FISHER, *Director, ATSSS Engineering Research Center, Lehigh University*
DENNIS J. FITZGERALD, *Executive Director, Capital District Transportation Authority*
DAVID R. GOODE, *Chairman, President, and CEO, Norfolk Southern Corporation*
DELON HAMPTON, *Chairman & CEO, Delon Hampton & Associates*
LESTER A. HOEL, *Hamilton Professor, University of Virginia, Department of Civil Engineering*
JAMES L. LAMMIE, *Director, Parsons Brinckerhoff, Inc.*
BRADLEY L. MALLORY, *Secretary of Transportation, Commonwealth of Pennsylvania*
ROBERT E. MARTINEZ, *Secretary of Transportation, Commonwealth of Virginia*
JEFFREY J. MCCAIG, *President and CEO, Trimac Corporation*
MARSHALL W. MOORE, *Director, North Dakota Department of Transportation*
CRAIG E. PHILIP, *President, Ingram Barge Company*
ANDREA RINIKER, *Deputy Executive Director, Port of Seattle*
JOHN M. SAMUELS, *Vice President-Operating Assets, Consolidated Rail Corporation*
WAYNE SHACKLEFORD, *Commissioner, Georgia Department of Transportation*
LESLIE STERMAN, *Executive Director of East-West Gateway Coordinating Council*
JOSEPH M. SUSSMAN, JR. *East Professor and Professor of Civil and Environmental Engineering, MIT (Past Chair, 1994)*
JAMES W. VAN LOBEN SELS, *Director, California Department of Transportation (Past Chair, 1996)*
MARTIN WACHS, *Director, University of California Transportation Center, Berkeley, California*
DAVID L. WINSTEAD, *Secretary, Maryland Department of Transportation*

MIKE ACOTT, *President, National Asphalt Pavement Association (ex officio)*
ROY A. ALLEN, *Vice President, Research and Test Department, Association of American Railroads (ex officio)*
JOE N. BALLARD, *Chief of Engineers and Commander, U.S. Army Corps of Engineers (ex officio)*
ANDREW H. CARD, JR., *President & CEO, American Automobile Manufacturers Association (ex officio)*
KELLEY S. COYNER, *Acting Administrator, Research & Special Programs Administration, U.S. Department of Transportation (ex officio)*
MORTIMER L. DOWNEY, *Chairman and President, National Railroad Passenger Corporation (ex officio)*
THOMAS M. DOWNS, *Chairman & President, National Railroad Passenger Corporation (ex officio)*
FRANCIS B. FRANCOIS, *Executive Director, American Association of State Highway and Transportation Officials (ex officio)*
DAVID GARDINER, *Assistant Administrator, Office of Policy, Planning, and Evaluation, U.S. Environmental Protection Agency (ex officio)*
JANE F. GARVEY, *U.S. Department of Transportation (ex officio)*
JOHN E. GRAYKOWSKI, *Acting Administrator, Maritime Administration, U.S. Department of Transportation (ex officio)*
GLORIA J. JEFF, *Acting Administrator, Federal Highway Administration, U.S. Department of Transportation (ex officio)*
T.R. LAKSHMANAN, *Director, Bureau of Transportation Statistics, U.S. Department of Transportation (ex officio)*
GREGORI LEBEDEV, *Acting President and CEO, American Trucking Associations, Inc. (ex officio)*
GORDON J. LINTON, *Federal Transit Administrator, U.S. Department of Transportation (ex officio)*
RICARDO MARTINEZ, *Administrator, National Highway Traffic Safety Administration (ex officio)*
WILLIAM W. MILLAR, *President, American Public Transit Association (ex officio)*
JOLENE M. MOLITORIS, *Federal Railroad Administrator, U.S. Department of Transportation (ex officio)*

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

Transportation Research Board Executive Committee Subcommittee for NCHRP

DAVID N. WORMLEY, *Pennsylvania State University (Chair)*
FRANCIS B. FRANCOIS, *American Association of State Highway and Transportation Officials*
LESTER A. HOEL, *University of Virginia*

ROBERT E. SKINNER, JR., *Transportation Research Board*
RODNEY E. SLATER, *Federal Highway Administration*
JAMES W. VAN LOBEN SELS, *California Department of Transportation*

Field of Special Projects

Project Committee SP 20-5

JON P. UNDERWOOD, *Texas Department of Transportation (Chair)*
KENNETH C. AFFERTON, *New Jersey Department of Transportation (Retired)*
GERALD L. ELLER, *Federal Highway Administration (Retired)*
JOHN J. HENRY, *Pennsylvania Transportation Institute*
GLORIA J. JEFF, *Federal Highway Administration*
C. IAN MACGILLIVRAY, *Iowa Department of Transportation*
GENE E. OFSTEAD, *Minnesota Department of Transportation*
EARL C. SHIRLEY, *Consulting Engineer*
J. RICHARD YOUNG, JR., *Mississippi Department of Transportation*
RICHARD A. MCCOMB, *Federal Highway Administration (Liaison)*
ROBERT E. SPICHER, *Transportation Research Board (Liaison)*

Program Staff

ROBERT J. REILLY, *Director, Cooperative Research Programs*
CRAWFORD F. JENCKS, *Manager, NCHRP*
DAVID B. BEAL, *Senior Program Officer*
LLOYD R. CROWTHER, *Senior Program Officer*
B. RAY DERR, *Senior Program Officer*
AMIR N. HANNA, *Senior Program Officer*
EDWARD T. HARRIGAN, *Senior Program Officer*
RONALD D. MCCREADY, *Senior Program Officer*
KENNETH S. OPIELA, *Senior Program Officer*
EILEEN P. DELANEY, *Editor*

TRB Staff for NCHRP Project 20-5

STEPHEN R. GODWIN, *Director for Studies and Information Services*
LINDA S. MASON, *Editor*

SALLY D. LIFF, *Senior Program Officer*

STEPHEN F. MAHER, *Senior Program Officer*

National Cooperative Highway Research Program

Synthesis of Highway Practice 248

Evaluating and Measuring the Effectiveness of Training

DAVID J. GIBER, Ph.D.
Linkage, Inc.
Lexington, Massachusetts

Topic Panel

JOHN L. COLEMAN, *Texas Department of Transportation*
ARNO D. GREY, *Illinois Department of Transportation*
JUDY GUERRERO, *California Department of Transportation*
LARRY JONES, *Federal Highway Administration*
PATRICIA L. LEES, *New York State Department of Transportation*
BARBARA G. MARTIN, *Montana Department of Transportation*
ELLIN REISNER, *Massachusetts Port Authority*
ROBERT H. STULL, *Pennsylvania Department of Transportation*
JAMES I. TAYLOR, *University of Notre Dame*
JON M. WILLIAMS, *Transportation Research Board*

Transportation Research Board
National Research Council

Research Sponsored by the American Association of State
Highway and Transportation Officials in Cooperation with the
Federal Highway Administration

NATIONAL ACADEMY PRESS
Washington, D.C. 1997

Subject Areas
Planning and Administration

Systematic, well-designed research provides the most effective approach to the solution of many problems facing highway administrators and engineers. Often, highway problems are of local interest and can best be studied by highway departments individually or in cooperation with their state universities and others. However, the accelerating growth of highway transportation develops increasingly complex problems of wide interest to highway authorities. These problems are best studied through a coordinated program of cooperative research.

In recognition of these needs, the highway administrators of the American Association of State Highway and Transportation Officials initiated in 1962 an objective national highway research program employing modern scientific techniques. This program is supported on a continuing basis by funds from participating member states of the Association and it receives the full cooperation and support of the Federal Highway Administration, United States Department of Transportation.

The Transportation Research Board of the National Research Council was requested by the Association to administer the research program because of the Board's recognized objectivity and understanding of modern research practices. The Board is uniquely suited for this purpose as it maintains an extensive committee structure from which authorities on any highway transportation subject may be drawn; it possesses avenues of communication and cooperation with federal, state, and local governmental agencies, universities, and industry; its relationship to the National Research Council is an insurance of objectivity; it maintains a full-time research correlation staff of specialists in highway transportation matters to bring the findings of research directly to those who are in a position to use them.

The program is developed on the basis of research needs identified by chief administrators of the highway and transportation departments and by committees of AASHTO. Each year, specific areas of research needs to be included in the program are proposed to the National Research Council and the Board by the American Association of State Highway and Transportation Officials. Research projects to fulfill these needs are defined by the Board, and qualified research agencies are selected from those that have submitted proposals. Administration and surveillance of research contracts are the responsibilities of the National Research Council and the Transportation Research Board.

The needs for highway research are many, and the National Cooperative Highway Research Program can make significant contributions to the solution of highway transportation problems of mutual concern to many responsible groups. The program, however, is intended to complement rather than to substitute for or duplicate other highway research programs.

Project 20-5 FY 1992 (Topic 26-09)

ISSN 0547-5570

ISBN 0-309-60104-0

Library of Congress Catalog Card No. 97-69822

© 1997 Transportation Research Board

Price \$15.00

NOTICE

The project that is the subject of this report was a part of the National Cooperative Highway Research Program conducted by the Transportation Research Board with the approval of the Governing Board of the National Research Council. Such approval reflects the Governing Board's judgment that the program concerned is of national importance and appropriate with respect to both the purposes and resources of the National Research Council.

The members of the technical committee selected to monitor this project and to review this report were chosen for recognized scholarly competence and with due consideration for the balance of disciplines appropriate to the project. The opinions and conclusions expressed or implied are those of the research agency that performed the research, and, while they have been accepted as appropriate by the technical committee, they are not necessarily those of the Transportation Research Board, the National Research Council, the American Association of State Highway and Transportation Officials, or the Federal Highway Administration of the U.S. Department of Transportation.

Each report is reviewed and accepted for publication by the technical committee according to procedures established and monitored by the Transportation Research Board Executive Committee and the Governing Board of the National Research Council.

The National Research Council was established by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and of advising the Federal Government. The Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in the conduct of their services to the government, the public, and the scientific and engineering communities. It is administered jointly by both Academies and the Institute of Medicine. The National Academy of Engineering and the Institute of Medicine were established in 1964 and 1970, respectively, under the charter of the National Academy of Sciences.

The Transportation Research Board evolved in 1974 from the Highway Research Board, which was established in 1920. The TRB incorporates all former HRB activities and also performs additional functions under a broader scope involving all modes of transportation and the interactions of transportation with society.

Published reports of the

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

are available from:

Transportation Research Board
National Research Council
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

and can be ordered through the Internet at:

<http://www.nas.edu/trb/index.html>

Printed in the United States of America

NOTE: The Transportation Research Board, the National Research Council, the Federal Highway Administration, the American Association of State Highway and Transportation Officials, and the individual states participating in the National Cooperative Highway Research Program do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of this report.

PREFACE

A vast storehouse of information exists on nearly every subject of concern to highway administrators and engineers. Much of this information has resulted from both research and the successful application of solutions to the problems faced by practitioners in their daily work. Because previously there has been no systematic means for compiling such useful information and making it available to the entire community, the American Association of State Highway and Transportation Officials has, through the mechanism of the National Cooperative Highway Research Program, authorized the Transportation Research Board to undertake a continuing project to search out and synthesize useful knowledge from all available sources and to prepare documented reports on current practices in the subject areas of concern.

This synthesis series reports on various practices, making specific recommendations where appropriate but without the detailed directions usually found in handbooks or design manuals. Nonetheless, these documents can serve similar purposes, for each is a compendium of the best knowledge available on those measures found to be the most successful in resolving specific problems. The extent to which these reports are useful will be tempered by the user's knowledge and experience in the particular problem area.

FOREWORD

*By Staff
Transportation
Research Board*

This synthesis will be of interest to transportation agency administrators, especially human resources development managers and training personnel, as well as to the "client" staff and functional area managers who are responsible for maintaining and improving the level of productivity and quality control within the agency. It will also be of interest to consultants and other organizations that develop training programs for transportation agencies. It presents basic information on the subject of training evaluation and describes examples of practice in several transportation agencies. The overall process for analyzing needs for training, the current evaluation models or processes, and techniques for measuring the results of training are presented.

Administrators, engineers, and researchers are continually faced with highway problems on which much information exists, either in the form of reports or in terms of undocumented experience and practice. Unfortunately, this information often is scattered and unevaluated and, as a consequence, in seeking solutions, full information on what has been learned about a problem frequently is not assembled. Costly research findings may go unused, valuable experience may be overlooked, and full consideration may not be given to available practices for solving or alleviating the problem. In an effort to correct this situation, a continuing NCHRP project, carried out by the Transportation Research Board as the research agency, has the objective of reporting on common highway problems and synthesizing available information. The synthesis reports from this endeavor constitute an NCHRP publication series in which various forms of relevant information are assembled into single, concise documents pertaining to specific highway problems or sets of closely related problems.

This report of the Transportation Research Board presents discussions of several models and techniques used both within the transportation agencies and in other business

settings for evaluating and measuring the effectiveness of training to both the individual and the agency affected. It describes the process of multilevel evaluation measures that begins with a needs analysis to determine desired outcomes of the training. This becomes more important as the training practice has evolved from the typical lecture style to more interactive participation.

To develop this synthesis in a comprehensive manner and to ensure inclusion of significant knowledge, the Board analyzed available information assembled from numerous sources, including a large number of state highway and transportation departments. A topic panel of experts in the subject area was established to guide the research in organizing and evaluating the collected data, and to review the final synthesis report.

This synthesis is an immediately useful document that records the practices that were acceptable within the limitations of the knowledge available at the time of its preparation. As the processes of advancement continue, new knowledge can be expected to be added to that now at hand.

CONTENTS

- 1 SUMMARY

- 3 CHAPTER ONE INTRODUCTION
 - Definition of the Problem, 3
 - Evaluation in Transportation Agencies, 3
 - Objectives and Scope, 3

- 5 CHAPTER TWO ANALYZING TRAINING NEEDS
 - Acting as a Consultant, 5
 - Levels of Analysis, 6
 - Competency Modeling, 7
 - Current Transportation Agency Practices in Analyzing Needs, 8
 - Overview, 8

- 9 CHAPTER THREE CURRENT EVALUATION PROCESSES
 - Introduction, 9
 - Kirkpatrick's Four-Level Model, 9
 - Brinkerhoff's Six-Stage Model, 11
 - Robinson and Robinson's Training for Impact Model, 12
 - Levels of Evaluation as Practiced by Transportation Agencies, 14
 - Overview, 15

- 16 CHAPTER FOUR MEASURING THE BUSINESS RESULTS OF TRAINING
 - Introduction, 16
 - Models for Evaluating Impact, 16
 - Overview, 19
 - Isolating the Impact of Training, 19
 - Overview, 21
 - Transfer of Training, 21
 - Overview, 21

- 22 CHAPTER FIVE EXAMPLES OF EFFECTIVE EVALUATION METHODS
 - Washington State Department of Transportation, 21
 - North Carolina Department of Transportation, 22
 - Illinois Department of Transportation, 23
 - Massachusetts Port Authority, 23
 - Noteworthy Measures Used for Evaluation, 23
 - Overview, 24

- 25 REFERENCES

26	BIBLIOGRAPHY	
27	APPENDIX A	SURVEY QUESTIONS
28	APPENDIX B	LIST OF RESPONDING TRANSPORTATION AGENCIES

ACKNOWLEDGMENTS

David J. Giber, Ph.D., Linkage, Inc., in Lexington, Massachusetts was responsible for collection of the data and preparation of the report.

Valuable assistance in the preparation of this synthesis was provided by the Topic Panel, consisting of John L. Coleman, Director of Training and Development, Texas Department of Transportation; Arno D. Grey, Manager of Technical Training, Illinois Department of Transportation; Judy Guerrero, Director of Administrative Service Center, California Department of Transportation; Larry Jones, Acting State Programs Officer, Federal Highway Administration; Patricia L. Lees, Director, Transportation Organization Development, New York State Department of Transportation; Barbara G. Martin, Chief, Organizational Development Bureau, Montana Department of Transportation; Ellin Reisner, Manager of Organizational Development, Massachusetts Port Authority; Robert H. Stull, Chief, Training Division.

Pennsylvania Department of Transportation; James I. Taylor, Associate Dean/Engineering, University of Notre Dame; Jon M. Williams, Senior Program Officer, Transportation Research Board.

This study was managed by Sally D. Liff, Senior Program Officer, who worked with the consultant, the Topic Panel, and the Project 20-5 Committee in the development and review of the report. Assistance in Topic Panel selection and project scope development was provided by Stephen F. Maher, P.E., Senior Program Officer. Linda S. Mason was responsible for editing and production.

Crawford F. Jencks, Manager, National Cooperative Highway Research Program, assisted the NCHRP 20-5 staff and the topic panel.

Information on current practice was provided by many highway and transportation agencies. Their cooperation and assistance are appreciated.

EVALUATING AND MEASURING THE EFFECTIVENESS OF TRAINING

SUMMARY

Training departments have come under pressure from upper management to provide evidence that training favorably impacts measures of business results. This pressure derives from issues that many transportation agencies are facing: need to continually retrain staff to keep skills current, the move to downsize organizations and accomplish more with fewer staff, the need for a flexible, multiskilled workforce, and the privatization of services. Traditional approaches to training have focused on the number of courses offered, or the number of people trained, as a measure of an effective training department. Evaluation of training has focused mostly on participant reactions to training, a test of what was learned in training, or an assessment of individual behavioral change after training is completed. Today, training departments must help connect organizational strategy to performance and the competence needed to drive performance. In order for training professionals to keep up with business demands, they must be proactive in identifying the performance gaps that keep an agency from meeting its goals, in aligning with staff and management in a consultative approach, and in conducting more effective evaluations of training.

A survey mailed to training directors of federal, state, and Canadian provincial transportation agencies asked respondents to describe their current needs analysis methods and training evaluation techniques. The results indicate that multiple sources of information are used to determine needs. Written comments indicated that training departments typically respond to needs from sources that assume that training will be the solution to performance problems. A new approach to identifying training needs, called performance consulting, recommends that training professionals become more proactive in identifying needs. According to this approach, problems in performance may result from a number of factors, including problems in the design of the workplace or work systems, organizational factors that inhibit the use of newly learned skills or behaviors, and the lack of specific skills and knowledge. Training can only address the last factor. By properly identifying the need and the performance problem, training professionals can better use their resources to train only when it is the appropriate intervention. Thus, both a targeted, soundly designed needs analysis, and a commitment to long-term, ongoing evaluation can help training departments become more efficient.

The survey also asked transportation agencies to report the methods of evaluation they currently use. The majority of respondents use at least one type of evaluation, the most common being a measure of employee reactions. Agencies not currently evaluating training cite a lack of resources as the reason for failure to evaluate. Agencies also collect information on learning after training, measured by administering skill tests or paper-and-pencil tests. Fewer agencies collect data concerning the application of training by employees back on the job. This level of evaluation is often conducted by observing trainees or having supervisors and coworkers assess the trainee on how often or how well new skills and behaviors are being applied to the job. The method most commonly used by transportation agencies is informal observation. The use of the different measures of evaluation in the transportation industry follows the same trend as other industries. As "higher" levels of evaluation are reached, the amount of time and resources necessary to conduct an effective evaluation increase.

The "highest" level of evaluation is the measurement of business impact as a result of training. This is the most difficult level of evaluation to attain because (1) it is difficult to identify proper measures to track the impact of training, (2) an adequate period of time must pass to realistically assess the effect of training on the business, and (3) the more time that passes, the more difficult it becomes to attribute changes in the business to training. New models of evaluation have been developed that provide methods to simplify the process for measuring business impact.

One model, developed by Robert Brinkerhoff, suggests anecdotal data may be useful for tracking the dollar impact of training. For example, a department manager might relate an incident where training of supervisors has resulted in more efficient time utilization, where work is being completed in 50 percent of the time previously required. This information alone may be enough to demonstrate impact rather than conducting a lengthy cost-benefit analysis.

Another model, the performance consultant model, recommends that training professionals focus their initial efforts on working with clients to identify indicators of business impact that will be meaningful and measurable. Also, an evaluation strategy is developed before the training is administered to ensure that all the resources and mechanisms, such as pretraining measures and posttraining measures, are in place to efficiently track change in performance.

Another model of training evaluation allows an estimate of return on investment. This model describes data collection methods to link dollar values to participant reactions to training, the amount of learning, and specific behavioral changes.

The results of the survey indicate that training departments in transportation agencies recognize the need to approach training and needs analysis from a new direction in order to achieve the goals of providing efficient and effective training. The synthesis provides some examples of evaluation methodologies from state, federal, and provincial transportation agencies.

INTRODUCTION

DEFINITION OF THE PROBLEM

Federal, state, and Canadian provincial transportation agencies, like other organizations throughout today's world, struggle with issues such as a continual need for retraining employees and updating skills, a need for a flexible and multiskilled workforce, and pressures to achieve greater cost efficiency. The focus of training departments is shifting from fulfilling requests for training courses, to being more proactive in addressing business needs and supporting strategies to enhance productivity. As a result, trainers are being held accountable for results from training and for demonstrating that training does, in fact, impact business results.

Traditionally, training departments have approached training by developing courses, publishing extensive catalogs of courses, and emphasizing the number of employees who pass through the courses. Robinson and Robinson (1) have termed this approach "Training for Activity." Using this approach, evaluation relied primarily on the reactions of participants, using questionnaires (sometimes referred to as "smile sheets") to evaluate a trainer's performance.

More recently, however, training departments are being pressured to produce measurable results from their training programs. A recent Transportation Research Board (TRB) publication, National Cooperative Highway Research Program (NCHRP) Synthesis 188: *Management Training and Development Programs*, described how, within transportation agencies, training is emerging as "an integral part of the strategy mission of many agencies" (2). Similarly, Phillips reports that training and development departments are being challenged by upper management to provide profit contributions and programs that produce results (3). Robinson and Robinson call this new approach "Training for Impact" (1). In essence, training departments are now becoming organizational development consultants whose focus expands to include performance management in addition to training. In performance management, Human Resource Development (HRD) professionals focus on developing the performance of employees.

The problem that exists in many organizations is that the processes for evaluating the impact of training on business goals are not in place. Trainers and HRD professionals have difficulty identifying the appropriate measures to determine impact and lack the tools to demonstrate that changes in individual and business performance result from training. This synthesis describes methods currently used by evaluation experts to determine the impact of training.

EVALUATION IN TRANSPORTATION AGENCIES

As part of the research conducted to compile this synthesis, a survey was mailed to transportation agencies throughout the

United States and Canada. Thirty-seven agencies responded to the survey, which asked questions concerning the processes for conducting needs analyses, the types of evaluation processes currently in use, and the funding allocated for training and evaluation.

Results of the survey items related to funding reveal that transportation agencies vary with respect to the amount of funding they receive to conduct training. Of the responding agencies, 15 received an increase in funding, 13 agencies were receiving less funding, and four experienced no change from the previous fiscal year. The average training budget was \$1,410,000 with a range from \$20,000 to \$15,390,000. In addition, only five agencies reported having a specific budget allocated to evaluation, the average being \$20,000. In follow-up discussions with various transportation agencies it was determined that, in addition to stagnant or decreasing training budgets for many agencies, training departments are experiencing cutbacks in staff and resources.

This trend differs from the national trend of training budget allocations that was reported in a 1993 industry report of training budgets by *Training* magazine (4). According to the article, training budgets increased steadily from 1990 to 1993. In 1993, an estimated \$48.2 billion dollars was spent on training, with 72 percent of that amount going toward training staff salaries.

Changes in the agencies themselves are creating more pressure for training departments: agencies are downsizing, the nature of the workforce is changing, and there is a greater need for professionals with multiple skills and flexibility. NCHRP Synthesis 163: *Innovative Strategies for Upgrading Personnel in State Transportation Departments*, reported that the fastest growing jobs in agencies will be professional positions (5). Furthermore, the current workforce is aging and turnover is high among professionals at higher levels. The need exists to effectively train entry-level professionals who can keep pace with the changing environment and can be groomed for succession through a professional career within the transportation agency.

Thus, the challenge of today's training department is to provide training that meets the needs of the swiftly changing business climate and provides this training within resource or budget restrictions. An HRD professional will have to use resources wisely and identify those programs that most benefit the agency. To accomplish this goal, the HRD professional will need effective methods for evaluating training programs and deciding which will successfully address the needs of the organization.

OBJECTIVES AND SCOPE

The purpose of this synthesis is to examine the processes used to evaluate the effectiveness of training programs. In

examining evaluation, the synthesis follows the evaluation process from the beginning, when the need for training is first identified, to the end of the process, when the impact of training on business results is measured. The objective is to provide the reader with a comprehensive methodology for program evaluation based on current theory and practice. Also, the synthesis suggests measures for improving evaluation and identifies pitfalls to avoid when developing evaluation strategies.

Chapter 2 addresses the needs analysis process, which is the starting point of evaluation. However, the focus should be on assessing performance needs, not just training needs. The most critical work is performed at this step to ensure a successful evaluation. The synthesis describes a new role for trainers, that of performance consultant. The critical tasks in this new role are identifying the correct needs, partnering with a client, linking training objectives to organizational strategy goals, and obtaining buy-in from stakeholders.

Chapter 3 details several models of evaluation that begin with needs analysis and continue through the evaluation of business impact. Evaluation of training has generally followed a four-level model of evaluation first developed by Kirkpatrick, where level 1 measures the reactions of participants to a program, level 2 measures the amount of learning resulting

from training, level 3 evaluates how behavior on the job has changed as a result of training, and level 4 evaluates the impact of training on business goals (6). Discussions throughout the synthesis refer to this model.

Chapter 4 examines methods for evaluating the impact of training. Several models are summarized that describe how to assess impact. Chapter 4 also lists several procedures for isolating the true effectiveness of training; that is, being confident that changes in performance or the level of business effectiveness can be attributed to training. A brief review of a model that addresses the factors that impact the transfer of training to the job is also presented.

Chapter 5 presents examples of evaluation practices seen in transportation agencies. In reporting these practices, the synthesis describes three rules of thumb that make the practices effective.

Upon reading the synthesis, the reader should have a better understanding of the issues that surround evaluation of training, especially at the business level. The HRD professional should be in a position to begin formulating a more successful evaluation strategy. This synthesis also provides a bibliography of resources that provide more detail on the models described and direct the reader to leading texts on measurement, evaluation, and training.

ANALYZING TRAINING NEEDS

A rigorous needs assessment is a valuable method of determining what gaps exist between current performance and desired performance, how these gaps impact the goals of the business, and what intervention will best close or narrow the gaps. An effective needs analysis looks beyond the request for training and asks what affects performance besides a lack of knowledge or skill. The HRD professional involves members of the agency who will be impacted by the change in performance in addition to the person who initially requested training. This approach to needs analysis, in essence, changes the role of the trainer to that of a performance consultant. The assessment must help the internal consultant to determine if training is the right intervention to solve a specific performance problem.

The needs assessment process also enables HRD professionals to establish guidelines for their evaluation strategy prior to the implementation of a training program. More specifically, the needs assessment can (1) provide baseline measures of individual, group, and organizational performance, (2) determine the purpose and goals of the evaluation, and (3) identify the procedures to be used for evaluating the training program.

This chapter provides an overview of the process that results from changing the role of trainers and HRD professionals. This change in perspective has been labeled "Performance Consulting" by Robinson and Robinson (7). The process of performance consulting describes both the consultative approach of the people involved in information gathering and the levels of analysis used to support an effective needs analysis. As an example of an information gathering approach, this chapter describes a popular method of workforce assessment, competency modeling, which is used to identify performance gaps. The chapter concludes with a summary of needs assessment practices used by surveyed transportation agencies.

ACTING AS A CONSULTANT

Traditionally, trainers have had a reactive role in organizations. They have responded to needs based on the assumed learning needs of employees. Trainers did not question the requests of line managers. In their new role as performance consultants, HRD professionals form partnerships with management to identify the interventions required to achieve superior performance. Thus, the focus becomes performance management, not training. The key becomes identifying what people must do (i.e., performance) rather than what they must learn. One difference with this approach is that the performance consultant works with others to identify solutions to the performance problem that may not include training. To be successful, the performance consultant must partner with numerous people within the agency to drive change in performance and question basic assumptions about what needs exist.

Partnering With Others

The first step in performance consulting is identifying stakeholders, clients, and sponsors associated with the performance problem. Identifying these people at the start helps ensure that everyone impacted by the performance problem is included in decisions made. This, in turn, increases buy-in and strengthens communication channels.

The person making initial contact with the performance consultant, referred to as "the contact," may or may not be the person who is directly impacted by the performance problem. The performance consultant's first task is to determine who is the client; that is, who is the person with a need that is not being filled. This is the consultant's "true" client. By identifying this person or persons, the consultant forms a direct channel to the person who understands the need and can provide business measures that are impacted by a successful change in performance.

The client, however, may not be in a position to make decisions concerning actions required to change performance. Therefore, it is also important to identify a sponsor. The sponsor becomes the highest-level advocate of the proposed intervention in the transportation agency. This person provides the power to assign resources to the program and makes ultimate decisions concerning the program. Identifying the sponsor ensures that the needs analysis and the resulting intervention are valued by the organization.

Another group of people to identify and include are those who will be directly impacted by the changes that the intervention produces (i.e., stakeholders). The stakeholders may be external customers or may be other departments within the agency. Stakeholders include the participants of the program and the participants' supervisors. Accurate identification of the participants helps guarantee that the program will meet their needs as well as their level of understanding and ability. By involving the participants' supervisors, performance consultants obtain their buy-in and can use them as sources of data for determining needs, identifying business metrics associated with the needs, and determining the extent to which performance has changed after the intervention.

It is critical to identify the client, stakeholders, and possibly the sponsor before conducting the needs analysis because they all have information that contributes to identifying the true need, which is the next step in the needs analysis.

Questioning Basic Assumptions

To identify the true need behind a performance problem, the performance consultant must question basic assumptions. That is, the consultant cannot simply assume that the need

identified by the contact, or client, is the true need that must be fulfilled to change performance.

The decision to deliver training is often based on the belief that training will fix most performance problems that exist within an organization. The role of trainer was to reply to the request with a customized or an existing program that met the stated need of the client. The trainer assumed training would fulfill the need. In the performance consulting role, the HRD professional questions that assumption. When approached with a request for training, the performance consultant must question not only if the need identified by the client is the correct need, but also if training is the best solution.

Robinson and Robinson have identified four levels of needs that exist in defining a performance problem: business needs, performance needs, training needs, and work environment needs (7). This categorization of needs is not new. Mager and Pipe advocated a structured approach to identifying performance problems and stressed the importance of differentiating between knowledge and skill problems versus environmental problems that affect performance (8).

Business needs are driven by the goals of the organization. They are the measures or indicators that define success for the organization (or department). Examples of a business need include a department's desire to attain a customer satisfaction level of 95 percent, or an agency's realization that it must update its staff on the latest bridge construction techniques.

Performance is the next level of need. A performance need refers to employees' behaviors, or the way they perform their job. A performance need would indicate that behavior has to change for employees to be more effective.

The last level of need contains both training needs and work environment needs. A training need exists when employees must learn a specific knowledge, skill, or process in order to be effective. A work environment need refers to a process or internal system that must change to allow employees to perform successfully. These last two needs are combined because both needs must be addressed to achieve the performance needs (although an intervention may not require change in both types of needs).

The levels of needs are represented by a set of nested boxes (Figure 1). Business needs are located in the largest outer box because they are the needs that drive the success of the business. All other needs impact business needs either directly

(i.e., performance needs) or indirectly (training needs and work environment needs impact business needs through performance needs). The purpose of questioning assumptions, then, is to help the client identify the true need and then link the true need to a business need. That is, once the true cause for a performance problem has been identified, the role of the performance consultant is to link the change in performance to a business need. By doing this, the client and the performance consultant understand how the business will be impacted by closing a performance gap. Also, the performance consultant and client will be better able to identify indicators to measure the business impact of an intervention. As discussed in chapter 4, identifying proper organizational indicators is a crucial step in evaluating the business impact of a training program or human resource intervention.

LEVELS OF ANALYSIS

Analyzing needs typically requires gathering information from several levels of the job: organizational level, task level, and person level (9). Combining information from all three levels shapes a comprehensive picture of what is needed to close performance gaps and can answer questions such as: what knowledge and skills are required to support the goals, what organizational barriers exist that may inhibit transfer of training, and what are the most appropriate indicators to measure business impact?

The purpose of the organizational analysis is to identify the goals and mission of the organization and the critical operational measures by which performance against those goals is assessed. The ultimate goal of closing a performance gap is to increase the effectiveness of the organization; thus, identifying the specific goals of the organization ensures the proper intervention is implemented. For example, a goal may be to increase the responsiveness to customers as measured by response time to transportation problems.

Organizational Analysis

The organizational analysis provides information crucial to a successful evaluation. As part of the organizational analysis,

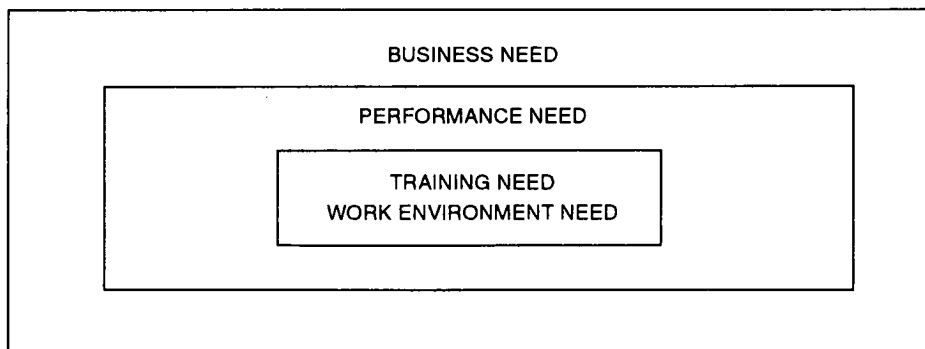


FIGURE 1 Robinson and Robinson's level of needs.

the performance consultant collects data that describe the resources available for training. This could include identifying personnel, physical space, and equipment required. Encompassing all of these is the budget allocated to training. Since transportation agency budgets are usually allocated on an annual basis, careful tracking of expenses for training (and evaluation) is required to determine if a training program, or other intervention, is cost-effective.

Another focus of the organizational analysis is determining the educational climate of the organization; that is, how is training perceived within the agency? Survey data collected for this synthesis indicate that financial resources available to trainers (especially for evaluation) have generally remained stable, or declined, as detailed in chapter 1. This may indicate that training functions are not the most highly valued functions within transportation agencies. A climate that is not supportive of training may not provide the necessary reinforcements after training to help employees practice and transfer their new skills to their jobs. Identifying an unsupportive climate indicates a work environment need that a training intervention will not solve.

Likewise, it is important to identify environmental constraints or supports that would impact how well learning transfers from the training medium back to the job. Agencies may need to create career ladders to reinforce the use of these skills and learning on the job. This may include systems within the transportation agency that may conflict with newly learned behaviors. For example, employees are being trained to work in teams, yet a compensation system provides bonuses to employees on an individual basis. On the other hand, a system that pairs up a new trainee with someone who has already been through training may facilitate transfer of new skills to the job. Identifying these barriers and supports to transfer of training prior to the program implementation would improve the effectiveness of the training and save training dollars. (Issues of transfer of training are discussed in more detail in chapter 4).

Task Analysis

The next level of analysis is task analysis. Traditionally, this information has been gathered through surveys and questionnaires to determine what employees do in their position. Results of this analysis indicate what knowledge and skills are required for successful performance on the job, thus helping to identify training needs.

A task analysis also provides information on the situational constraints of a position, according to Ostroff and Ford (10). Situational constraints include limitations of equipment, technology, or resources that impact the performance of an employee. In this context, the task analysis provides information on the work environment needs.

Person Analysis

The third level of analysis is the person level, which focuses on the level of skill or knowledge of the individual performer. The goal of this analysis level is to determine if a

training need exists by comparing the level of skill or knowledge currently held by employees against the levels required for effective performance. Methods of data collection include interviews, focus groups, direct observation, and the use of questionnaires. Robinson and Robinson prefer the use of questionnaires for several reasons, including their ability to (1) provide quantitative data that are easily tabulated, (2) reach a small or a large number of people, and (3) present questions in a consistent manner to all respondents (7).

An important property of the person level of analysis is the ability to obtain a baseline measure of performance prior to the training intervention. Basarab and Root state that an evaluation of the business impact of an intervention cannot be successful without first establishing a level of performance prior to the intervention (11). By collecting baseline information, in combination with measuring both a control group (a group that is not exposed to the intervention) and an experimental group (a group that is exposed to the intervention), a desired change in the level of performance of the experimental group after the intervention indicates that the intervention was effective at changing behavior or performance. This level of analysis is important in assuring that the training was matched to the right audience, delivered at the right level, and focused on the right objectives.

COMPETENCY MODELING

An approach to performance management that has recently become very popular is competency modeling. Developed in the 1970s by David McClelland, competency modeling is a method of identifying patterns of behaviors, beliefs, values, traits, and skills that characterize high performers in an organization (12). The advantages of competency modeling include the ability to (1) form a common language of performance requirements across similar positions or even across an entire organization, (2) identify systems and structures that support or inhibit competent action, and (3) build an architecture of HRD applications based on the competency model, including needs assessment, identification of developmental opportunities, and performance management.

In the context of needs assessment, competency modeling provides the performance consultant with a tool for collecting information for all the levels of needs identified above by Robinson and Robinson (7). During the development of the model, senior level executives and managers are interviewed to identify the goals and missions of the organization. Managers are also asked to identify high performers who exemplify successful performance that is aligned with the goals of the organization. In this way, the implications of organizational strategies for the skills and training required of individuals and teams become clear.

The next step of the model building process consists of interviewing the high performers to identify the behaviors, beliefs, values, traits, and skills that make them high performers. By interviewing multiple high performers, a pattern of the beliefs, values, and other factors emerges that describes the competencies necessary for high performance. The interview

process also elicits information on cultural barriers and facilitators to high performance.

Once a model has been identified, the process then allows individuals to be assessed against the model of high performance. This assessment, equivalent to a person-level analysis, allows a performance consultant and his or her client to identify where gaps exist in performance and helps target specific performance needs of the employees. Thus, the modeling process identifies needs at all levels and can be used to create a career development system as well as training to establish an integrated system for human resources.

CURRENT TRANSPORTATION AGENCY PRACTICES IN ASSESSING NEEDS

In a survey sent to transportation agencies in the United States and Canada, respondents were asked to indicate what sources of information they used in assessing training needs in their agency. Respondents were allowed to select more than one option. The results appear in Table 1. The data indicate that training needs are most often based on management directives. Needs are also frequently assessed by determining employee performance gaps and are aligned with the mission and vision of the agency. Nine agencies used competency modeling as a tool for assessing needs, including Departments of Transportation from Arizona, New York, Delaware, California, Florida, North Carolina, and the District of Columbia, as well as the Federal Highway Association (FHWA), and the province of Alberta.

Agencies were also asked to name sources of information for needs analysis that did not appear on the list. Regulations and statutes comprised the most common source of information

from written responses (by five agencies). Other responses included safety concerns of the agency, employee development plans, directives from training committees and training coordinators, and informal needs assessment.

TABLE 1
SOURCES OF INFORMATION USED IN THE NEEDS ASSESSMENT PROCESS

Source of Information	Number of Responses
Management directives	35
Employee performance gaps	29
Organizational mission/vision	29
Employee opinions and/or attitudes	23
Strategic goals	22
Customer satisfaction data	12
Competency model(s)	9
Other	11

OVERVIEW

The needs assessment process provides important information to the performance consultant on a variety of issues. The consultant is able to identify who the "players" are for a given request for performance management. Also, the consultant has a process for determining the true need of the client. The needs assessment process also provides the performance consultant with information that will lay the foundation for a successful evaluation. The next chapter discusses specific models of evaluation that begin with information obtained during the needs analysis and continues through evaluation of the impact of training on the agency and its business.

CURRENT EVALUATION PROCESSES

INTRODUCTION

This chapter introduces three types of evaluation models, then describes and compares them in more detail.

Evaluation of training has traditionally revolved around a model developed by Kirkpatrick (13–16). His model of evaluation, which consists of four levels for measuring the effectiveness of training, focuses on evaluating training after a program has been conducted and participants have completed it. A more recent model of evaluation, by Brinkerhoff, includes not only steps for evaluating impact of training but also stresses the importance of evaluating the training method itself, and the process used for identifying needs and goals (17). His model contains six stages. The difference between the two models is their focus. Kirkpatrick's model is considered a summative model of evaluation whereas Brinkerhoff's model includes both summative and formative evaluation.

Scriven coined the terms formative evaluation and summative evaluation (18). Formative evaluation occurs prior to and during implementation of a program. The purpose is to gather information about program operation in order to improve the training process. This includes conducting a proper needs analysis, identifying the correct objectives for a training program, determining the most effective method of training, and targeting the correct people to receive training. Summative evaluation occurs after a program has been completed. Its focus is to determine how effective the program was at meeting its objectives.

A third model of evaluation, by Robinson and Robinson, incorporates the levels of summative evaluation, as described by Kirkpatrick and Brinkerhoff, but also includes important steps for creating a consultative partnership (1).

In addition to describing each of the models and explaining the types of information collected at each stage within the models, survey results are presented that illustrate what methods of evaluation transportation agencies are currently using.

KIRKPATRICK'S FOUR-LEVEL MODEL

Kirkpatrick's model of evaluation is the most familiar and most commonly used model for training evaluation. It contains four levels at which data can be collected to measure the effectiveness of training (Figure 2):

1. Reaction, which captures participant's reactions to the program,
2. Learning, which measures how much a participant has learned in the program,

3. Behavior, which measures behavioral change due to the program, and
4. Results, which measures the impact the program has had on business indicators.

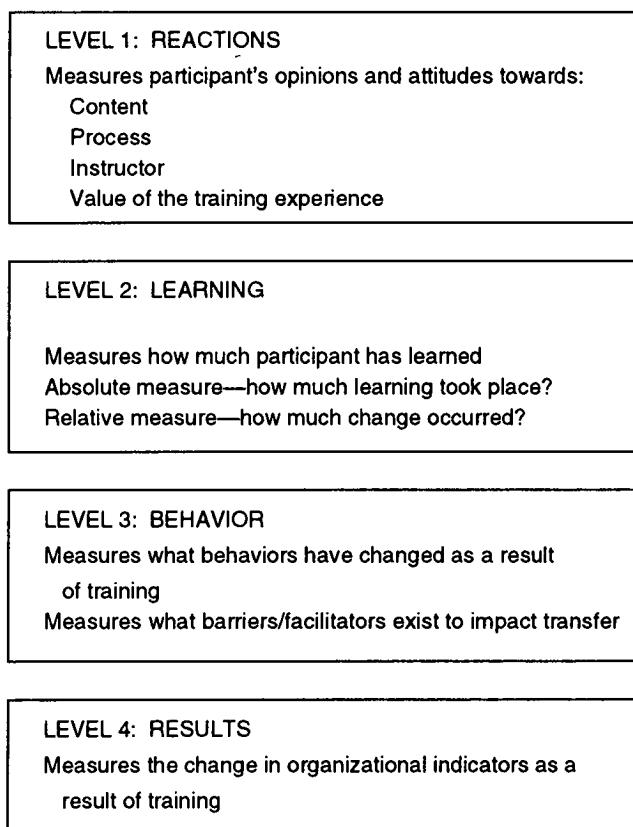


FIGURE 2 Kirkpatrick's (1959, 1960) four-level model of evaluation.

Level 1: Reactions

Information gathered at the "reaction" level of evaluation captures participants' opinions and attitudes toward the program including content, process, instructors, and facilities. Data are typically collected using survey forms that have both open-ended questions and rating scales. Information is most often gathered at the end of the program before participants leave the training environment.

Reaction level data were gathered by more transportation agencies responding to the survey than any other level of data, as shown in Table 2. The ease of data collection makes this level popular. Reactions to programs are often used as a tool to

TABLE 2
LEVELS OF EVALUATION USED BY RESPONDING
TRANSPORTATION AGENCIES

Level of Evaluation	Number of Responses (out of 37)
Level 1: Reactions	35
Level 2: Learning	18
Level 3: Behavior	17
Level 4: Results	12

make decisions about programs in terms of allocation of resources (i.e., which programs should be kept or revised). Also, reaction data provide performance information for the deliverers of training. One state department of transportation uses reaction level data to evaluate training conducted by outside contractors. This information contributes to decisions concerning continued use of the contractors.

Kirkpatrick provides several guidelines for performing effective evaluations at the reaction level (19). Guidelines include the following:

1. Determine what you want to find out—decide if there are specific aspects of the training program you need to investigate.

Time constraints and space on a form may limit the ability to obtain reactions to all aspects of the course.

2. Design a form that will quantify reactions. Although qualitative information is useful for providing rich detail, an effective evaluation form should contain items using a rating scale that allows quick, efficient collection of data. Time is saved both collecting data (participants are often eager to leave at the end of a program) and summarizing results (deliverers of training want quick feedback on their performance). Use of a quantitative scale also allows comparisons across different programs.
3. Encourage written comments. Rating scales provide some, but not all, of the reaction information. Providing a comment section allows participants to write about issues that might not have been included in other items. Written comments may also provide explanations for exceptionally poor or high ratings.
4. Get 100 percent response immediately. Allow participants the opportunity to complete reaction forms before leaving the program. Announcing at the beginning of a program that evaluations will be collected allows participants to reflect on important issues as they arise during the course of the program.
5. Get honest responses. The honesty of participants' responses is sometimes compromised if they are asked to sign the form or place their name on the form. Including a name allows the deliverer to measure attendance or to follow-up with participants for more information. However, these reasons for asking participants to include their names should not outweigh the importance of getting honest responses. An important guideline is to keep reaction forms anonymous.

Level 2: Learning

This level of evaluation measures how much the participant has learned from the program. Learning evaluations usually take the form of a paper and pencil test (in which participants must answer questions concerning the knowledge or skill they have acquired) or a skills test (which has the participant perform an activity that demonstrates their ability).

Level 2 evaluations are sometimes omitted because evaluators fear potential legal repercussions from using tests. These repercussions, generally seen as lawsuits filed by participants, may claim a test was not a valid measure of the participant's ability. The claim is most commonly seen when personnel decisions, such as promotions or pay raises are tied to performance in training. Also, the resources and time necessary to construct valid tests may be prohibitive.

Some alternatives exist that evaluators can use to side-step the issues described above. First, evaluators can create in-class exercises that allow the trainer to observe behavior similar to what is seen on the job. Case studies and exercises conducted near the end of the program can be used to measure learning and are less obtrusive to participants than a formal test given after the end of the program.

Another alternative is the use of a competency-based assessment (Figure 3). If an agency has developed a competency model that describes a level of performance necessary for success on the job, an assessment tool, or test, that measures the participant's performance against the model of competence makes it easier for the participant to understand how his or her performance on the test is related to performance back on the job.

No matter what form the test takes, Kirkpatrick offers several guidelines to make an evaluation at the learning level more effective (19):

1. Use a control group if practical—Compare the test results of a group of employees who have not received training to the results of the group that went through training. This provides stronger evidence that training has had an effect on the level of knowledge or skill for those who received training.
2. Obtain a baseline measure of performance prior to training—obtaining a measure of performance, skill, or knowledge before training, then comparing that measure with a repeated measure after training allows the evaluator to identify a change in performance. This change may be attributed to training. (Chapter 4 describes how change in this measure may not necessarily accrue to training, and describes methods for controlling for outside influences on measures of performance).

Level 3: Behavior

Level 3 evaluation assesses the degree to which behaviors have actually changed as a result of a training program. Although learning may have been occurring, as assessed through the level 2 evaluation, physical or cultural barriers in the agency may prevent transfer of training. Level 3 evaluations

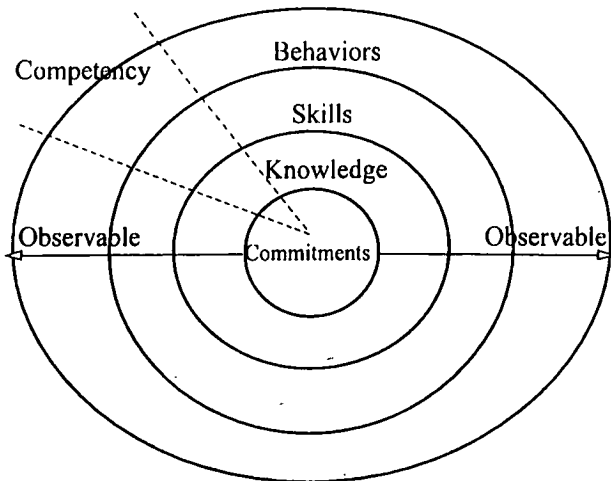


FIGURE 3 Competency model.

examine behavior, typically 3 to 6 months after a training program, to determine if transfer did occur. If transfer occurred, the evaluation may also identify factors in the agency that facilitate transfer. Likewise, noticing lack of transfer may lead to identification and removal of barriers to transfer. (Chapter 4 discusses factors that hamper and facilitate transfer of training.)

Data are often collected by surveying past participants and asking what new behaviors they have used on the job as a result of training. Also, supervisors, peers, and subordinates may be contacted to obtain others' perceptions of change in behavior. This latter method, often called a 360-degree assessment, is a popular form of evaluating training based on competencies. The assessment tool consists of a number of behavioral statements and asks the respondent to indicate if the trainee is exhibiting the new behavior, how often they exhibit the behavior, and how effective they are when using the behavior. The use of multiple sources of ratings results in a more objective assessment since different sources may have different levels of exposure to the trainee or the specified behavior.

The use of multiple sources of information is one of the guidelines that Kirkpatrick suggests for conducting more effective behavior evaluations (19). Two other suggestions include:

1. Use of control groups if practical—similar to level 2 evaluations, the use of control groups allows the evaluator to compare the behaviors of people who have not received training with those who have in order to compare the difference between the two groups.
2. Allow time for a behavioral change to take place—although learning can be demonstrated quickly, change in behavior takes more time. An immediate change in behavior could occur after training, but the important indicator of effective training is persistence of the change over time. For this reason, the gap mentioned above, 3 to 6 months, is often used to allow behavior to become consistent.

Level 4: Results

This level of evaluation measures the impact the training program has had on the organization, its key operational metrics and results. For example, has productivity increased? Has customer satisfaction increased? Has the number of accidents decreased? Often, a dollar amount of savings (or expense) is tied to the answer to these questions.

As training departments are trying to prove their "worth" to the organization, this level of evaluation has become the focus of many evaluation efforts. However, the ability to obtain clear results at this level of evaluation is difficult for three reasons: (1) evaluators do not know how to measure results or what to measure, (2) the results of this level of evaluation may not provide clear evidence that training was the cause for the change in the measure, and (3) the time span required to see a change in the business measure, if the same measure is still being used, is sometimes so long that data become elusive and meaningful interpretation of a change in the measure is difficult.

Kirkpatrick's guidelines for conducting successful level 4 evaluations are similar to those described for level 3. However, Kirkpatrick adds an important caveat to the measurement of impact that evaluators sometimes overlook: absolute evidence that training was the cause for change, or obtaining a pure measure of the amount of impact, is impossible. The best efforts of the evaluator are directed to controlling the influence of as many other factors as possible that could cause change in performance.

Because of the emphasis on measuring impact, more detail on controlling for other factors, as well as specific methods for evaluating impact, are reserved for chapter 4, which specifically describes measuring the business results of training.

BRINKERHOFF'S SIX-STAGE MODEL

Brinkerhoff's model of evaluation is based on several important tenets: First, the goal of training evaluation is to determine what impact a training program has had on the organization (17). Although the model contains many of the same methods of data collection as Kirkpatrick's model, and uses similar sources of data, the goal of all of the evaluation data is ultimately to determine how training has impacted business results. Each step contributes to this end. Thus, the second tenet of the model is that evaluation is a cyclical process (Figure 4). That is, the results of evaluating payoffs of the training (Stage 6) do not get "filed away" but, rather, are used as data for the next cycle of needs analysis for the particular program. The evaluation process is constantly providing information to other stages. Finally, Brinkerhoff maintains that financial data are not the only source of information available to demonstrate impact. Anecdotes about the success of training programs are also a rich source of information.

The first stage is an evaluation of the needs and goals of training. As with a typical needs analysis, this stage asks how great the need or problem is, and whether training is the most effective solution. Also, information is gathered to determine what other approaches to changing behavior are available and

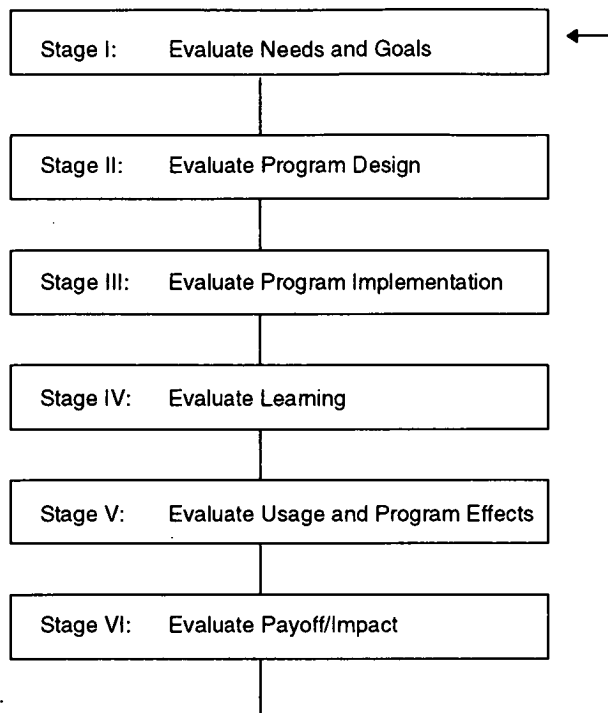


FIGURE 4 Brinkerhoff's (1987) six-stage model of evaluation.

what would be the likely payoff as a result of training. Sources of information are performance analyses, observation, surveys, and organizational audits.

The second stage evaluates the design of the training program. If training, in fact, has been identified as the best alternative for performance improvement, what design would work best? Evaluation at this stage not only asks if a program design would work, but also examines why other designs would not work. Useful data would come from literature reviews describing the effectiveness of training designs, pilot tests of current program designs, and reviews by current participants.

The third stage examines how the training program is operating. This stage is most similar to Kirkpatrick's level 1 (reaction). At this stage, evaluators determine if the program is operating as it should and, if not, identify what problems are arising. Participants are surveyed to determine what they liked and did not like. The cost of training is also determined.

The fourth stage of the model evaluates learning. The measurement of learning takes place immediately after the training program. Knowledge and performance tests, observation of participants, and perhaps work samples are used to judge if participants learned the material, how well they learned it, and exactly what they learned.

The fifth stage measures how the material from the program is being used and also if learning is enduring. The purpose of this stage is to determine if participants are keeping up with and using what they learned. Evaluation results help determine if only some of what a participant learned is being used and measure how well the knowledge or skill is being used. Methods for data collection include reports from peers, subordinates, and supervisors, observation, and an examination of work samples.

The sixth stage of Brinkerhoff's model evaluates the payoffs associated with the program. More specifically, this evaluation level answers three questions: (1) What difference does having implemented the program make? (2) Has the identified training need been met? and (3) Was it worth the cost of training? To answer these questions, some of the same measures identified for needs analysis are used to measure impact: organizational audits, surveys, performance records, and cost-benefit comparisons.

Although this is the final stage of a six-stage model, it is important to reiterate that the evaluation process does not end there. As demonstrated by the type and sources of data that are used for evaluating both stages one and six, the two stages form the ends of a loop that feeds information from the end of one evaluation process to the beginning of another.

The two models of evaluation presented above share some obvious similarities. However, the differences between the two, with Brinkerhoff's model being more recent, illustrate how the focus of evaluation has grown to include pre-implementation study as well as providing alternatives to demonstrating impact. This cyclical model ties assessment, measurement, and evaluation together into a continuous improvement effort. The next model examined focuses, again, on the formative stage of evaluation to provide a more effective summative evaluation, that is, a stronger demonstration of business impact.

ROBINSON AND ROBINSON'S TRAINING FOR IMPACT MODEL

The importance of conducting careful and extensive "up-front" work prior to implementing a training program is illustrated in Robinson and Robinson's Training for Impact model of evaluation (1). Similar to the two previous models, their model (Figure 5) includes the collection of data after training that describe reaction, learning, behavior, and operational changes resulting from training. What differs, however, is the importance of identifying a business need and clients, forming a collaborative relationship with clients, and reporting to clients. Their entire evaluation model is summarized in a 12-step process:

1. Identify a business need. Business needs can be identified through a request from a person within the organization or as a proactive effort by the HRD professional (i.e., performance consultant) to determine future needs of the business.
2. Identify and form a collaborative relationship with a client. Once a need is identified, the performance consultant identifies a client, a person with the power to make decisions and who has something at stake based on the effectiveness of the training program. Identifying the correct client is crucial to securing resources and having someone to provide appropriate business measures. Forming a collaborative relationship also emphasizes the client's accountability for the results of the program.
3. Conduct an initial project meeting. The goal of the initial project meeting is to prevent the client from immediately jumping into a training program. The role of the

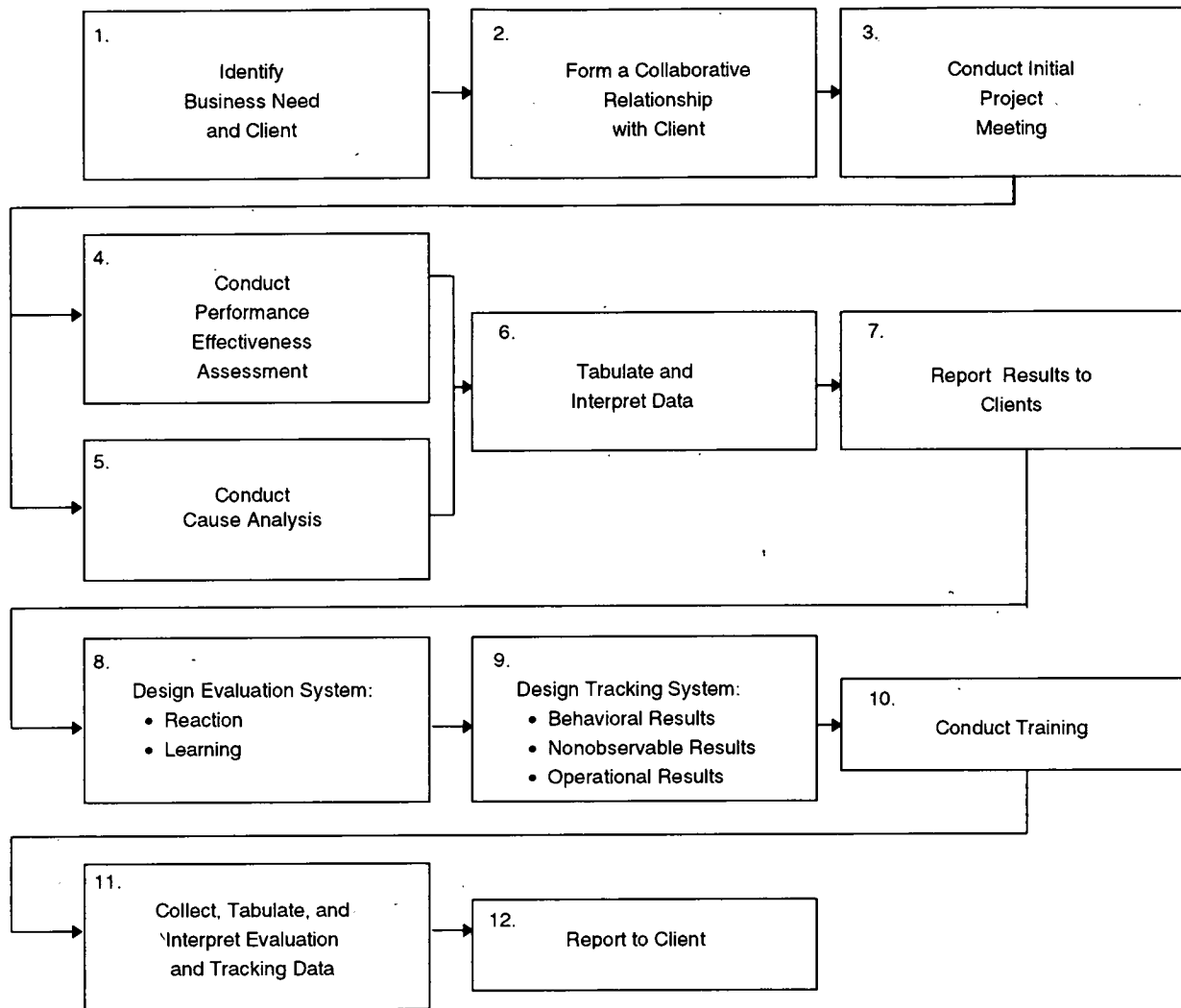


FIGURE 5 Robinson and Robinson's Training for Impact model.

performance consultant is to help the client become aware that training may not be the best solution. In this meeting, roles are determined and, ideally, a partnership is formed between the consultant and the client.

4. Conduct a performance effectiveness assessment. In essence, this is a pre-measure of knowledge, skill, or performance. This step allows the client and consultant to identify exactly where gaps exist that are causing noneffective performance. The goal is to determine the difference between what is and how things should be.
5. Conduct a cause analysis. Once performance gaps have been identified, the next step is to determine why performance is not where it should be. (This step ties to the various levels of needs presented in chapter 2).
6. Tabulate and interpret results. At this step, the consultant has obtained and summarized information to identify causes and effects of performance deficiencies.
7. Report results. Rather than provide the client with recommendations on courses of action, the consultant meets with the client to jointly discuss and determine

what the proper course of action will be. This allows the client to be involved in decision making and reduces resistance to any particular course of action. It also protects the consultant from blame resulting from unsuccessful programs since decisions were arrived at jointly.

8. Design the reaction and learning evaluation systems. This process begins before the implementation of the training program. The surveys and tests used to measure reactions and learning should be designed as the course is being designed to allow evaluation tools to accurately measure objectives of the course.
9. Design tracking systems: behavioral, nonobservable, and operational changes (and impacts). Identification and design of these systems, prior to training, focuses the client and the consultant on the outcomes of training before training is actually implemented. This not only helps to keep the objectives of the program focused but also allows the client and consultant to determine what resources are required to conduct evaluations. Design of these systems is done by both the client and the consultant in a collaborative fashion.

10. Conduct training. Only after all nine steps of preparation are completed is training conducted. All the systems for evaluation have been identified and designed. All the appropriate measures have been identified as well.
11. Collect, tabulate, and interpret evaluation and tracking data. This step includes both pretraining and post-training measures of performance. Data are collected using the systems identified in step 9 and interpretation of data allows the consultant to determine if performance was impacted.
12. Report to the client. Finally, the results of the evaluation are presented to the client. The client is involved in final interpretation of the results and in determining what impact training had on performance and on business measures.

Robinson and Robinson's model describes a process of evaluation that asks the trainer to take on a role that goes beyond traditional training. That is, by playing a more consultative role, the trainer shifts the focus of evaluation from a summative approach to that of a formative approach and can play a greater role in affecting performance and business results. Even if all 12 steps are not followed exactly, this model addresses key issues in the role of the consultant and the use of evaluation as a continuous improvement intervention.

LEVELS OF EVALUATION AS PRACTICED BY TRANSPORTATION AGENCIES

The last section of the chapter describes some tools that transportation agencies submitted or described as part of the distributed survey. Many agencies indicated that they used at least one of Kirkpatrick's four levels of evaluation, as indicated in Table 2. (Kirkpatrick's model is used because it is the standard for evaluation of training). The section is not a comprehensive review, however, because agencies either did not provide a sample of tools, or provided samples that were similar across more than one agency. The samples are presented in order of Kirkpatrick's model.

Level 1: Reaction

Of the 37 agencies responding to the survey, 35 conduct reaction level evaluations. Both managerial and technical programs were evaluated. Items on the surveys were used to assess the effectiveness of the instructor(s), a particularly useful tool for agencies that contract instructors. Also, items assessed the usefulness of course material on the job, the format of the course, the depth of the course content, and opportunities for participants to practice learned material while still in the program.

Open-ended questions on the reaction surveys helped to provide information at other levels of evaluation. For example, one item that provides needs analysis information read, "What alternative sources of training would accomplish the objectives of this course or your objectives in attending?" This item allows

participants to suggest methods of training that might be more effective for transfer and less time consuming than current methods (e.g., on-the-job training versus classroom training). Another item asks if participants have had an opportunity to apply learning from the program. The information from this item identifies potential barriers to transferring learning to the job, an issue that arises when evaluating at level 3 (behaviors). For example, responses from this item could indicate that the environment did not support transfer, or that training was not given at the most opportune time.

Agencies that did not evaluate reactions cited lack of resources as the reason why data were not gathered. Also, some agencies were in the process of restructuring their training departments and were not currently conducting reaction level evaluations.

Level 2: Learning

Eighteen of the 37 agencies indicated that they used a learning measure of one form or another. The majority of measures are for technical training courses primarily because technical skills are easily observed and measured.

One agency uses a test to measure learning and makes personnel decisions based on the results. Test results also indicate the program's effectiveness. Equipment operators are given a written exam at the end of the program and must receive a passing grade to operate the equipment. The test is coupled with supervisory observation. The operators must demonstrate, to their supervisor's satisfaction, that they are capable of operating the equipment.

Another agency detailed how tests administered to participants were used to measure the validity of the training program. Trainers tracked scores for participants who had passed and not passed the curriculum. Statistics concerning promotion rates were also collected. The results of their analysis indicated that there was a relationship between passing the course and promotions, such that participants who passed the curriculum were promoted at a greater percentage rate than participants who did not pass the curriculum. Although a relationship exists, it is important to realize that the presence of the relationship does not indicate cause and effect: that is, there is no proof that the training caused the promotions. Factors other than success in the training may have impacted promotion decisions.

Several agencies described a pretest/posttest methodology. Learning measures are taken both before and after training to identify how much participants learned as a result of the training. This method helps isolate the effect the training has on learning. Changes in scores can be attributed to learning from the program (although it is important to recognize that other factors may also cause a change in scores).

Level 3: Behaviors

Seventeen agencies surveyed assess training programs at level 3. Behavioral assessment is more often seen in management

training partly because no simple learning measures exist to determine the amount of learning of "soft" skills that are typical of management training programs. Behavioral assessments are also common methods for evaluating competencies, or patterns of behavior, motives, traits, knowledge, and skills. Learning tools cannot assess all the components of competencies.

The most common method of data collection is through supervisory observation and feedback. Several agencies specifically described this method of data collection. Agencies collect data from participants, supervisors, upper management, and customers to determine how behavior has changed. Some methods were slightly more structured, with data being collected both during classroom sessions (through observation) and in follow-up interviews with participants and supervisors. One agency described a Proficiency Guide that lists behaviors specific to a piece of equipment, on which each participant is rated.

Other guides include lists of behaviors that are linked to performance appraisals and to course objectives. The lists provide participants with expectations and goals for their performance development that allow them to know how they will be evaluated on the course and on their job.

One department of transportation developed a method for determining baseline measures of performance. A participant's supervisor is asked to rate the participant's current level of behavior immediately following the program and is informed that he or she will be contacted again in several months to re-evaluate the participant's behavior. Each item on the survey is linked to a specific skill area and objective of the course.

Many of the questions on level 3 evaluations focus on behavioral changes that have (or have not) occurred. Another focus of this level of evaluation is the identification of barriers that may prevent learning and of supporters that facilitate learning. In conducting an evaluation of an assessment center, an agency asked participants to evaluate transfer with the following item: "The exercises in the assessment center accurately reflected current or future job demands." This item helps trainers identify other factors beyond the influence of the trainer that can impact the effectiveness of training. In this example, the timing of the training program has an impact on the ability to transfer knowledge or skill. In order for participants to practice and apply learning, they must return to a job that allows them opportunities to practice (the reader is referred to

Broad and Newstrom (20) for a comprehensive list of factors that influence transfer of training).

Level 4: Results

Due to the specific interest in evaluating at level 4, a separate chapter is devoted to issues of evaluating results and impact of training on business needs. Therefore, examples of level 4 evaluations provided by survey respondents are available in chapter 4.

Overall, the results of the survey of transportation agencies support research by Robinson and Robinson (1). They found that across many organizations, level 1 evaluations are routinely conducted. The use of level 2 evaluations, however, varies widely across organizations, depending on the types of positions that exist within the organization and the ease with which learning can be evaluated. In measuring behavioral change (level 3), transportation agencies were more active than other types of organizations. This relates to the large number of technical positions that require training and the ease of observing behavior on equipment. Transportation agencies also rely a great deal on supervisory interviews (although often informal) to capture data concerning behavioral change. The difficulty in evaluating at level 4 in transportation agencies is similar to and consistent with that of other organizations.

OVERVIEW

The models of evaluation presented in this chapter share sources of data, methods of data collection, and methods of data analysis. The trend in evaluation, however, is toward a more collaborative method of identifying needs and measures of business impact. Where evaluation once started after the training program was completed, newer models conduct significant parts of the evaluation work prior to program implementation. The ability to identify relevant measures prior to training, obtain baseline measures before training occurs, and create a collaborative relationship with the client all are steps in reaching the critical element of training evaluation: an evaluation of business impact. The next chapter details specific methods of evaluation of impact, and provides steps for isolating the true impact of training.

MEASURING THE BUSINESS RESULTS OF TRAINING

INTRODUCTION

A recent study by the American Society for Training and Development (ASTD) indicates that approximately two-thirds of training managers feel pressure to show that training programs are producing “bottom-line” results (21). However, surveys of industry show that measuring the impact of training on business results is still the least commonly used method of evaluation (1). The primary reason for this is the difficulty in conducting a proper evaluation and being able to attribute changes in performance to training. The purpose of this chapter is to simplify the process of conducting evaluations of the impact training has on business results and to describe methods for ensuring greater effectiveness of training.

The chapter describes several methods for evaluating business impact and explains how to isolate the effects of training from other potential factors that influence performance. A model for assessing factors that influence the transfer of training is also presented.

MODELS FOR EVALUATING IMPACT

The ultimate goal of most HRD program evaluations is to demonstrate that the program has had an impact on business results and has met the stated needs. Calculations of cost-benefit ratios and returns on investment (ROIs) often accompany this level of evaluation, although Phillips advocates the addition of a fifth level of evaluation that encompasses ROI calculations (22). (His method is discussed in more detail below). What follows are descriptions of evaluations of impact that correspond to the models of evaluation discussed in chapter 3.

Brinkerhoff's Stage Six: Evaluating Payoff

As detailed by Brinkerhoff, the purpose of Stage Six of program evaluation is to assess the value of an HRD program's payoffs (17). The two main questions that must be answered at this stage are: Did the program achieve its goals in addressing the stated need and, did the program achieve its goals in a cost-effective manner? To answer these questions, the HRD professional or performance consultant must carefully answer four more detailed questions outlined below.

The first question to answer is: What benefits have resulted from the training or HRD program? Part of the answer for this question emerges from the results of Brinkerhoff's Stage Five, which examined use and endurance of training. In other words, data that demonstrate that new skills or new behaviors are being performed and are lasting over time, provide a

starting point from which to identify benefits. Work done during the needs assessment (Stage One) provides the HRD professional with the expected benefits to which behaviors are linked. Also, Brinkerhoff developed a taxonomy of levels of organizational benefits that result from improved performance to help HRD professionals determine the appropriate benefits. The taxonomy includes: (1) survival of the organization, (2) profits and profitability, (3) growth and expansion, (4) employee welfare, and (5) social welfare. Each level can potentially benefit from an HRD program, and a program can benefit more than one level. Thus, by identifying which level the program impacts, the HRD professional can more easily identify direct measures of impact. Brinkerhoff reminds the HRD professional that the benefit linked to a specific behavior or skill may vary among organizations. For example, two agencies may implement a similar HRD program for the purpose of changing the behavior of managers so that they are more sensitive to employees' needs. However, one agency may see the benefit of that change to be decreased number of employee grievances (which is linked to time lost on the job to attend grievance hearings) while another agency sees the benefit to be improved employee morale (which is tied to higher productivity).

Once the benefits of HRD interventions are identified, the next question is: What is the value of each of the benefits of HRD? This question is a natural follow-up to the first question because an intervention must have some value to the organization to be considered a benefit. The goal of this question is to quantify that value.

The goal of quantifying the value of benefits of an HRD program to an organization is often identified as the most difficult part of this level of evaluation, especially for “soft” skills training. Success at this level relies on extensive work conducted prior to the design of the HRD program. Proper identification of needs and close collaboration with the client lead to better identification of benefits and to indicators that can measure the value of the program.

Brinkerhoff suggests that, in some situations, extensive efforts to determine quantifiable measures or dollar values to benefits may not be worth the trouble. For example, a program designed to increase safety may result in three lives being saved over the course of a year. The cost of delivering that program seems inconsequential when compared to the benefit of saving three lives. Also, a combination of qualitative and quantitative data might be adequate to demonstrate benefit. A supervisory training program implemented to increase efficiency of a performance appraisal system may be evaluated with quantitative data that describe how much time is saved during performance appraisals, thus providing a dollar value of time saved. These data can be supplemented with qualitative data (e.g., interviews with supervisors, direct reports) that

describe a much more efficient and useful process for performance appraisal. Finally, other sources of information could be used to acquire value estimates.

The third question in the process of evaluating the impact of an HRD intervention is: How do the benefits of the program compare to the costs of the program? The answer to this question comes easier than the answer to the previous question because costs associated with implementation are easier to track. The difficulty with this step, however, is identifying all the associated costs of a program, both direct and indirect. In addition to the costs typically associated with training, such as trainer's time, cost of materials, and participant's time, programs also entail opportunity costs. Opportunity costs include money or time spent on training that could have been used elsewhere. For example, a participant's time spent during training could include lost revenue to the agency. Works by Cascio, Spencer, and Fitz-enz provide more information on costing HRD functions (23, 24, 25).

The fourth question that Brinkerhoff addresses to help determine the impact of training is: To what extent has the initial HRD need been resolved? The reactions to a program may be positive, participants learned new skills and were able to apply them on the job, but the program may not have addressed the need. To answer this question, the HRD professional completes the cycle of Brinkerhoff's model by returning to the needs analysis conducted in Stage One, and comparing results from a new analysis to the results of the initial analysis. The answer to this question might reveal that the need was addressed and the program ended. The results might also indicate that the program must be revised because the needs were not adequately met. The results might unexpectedly demonstrate that the program impacted other areas of the business that were not initially anticipated. The answer to the fourth question ultimately leads to decisions about the program itself, which begin a new cycle of evaluation and program implementation.

Robinson and Robinson's Training for Impact Model

Robinson and Robinson describe a method for tracking the impact of HRD programs on the organization using the approach of a performance consultant (see chapter 2 for details on the role of a performance consultant) (1). The tracking of impact uses operational data to calculate the costs of the program and compares them to the benefits of the program. Identifying costs associated with a program is generally easy, although there is not a set method for doing so. Different organizations include different measures. Determining benefits is much more difficult because methods for doing so differ from situation to situation and measures are not always easy to identify. The key to successful cost-benefit measurement is close work with the client. In the cost calculation, it is crucial to reach mutual agreement with the client with respect to what costs will be included in the calculation. This, in turn, promotes buy-in from the client, and increases the chance for successful evaluations because the client has a stake in the

process. On the benefit side, the client provides information on what operational indicators will accurately measure impact of training. Robinson and Robinson also stress the importance of identifying benefits from a program that do not have a dollar value attached to them.

To effectively track the results of a training intervention, the Training for Impact model lists seven questions to ask (Table 4). The first is: Who is the client, and what is the business need for the training effort? As outlined in chapter 2, the purpose of identifying the client is to be sure that the person supporting the program has a stake in its outcome. The client will want to see positive results from the program and will work with the consultant to reap results. Also, identifying the business need at this early stage of a needs analysis ensures that the program is based on objectives that will address the needs of the business. Identifying business needs makes identification of accurate operational indicators easier as well.

TABLE 4
TRAINING FOR IMPACT MODEL

Guide Questions	Actions
1. Who is the client? What is the business need?	Results measurement; objective-based; identification of accurate business needs
2. What is the cause of the problem?	Determination of training intervention can address the problem
3. What are the operational results that the client and the performance consultant want to track?	Identification of measures linked to the problem
4. What knowledge or skills are operationally linked to the desired results?	Benchmarking from previous results; observe successful performers; pilot testing
5. What is the total cost of developing and implementing the program?	Partnering with the client to identify costs
6. What information will be used to determine whether the desired results occur?	Identification of measures by client of meaningful and credible measures; parameters for data collection
7. What is the appropriate waiting period to determine effectiveness?	Affected by seasonality, fluctuation in operational measures

The second question asks: If training is addressing a business need, what is the cause of the problem? For a training program to have impact, it must be determined that the cause of the problem can be addressed through training. A lack in an employee's level of skill or knowledge must be at least part of the problem in order for training, as an intervention, to have an impact.

The third question is: What are the operational results that the client and the performance consultant want to track? Literally hundreds of measures exist to track the operations of the business. The key to this step is identifying the measure that is linked to the identified problem. Often, this measure was used during the needs analysis and can be used again to track changes after training. In cases of some managerial training,

however, Robinson and Robinson recognize that identifying and tracking measures that are linked to training is sometimes difficult, if not impossible. In these cases, they recommend interviewing managers who are impacted by training and obtaining anecdotal information concerning impact of training. This will provide data at the managerial level that may still be translated to dollar values.

Although measures have been identified, it is also important to be sure the measures are linked to training. The fourth question is: What knowledge or skills being taught in training are causally linked to the operational results to be tracked? In other words, how can the performance consultant be sure that the knowledge or skills being trained for are the ones that will cause a change in the operational measure? Robinson and Robinson outline three methods to determine a causal relationship.

The first method for identifying causality is a review of the literature pertaining to the skill or knowledge being trained for. Previous research may demonstrate that a specific skill will impact operational results in a predictable way. This type of benchmarking is invaluable whether the examples are from transportation or from other industries.

Another method to determine causal links is to observe successful performers. By observing someone who is successful at reaching a specific operational result, such as efficiently managing a road crew, the consultant can determine what behaviors, skills, and knowledge lead to the successful performance. The observational method results in direct links between action and results. This is similar to the competency modeling approach described earlier. This model of performance can then be used to assess other individuals to determine what skills or knowledge they lack.

The third method described involves pilot testing a training program and tracking the results. In piloting, the program is administered to a small group whose performance is tracked over a short period of time, perhaps several weeks. A change in performance that does not occur for individuals who are not in the pilot program (i.e. the control group) can be strongly attributed to the program. Piloting a program also allows the performance consultant to make adjustments to the program prior to its agencywide implementation.

The fifth question to ask when tracking operational results is: What is the total cost of developing and implementing the program? As with Brinkerhoff's model, the items included in determining cost vary from situation to situation and between organizations. At this step, it is important to partner with the client to reach agreement as to what costs will be included.

Question six asks: What information will you use to determine whether the desired operational results are occurring? Front-end work becomes crucial because the client must assist the consultant in identifying the measures that will be most meaningful and credible. If possible, use measures that already exist and information that is being collected. This will make data collection easier and usually demonstrates that the organization sees that indicator as important and worthwhile.

The seventh, and final, critical question to ask for tracking operational results is: How long must you wait to determine whether the desired operational results are occurring? Typically,

waiting periods are 3 to 6 months to see change at the operational level. Sometimes, the wait is even longer. There are several reasons for this delay. First, performance may change immediately, but the full impact of this change on the organization does not become apparent for some time. Second, the organization may experience seasonal fluctuations in operational measures. Waiting to take measures increases the chance of properly attributing change in operational measures to training. To circumvent the problem of seasonal variation, Robinson and Robinson suggest taking measures of operational indicators in the same quarter as the pretraining measure, but one year later.

In the Training for Impact model, the success of tracking the impact of training relies heavily on work conducted before the training intervention. Working with the client to identify the proper needs and the proper operational indicators increases the success of the evaluation in measuring impact.

Phillip's Level 5 ROI Model

While the two previous models discussed alternatives to obtaining dollar values for business indicators to calculate return on investment (ROI), Phillips has developed a model for ROI based on Kirkpatrick's levels of evaluation (22). In his model, Phillips also advocates a fifth level of evaluation that is exclusively devoted to ROI. The purpose of adding a fifth level is to compare the costs of a program to the benefits of a program.

In discussing his model of evaluation, Phillips lists several myths of cost-benefit analysis: (1) it takes too much time, (2) it is too complicated, (3) it is inaccurate, and (4) there is no effective method of assigning a value to performance improvement. His model demonstrates that these myths do not hold. The basic tenet of Phillip's model is that ROI data need not be collected for only level 4 evaluations. ROI can also be demonstrated for other levels of evaluation.

Prior to the implementation of a program, ROI can be calculated using the process typically used for cost-benefit analysis. The costs of the program are calculated, the benefits of the program are calculated using estimates of what change in performance or business measure is anticipated, and the return on investment is determined from the ratio of costs to benefits.

ROI can also be estimated using reaction data. At the end of the program, four questions are asked of participants to estimate the potential ROI.

- 1) Participants identify what knowledge or skill has been improved.
- 2) Participants describe what actions are planned using the newly acquired skills and knowledge.
- 3) Participants describe what impact the improvement will have on their work and what the dollar impact of this change is.
- 4) Participants are asked to provide the basis for obtaining their estimates of dollar impact and their level of confidence in their estimate.

The result of this analysis is a comparison of the cost of the program with estimates of the benefits of the program. Although it may be argued that this does not provide the most accurate estimate of benefit, this method does provide a more accurate picture than estimates obtained during the pre-program ROI calculation.

Assessing ROI using Level 2 data becomes more involved because an empirical link must be demonstrated between learning data and performance. The first step in the process is the development of a test, to be administered at the end of the program, that accurately reflects the objectives of the program and is job relevant. The test could be a skills test or a list of behavioral indicators. Next, a relationship is established between the scores from the test and measures of the performance of individuals. This typically is a correlation that demonstrates a positive relationship between scores and performance. A high correlation indicates that participants scoring high on the test are also high performers on the job. Based on the relationship, it is also possible to make predictions about the level of performance of each participant given their test scores. The next step is to convert performance data to a monetary value. That is, for a specific level of performance or increase in performance, what is the dollar value of benefit associated with that level of change in performance. Finally, the predicted value of the program, based on the performance data, is compared to the cost of the program.

Level 3 measures change in behavior on the job. The following steps describe the process used to determine ROI for behavioral change using Phillip's model:

- 1) Develop competencies for the job. The competency is the combination of knowledge, skill, and behavior that leads to successful performance. A job may have multiple competencies describing successful performance.
- 2) Determine what percentage of the competencies that describe the job are being addressed in the training program.
- 3) Using salaries, employee benefits, and market value information, determine the monetary value of the competencies. That is, determine how much it costs the organization to pay someone who has the required level of competence.
- 4) Calculate the worth of skill levels of participants before and after training. Measures of skills can be obtained using surveys of behaviors. The difference between post- and pretraining scores reflect the change in performance attributable to training.
- 5) Subtract the posttraining values from the preprogram values to determine an estimate of the added value of the change in competence.
- 6) Compare the total added benefits with costs of the program.

The fourth level of evaluation, results, is the level at which ROI calculations are usually conducted and Phillips does not add new methodologies to this level.

His model does include suggestions on making the ROI calculation process more effective and efficient. Because the

time-intensive nature of some of the data collection techniques, he advocates that evaluators be selective in choosing programs for which to calculate ROI. Programs that are likely candidates for conducting Level 4 and Level 5 (ROI) evaluations are those that (1) will be run for a long period of time, (2) are important to meeting the organization's goals, (3) are higher in cost to implement and deliver, (4) are highly visible, and (5) have large target audiences.

Phillip's model also stresses the importance of obtaining management buy-in when conducting ROI calculations. Management must believe in the methods being used to calculate dollar values in order to have a useful evaluation.

OVERVIEW

Numerous models, and countless variations, exist for measuring the impact of a training program on business measures. Models are based on careful identification of needs, proper identification of the business measures that will be impacted, and estimations of dollar values associated with change in the business measures and change in performance. In-depth work performed prior to training will lead to more effective measures of impact.

ISOLATING THE IMPACT OF TRAINING

Implicit in all the models mentioned above is the need to isolate the true impact of training as the cause for change in performance. Although the operational measures of a business need may change after training, it is not always possible to directly attribute the change to training. The following sections describe several common influences to behavior or performance that may mistakenly be attributed to training, as well as methods for separating the effects of training from other factors.

Factors Influencing the Validity of Training

Goldstein describes two conditions that may exist when measuring performance: criterion deficiency and criterion contamination (26). Criterion deficiency exists when the tool being used to measure performance does not adequately measure all the behaviors or aspects of performance that have been impacted by training. For example, a questionnaire is administered to supervisors of employees who have participated in a time management seminar. The questionnaire asks supervisors to rate how many daily goals employees are reaching since attending the seminar. Although the questionnaire captures some aspects of time management, the tool is deficient in that it does not measure other behaviors, such as use of daily logs. The other condition is criterion contamination, which occurs when the tool used to measure performance is also measuring other factors. A common example of criterion contamination is a mathematical test sometimes used as a measurement of basic level math skills. In addition to testing

for math, performance on the test is also influenced by the participant's ability to read. Poor performance on a test that contains word problems might indicate either poor math skills or poor reading skills.

Evaluators need to use care in considering only general measures of performance against which to assess the effectiveness of training. Campbell notes that overall performance is affected by numerous factors, only one of which is training (27). In essence, using a global measure of performance is similar to assuming that a lack of knowledge or skill is always the problem when identifying performance gaps and that training is always the solution.

While the examples above demonstrate potential errors in measuring performance, the following factors influence performance itself. That is, the factors described below might lead an evaluator to erroneously conclude that training impacted performance when, in fact, change in performance was due to another factor (5).

History—Performance can be affected by changes that occur through the passage of time. This factor includes the work environment factors described in chapter 2. Changes to the organization, job redesign, or implementation of new work systems are examples of factors that can influence performance. These factors can influence both behavioral measures and learning measures if the learning measure is not administered immediately after the program.

Maturation—Participants' performance may change simply as a result of physiological, biological, or psychological factors that impact their behavior. For example, if a long period of time passes between training and behavioral assessments, changes may be due to increased strength, decreased mobility, or increase in maturity or experience. Similarly, participants may learn from other sources, thus behavior changes, but not because of the specific training program.

Pretesting—Pretesting participants is a useful method for identifying performance gaps and assessing pretraining levels of performance. However, administering a posttraining test that is similar to the pretest may result in inflated scores due to the participants' familiarity with the test or due to the opportunity to practice the desired behaviors before the posttest. As a result, higher scores on the posttest may be erroneously attributed to the training program when the improvement resulted from practice. This problem, however, does not imply a pretest should not be used. Rather, a control group should be given the pretest as well.

The examples above demonstrate that a number of factors may lead evaluators to inappropriately attribute changes in performance scores to training. The following section describes methods for isolating the impact of training and increasing confidence in the fact that training was responsible for change in performance or change in the organizational measure.

Separating Effects of Training from Other Influences

Pretest-Posttest—One of the most common methods for isolating the impact of training is through a pretraining and

posttraining assessment of performance and skill. Participants are assessed prior to the training program and then after training has been administered. Changes in the scores of participants can be attributed to training. This assessment can be conducted at both learning and behavioral levels of evaluation. This method does have some limitations, as were described above. For this reason, it is sometimes advantageous to use a control group as well.

Control groups—Another common method for isolating impact is through the use of control groups. A control group is a group of employees who do not receive training but are similar to the group receiving training. They are similar in the sense that they may all come from the same department, have similar levels of knowledge and skill, or have the same tenure. Both groups are measured using the pretraining assessment tool and then again on the posttraining assessment tool. If training is the cause of change in performance, then the control group scores will not change, while the scores for the training, or experimental, group will change. A small change may occur in both groups for the reason described above: a practice or learning affect as a result of having taken the pretraining test. However, the experimental group should experience a larger change in test scores. This issue also applies to behavioral measures. The purpose of comparing groups that are similar is to isolate impact even more. Differences in group scores are less likely to be attributed to differences in the nature of the group themselves if they are matched on important variables such as experience, skill, and tenure. The alternative to creating matched groups is to create two randomly assigned groups. According to this method, participants are randomly assigned to either the experimental or the control group. It is then assumed that important factors such as those listed above will be approximately equal for both groups.

Several potential limitations exist when using control and experimental groups. First of all, the nature or size of the business or department might preclude an evaluator from assigning groups randomly or evenly matched. A department might not be large enough to create two groups. Also, the use of control groups might be somewhat amenable to classroom settings where a quantifiable measure of learning or performance is obtainable. However, other measures of learning or performance might be too difficult to assess in a classical control-group design due to practical business constraints. In these cases, it may be argued that a one group pretest-posttest design may be adequate to attribute change in performance to training. Finally, bad feelings can arise in control groups that do not have access to training. It is typically suggested that once the link between training and impact has been demonstrated, all participants receive training (19).

Trend line—Trend line analysis provides an empirical and graphical method of isolating changes resulting from training. Trend line analysis involves collecting performance and operational data for a period of time that includes the training intervention as well as a period of time after training. The performance measures and the operational measures linked to performance are identified with the client. On a regular basis, data are collected from both measures and plotted. If training impacted performance or operational measures, then the line

plotting each data point would demonstrate a change in the shape of the line following training that is not consistent with the general trend of the line before training. This change would also be apparent when seasonal fluctuations are evident in the operational indicator. The longer data are collected prior to and following training, the more confident the evaluator can be in identifying training as the cause of the change.

Multiple sources of data—Using multiple sources of data helps to isolate the impact of training because this decreases the possibility that changes in performance were due to random fluctuations or problems with the measures themselves. Kirkpatrick's four levels of evaluation provide an ideal framework for obtaining multiple sources of data.

The collection of data on learning, behavior, and business results on the same group of training participants provides evaluators with a method of isolating impact because they can track changes over time and across all the measures. (Reaction data are not included because research has demonstrated that reaction data are not directly related to the other levels of evaluation). More specifically, if a change in behavior occurs (measured 3 months after training) and a learning measure demonstrates an increase in the level of knowledge or skill, the HRD professional is more confident that training was the cause for change because the change was evident in two measures. However, change in behavior but no change in learning suggests that another factor outside of training may be influencing behavior. It may also be argued that the tool used to measure learning was not valid. In either case, it warrants further investigation and limits the HRD professional's ability to state confidently that training was the cause. Similarly, a change in the business indicator but no change in behavioral measures for the same group suggests that skills or behaviors that should have been learned during training are not being used, but the business is still being impacted, by another factor. It could be the case that training was the cause for impact, however the wrong behaviors were identified as being linked to the business measures. Again, this would warrant further investigation by the HRD professional and the client (27).

OVERVIEW

Several issues and methods have been presented that address the issue of determining if training was the true cause of change in performance or business results. The HRD professional can never be 100 percent sure that training was the only cause for change. However, awareness of some of the issues that arise when measuring learning, behavior, and results, and practicing methods to isolate the effect of training can significantly increase the level of confidence significantly that training did, in fact, impact performance and the business.

TRANSFER OF TRAINING

In addressing the effectiveness of training, the focus of evaluation is often on achieving a measure of results. The conclusion that no change in business results indicates that training was not effective is only partly true. Participants may have learned new skills or knowledge required to impact the business; however, something else is preventing them from applying the new skills to the job. The last section of this chapter discusses the issue of transfer of training. The goal, however, is not to provide an exhaustive list of potential barriers but rather make the reader aware of where barriers may exist. A comprehensive list of factors that influence transfer may be found in the book *Transfer of Training*, by Broad and Newstrom (20). The authors also describe strategies for reducing barriers to transfer.

According to Broad and Newstrom, transfer of training can be impacted in one of nine categories. On one level, transfer is influenced by the trainer, the participant, and the supervisor. On the other level, transfer can be impacted before, during, and after training. Figure 4 illustrates this matrix. For example, a supervisor who tells employees, "You can attend training but I don't see how it's going to help you" is creating a barrier prior to training by forming doubts in the participant's mind as to the usefulness and effectiveness of the training. Similarly, a trainer who provides little opportunity for practice during the training session is blocking transfer by decreasing the chance that the participant will properly learn the skill or knowledge. Finally, a participant may inhibit transfer after training by not putting the new skills or knowledge into use and reverting to old methods of conducting business. In the matrix, each player impacts transfer at all three points in time.

The role of the performance consultant allows the consultant to identify barriers and assist the client in formulating strategies for circumventing them. In fact, the result of the needs analysis may indicate that training is not the solution, but simply removing a barrier will close performance gaps and impact business needs. For example, the nature of the workplace and a company's reward structure may be in opposition to training. A structure that provides reward for individual contributions conflicts with training that attempts to build teamwork.

OVERVIEW

Adequately evaluating the impact of training and being able to attribute change to training has been an elusive goal for HRD professionals. The models and methods described in this chapter provide a structure for conducting a Level 4 evaluation and for assessing the contributions training has had on that impact. On review of these methods, the HRD professional may realize that successful evaluation is not so elusive after all.

EXAMPLES OF EFFECTIVE EVALUATION METHODS

This chapter provides a summary of training evaluation practices by reviewing examples of evaluation approaches used by state, federal, and provincial transportation agencies. It is important to note that the survey results, as well as the limited follow-up interviews, provided many examples of practices that demonstrate innovative methods and practical applications. The following examples are presented here because they are not only innovative and possibly useful for other transportation agencies, but also have been successful in demonstrating the value of training.

WASHINGTON STATE DEPARTMENT OF TRANSPORTATION

The training department at the Washington State Department of Transportation recognizes the necessity of having efficient training and also demonstrating the value of this training. Training personnel assume that if training departments cannot demonstrate their link to organizational performance, it is a prescription for downsizing. Hence, they have taken a three-pronged approach to ensuring that training is positively affecting organizational performance: they ensure that training programs offered are linked to the strategic directions of the organization, that the right employees are attending the training programs, and that they can demonstrate the value of the training programs they offer.

As part of a total quality management effort, all training programs are linked to organizational strategic objectives. In detail, when designing a program, the relationship between each behavioral or learning objective in that training program and the strategic objectives of the department are specified. This specification ensures that training is aligned with the strategic imperatives of the organization. These specifications are reviewed by the training oversight committee, which includes two deputy-secretaries of transportation. The combination of specifying the link between the organization's objectives and the review of the objectives by the training oversight committee ensures that the correct training programs are offered, and that they are approved by senior management.

To ensure that the correct classes of employees attend these strategically aligned training programs, the department has installed an Automated Training Management System (ATMS). The ATMS is a computer-based training participant prioritization, scheduling, and tracking system. The system, which resides on the DOT's IBM mainframe, has five major functions: needs identification, course scheduling, registration, course confirmation, and report generation. Based on a needs hierarchy for each program and job class, which was approved by the training oversight committee, classes of employees that need

training are specified. When employees call up the system, a curriculum for their personal development is suggested based on the strategic needs of the organization. The system also attempts to ensure that classes are full, which decreases the unit cost of the training program. ATMS ensures that training is efficiently offered to the individuals who need it.

The training department uses Kirkpatrick's four levels of measurement to evaluate training, although to date the most stringent evaluations have focused on behavioral evaluations (using 360-degree assessment methodology). The evaluation of participant reaction to all training programs showed that the results are sufficient for evaluating the performance of outside consultants, and testing applicability of training program content to a specific job. However, they recognize that in-depth training evaluation is costly both in terms of time and materials. Thus, they selectively evaluate for knowledge, behavior, and results. Offering more than 570 courses, it is necessary to have some criteria that specify which courses should be evaluated. Courses that are seen as strategically important by decision makers, that are costly, or are based on complicated material are more stringently evaluated. Course cost is also taken into consideration; department personnel see a trend of increasing residential courses, which will be expensive. This strategy of identifying criteria that are critical to clients offers clear directions on how to select courses for evaluation. Additionally, departmental experience has found that quality instructional design makes evaluation of knowledge and behavioral criteria easier to conduct. Thus, they have found that simultaneously developing an evaluation strategy along with designing the course facilitates the evaluation process.

NORTH CAROLINA DEPARTMENT OF TRANSPORTATION

The North Carolina Department of Transportation (NC DOT) uses a number of different tools for evaluating training. The tools encompass several levels of training evaluation and, taken together, provide a multifaceted assessment of training programs that allows progressive management of human resource development.

NC DOT provides training course participants with a standard form with which to measure their reactions. This evaluation form, rated on a quantitative scale, provides questions pertaining to course material, audio visual materials, instruction, classroom setting, and the instructors. In addition to these rated items, the form contains open-ended questions allowing the participant to include comments on (1) what areas in the course should be added, deleted, or emphasized, (2)

what additional related training should be made available, and (3) general comments. The form does not address expected behavioral or performance changes as a result of the training.

Along with measuring reactions at the end of the course, NC DOT also provides 6- and 12-month follow-up surveys to participants and their supervisors. Participants are asked if they have been able to apply new concepts in their work and supervisors whether they have observed any change. These follow-up surveys are a good measure of behavioral change (Level 3).

To manage the economic impact of its training department, NC DOT adopted a product focus, which uses the costs of external training as a benchmark for the economic efficiency of the Training and Development Department. This focus realigned trainers with the programs they offered by redefining their jobs as product managers who are responsible for managing the efficiency of their training product. Using the costs of a similar (external) course as a benchmark, they calculate the overall costs of their course, including fixed and variable costs associated with marketing, presenting, and evaluation. This evaluation of the costs and benefits of the training department by product was implemented to proactively address the need for accountability in local government.

ILLINOIS DEPARTMENT OF TRANSPORTATION

Constructing separate evaluation strategies geared to different types of training is useful because the focus of the evaluation may differ from program to program. The Illinois Department of Transportation (IDOT) demonstrates this practice in two separate training tracks: management development training and technical training. Evaluation for each track is handled separately.

Management development training, called the Growth, Recognition & Employee Advancement Training (GREAT) (detailed in *NCHRP Synthesis 188: Management Training and Development Programs*) (2), is evaluated through a number of tools, starting with a reaction-level survey that asks respondents to rate the course and instructor. Another piece of the GREAT evaluation is a team evaluation of an entire program. This allows members of a team to evaluate both the effectiveness of the program in meeting its objectives and the applicability of learned material to the job. Finally, informal feedback is collected from executives who have had subordinates attend the programs as well as from past participants to determine how behavior has been impacted as a result of the training program.

The result of examining the reaction data, behavioral change data, and instructor effectiveness data is the fine-tuning of each targeted program curriculum. IDOT estimates that approximately 7 percent of its staff time is allocated to evaluation in one form or another. This effort is limited due to the large training demand of the GREAT program.

The technical training covers a wide variety of skills and competencies under engineering topics. As such, learning is easily measured through comprehensive tests of the subject material. In some cases, future work assignments or reassignments are contingent on passing the test. In all cases, the test score becomes a part of the employee's permanent record.

MASSACHUSETTS PORT AUTHORITY

Massachusetts Port Authority (Massport), the agency responsible for management of the Ports of Boston, demonstrates the value of expanding the definition of training evaluation. The training department at Massport has integrated evaluation into training by ensuring that the training addresses aspects of employee performance that are critical to organizational performance. Massport evaluates 100 percent of its training programs with postcourse student reactions and some courses with pre- and posttraining knowledge tests. However, Massport is transitioning to competency-based training, which is most noteworthy. This profile does not define evaluation as a demonstration of a change in employee performance or in the value of employee performance. Rather, the emphasis is on ensuring that the content of the training is aligned with the performance required by the organization. The basis of the competency-based learning system is a validated strategic competency model.

At the direction of its board of directors, Massport has developed a competency model that drives the performance management system factors in personal development plans, and informs the design of training. This competency model is the basis of a seamless system designed to create a learning organization, where employees will be responsible for their development, and where the role of the training department is to support this development. The competency model summarizes the behaviors that differentiate high performers from average performers. The model was developed by interviewing high-performing managers about critical success incidents, and identifying common behavioral themes that caused success in the incident. This development method yielded two outputs: a competency model and examples of the competencies in action.

The competency model was approved by senior management and was used to build both the performance-appraisal/reward system and the personal-development-plan systems. These assessment systems enable individual employees to recognize individual strengths and developmental needs. The performance appraisal system both rewards and records performance of competencies.

Future applications of the competency model include recruitment and selection, as well as training needs analysis.

The examples of competencies in action are used to inform training design. Specifically, the examples are used as case studies and lecture examples in training programs for managers, supervisors, and individual contributors. This use of examples of competencies in action links training to the competencies that are critical to organizational strategy and are the basis of employees' developmental focus. Because the training dovetails with both the strategy of the organization and the rewards of participants, its relevance and importance are assured.

NOTEWORTHY MEASURES USED FOR EVALUATION

Each of the above examples considers the values of both clients and stakeholders, and demonstrates that evaluation can

be successful if necessary planning is conducted early in the training-program development process. Many of these examples also demonstrate the use of innovative measures or metrics, on which the evaluations are built. It is in the selection of measures that creativity and open-mindedness play a major role in evaluation. A few additional examples of creative measures deserve attention.

Sometimes, appropriate measures of behavioral change are best observed by people outside of the organization. Departments of Public Works often use customer complaints, requests, and recommendations to determine the impact of training or to assess the need for training. For some specific types of training, which may not be observable by customers (e.g., safety training, and driver training), an examination of the condition of equipment being used by the employee helps determine the impact of training.

The Kansas Department of Transportation has made the employee an integral part of the evaluation process. To evaluate behavioral changes as a result of training, employees are contacted 3 to 6 months after completing training. Participants receive a copy of an action plan they developed during training and are asked to compare their plans with their current performance or outputs. This example seamlessly connects the training evaluation and personal development plans by ensuring that the employee attempts to transfer the trained skills back to the job.

The Virginia Department of Transportation is using an anonymous opinion survey of workplace competency to evaluate training. The survey, which was first distributed in 1995, asks employees to rate the extent to which other employees demonstrate key competencies on the job. The results from the 50 percent of surveys that were returned provided some surprises for management, and demonstrated a need for training. Goals to increase specific competencies have been set, and achievement of these goals will be measured in a second wave of the survey that will be distributed 18 months after the original survey.

OVERVIEW

As described in chapter 3, the current training evaluation practices of the DOTs do not differ drastically from other industries. In fact, some methods of evaluation are used more frequently in DOTs. DOTs also face the same issue as most organizations: discovering better methods for determining the effectiveness of training programs. As demonstrated from the examples above, there are great differences among agencies in terms of the sophistication with which the training evaluations are conducted. It should be clear that within the federal and state DOTs, it is possible to be highly sophisticated. This review of best practices provides insights into strategies that can be used to demonstrate the impact of training.

Trainers and evaluators have traditionally become involved in the evaluation process after a training program has been implemented and even after the program is complete. Responses from the survey suggest that the same holds true for transportation agencies. However, as seen from the above examples, this is not the case in the more successful training evaluation practices, where the role of the trainer has been transformed into a performance consultant. In these cases, trainers work closely with their clients to achieve the strategic goals of the agency, creatively use metrics that are considered critical to decision makers as well as stakeholders, and begin the evaluation efforts even before the program is piloted. Indeed, in some cases, the evaluation is built into program design.

While this synthesis has presented various models for program evaluation, it should be highlighted that there is no standard evaluation model. Training program evaluations must be customized to the needs of the client and the agency. The examples presented may provide partial evaluation architecture templates, but ultimately evaluations, like all aspects of HRD, must be customized to the situation. Some of the practices highlighted in this synthesis may provide tools and processes that will enable further evaluation of work and an understanding of the impact that training, and other human resource development interventions, have on transportation agencies.

REFERENCES

1. Robinson, D. G. and J. C. Robinson, *Training for Impact: How to Link Training to Business Needs and Measure the Results*, Jossey-Bass Publishers, San Francisco (1990).
2. Harder, B. T., *NCHRP Synthesis of Highway Practices 188: Management Training and Development Programs*, Transportation Research Board, National Research Council, Washington, D.C. (1994).
3. Phillips, J. J., *Handbook of Training Evaluation and Measurement Methods (2nd Edition)*, Gulf Publishing Company, Houston, TX (1991).
4. Filipczak, B., "Training Budgets Boom," *TRAINING*, (October, 1993) pp. 37-47.
5. Poister, T. H., L. G. Nigro, and R. Bush, *NCHRP Synthesis of Highway Practices 163: Innovative Strategies for Upgrading Personnel in State Transportation Departments*, Transportation Research Board, National Research Council, Washington, D.C. (1990).
6. Kirkpatrick, D. L., *How to Train and Develop Supervisors*, AMACOM, New York (1993).
7. Robinson, D. G., and J. C. Robinson, *Performance Consulting: Moving beyond Training*, Berrett-Koehler Publishers (1995).
8. Mager, R. F., and P. Pipe, *Analyzing Performance Problems or "You Really Oughta Wanna,"* Fearon/Lear Siegler, Belmont, CA (1970).
9. Goldstein, I. L., *Training in Organizations: Needs Assessment, Development, and Evaluation (2nd Edition)*, Brooks/Cole Publishing Company, Monterey, CA (1986).
10. Ostroff, C., and J. K. Ford, "Assessing Training Needs: Critical Levels of Analysis," in *Training and Development in Organizations*, I. L. Goldstein, ed., Jossey-Bass Publishers, San Francisco, CA (1989).
11. Basarab, D. J., and D. K. Root, *The Training Evaluation Process*, Kluwer Academic Publishers, Boston, MA (1992).
12. McClelland, D. C., "Testing for competence rather than for 'intelligence'," *American Psychologist*, Vol. 28 (1973), pp. 1-14.
13. Kirkpatrick, D. L., "Techniques for Evaluating Training Programs," *Journal of ASTD*, Vol. 13, No. 11 (1959), pp.3-9.
14. Kirkpatrick, D. L., "Techniques for Evaluating Training Programs: Part 2—Learning," *Journal of ASTD*, Vol. 13, No. 12 (1959), pp.21-26.
15. Kirkpatrick, D. L., "Techniques for Evaluating Training Programs: Part 3—Behavior," *Journal of ASTD*, Vol. 14, No. 1 (1960), pp.13-18.
16. Kirkpatrick, D. L., "Techniques for Evaluating Training Programs: Part 4—Results," *Journal of ASTD*, Vol. 14, No. 2 (1960), pp.28-32.
17. Brinkerhoff, R. O., *Achieving Results From Training*, Jossey-Bass Publishers, San Francisco, CA (1987).
18. Scriven, M. B., "The methodology of evaluation," in *Perspectives of Curriculum Evaluation*, American Educational Research Association Monograph, No. 1, Rand McNally, Chicago (1967).
19. Kirkpatrick, D. L., *Evaluating Training Programs: The Four Levels*, Berrett-Koehler Publishers, San Francisco (1994).
20. Broad, M. L. and J. W. Newstrom, *Transfer of Training*, Addison-Wesley Publishing Company, Reading, MA (1992).
21. Carnevale, A. P. and E. R. Schulz, Evaluation practices, *Training Development Journal*, Vol. 44 (1990), pp. S-23-S-29.
22. Phillips, J. J., "Evaluating the Effectiveness of Training," *Proceedings of the Fourth Annual Assessment, Measurement and Evaluation of Human Performance Conference*, Linkage, Inc., and Lesley College (June, 1995) pp. 125-153.
23. Cascio, W. F., *Costing Human Resources: The Financial Impact of Behaviors in Organizations*, Kent, Boston (1982).
24. Spencer, L., *Calculating Human Resources Costs and Benefits*, Wiley, New York (1986).
25. Fitz-enz, J., *How to Measure Human Resources Management*, 2nd edition, McGraw-Hill, Inc., New York (1995).
26. Goldstein, I. L., Training in Organizations, in *Handbook of Industrial/Organizational Psychology*, Vol. 2, 2nd edition, M. D. Dunnette and L. M. Hough, eds., Consulting Psychological Press, Palo Alto, CA (1991).
27. Campbell, J. P., Personnel Training and Development, *Annual Review of Psychology*, Vol. 22 (1971) pp. 565-602.

BIBLIOGRAPHY

- American Society for Training and Development, *ASTD's Trainer's Toolkit: Evaluating Results of Training*, ASTD, Alexandria, VA (1992).
- Baldwin, T. T., and J. K. Ford, "Transfer of Training: A Review and Directions for Future Research," *Personnel Psychology*, Vol. 41 (1988), pp. 63-105.
- Block, P., *Flawless Consulting: A Guide to Getting Your Expertise Used*. Learning Concepts, Austin, TX (1981).
- Dixon, N.M., "Meeting Training's Goals Without Reaction Forms," *Personnel Journal*, Vol. 66, No. 8 (August 1987), pp. 108-115.
- Gordon, J., R. Zemke, and P. Jones (eds.), *Designing and Delivering Cost-Effective Training—and Measuring the Results*. 3rd edition, Lakewood Books, Minneapolis, MN (1993).
- Grenough, J.L. and R. Dixon, "Using 'Utilization' to Measure Training Results," *Training/HRD*, (February 1982), pp. 40-42.
- Head, G. E., *Training Cost Analysis: A How-To Guide for Trainers and Managers*, ASTD, Alexandria, VA (1993).
- Laird, D., "Four Ways to Evaluate Your Training Program," *Successful Meetings*, Vol. 33, No. 1 (January 1984), p. 79.
- Jamieson, D. and J. O'Mara, *Managing Workforce 2000: Gaining the Diversity Advantage*, Jossey-Bass Publishers, San Francisco (1991).
- Mager, R. F., *What Every Manager Should Know About Training*, Lake Publishing, Belmont, CA (1992).
- Mager, R. F., *Measuring Instructional Results*, 2nd edition, Lake Publishing, Belmont, CA (1984).
- Patton, M. Q., *Qualitative Evaluation Methods*, Sage Publishing, Newbury Park, CA (1980).
- Phillips, J. J., ed., *In Action: Measuring Return on Investment*, ASTD, Alexandria, VA (1994).
- Zemke, R. and J. Gunkler, "28 Techniques for Transforming Training into Performance," *TRAINING*, (April, 1985) pp.48-63.
- Zemke, R. and T. Kramlinger, *Figuring Things Out: A Trainer's Guide to Needs and Task Analysis*, Addison-Wesley Publishing Company, Reading, MA (1982).

APPENDIX A

Survey Questions

ASSESSING TRAINING NEEDS

1. In our organization, the need for a training program is based on (check all that apply):
 - employee performance gaps
 - competency model(s)
 - organizational mission/vision
 - strategic goals
 - management directives
 - employee opinions and/or attitudes
 - customer satisfaction data
 - other

2. What is the title, name, and phone number of the person who directs the training needs assessment process?

3. Who else, if anyone, is involved in the needs assessment process? (check all that apply)
 - Upper management
 - Middle management
 - Manager of Training
 - Manager of Human Resources
 - Trainer/Facilitator
 - Training Evaluator
 - Supervisors of program participants
 - Program participants
 - Other

ASSESSING TRAINING EFFECTIVENESS

1. Please list major training programs that are currently being offered in your organization. Include the following:
 - the delivery method used for that course
 - the evaluation method used for that course

2. For courses that you have evaluated in terms of their impact on behavior, performance, or organizational results, please describe the methods and measures you use.

3. Please describe any other innovative methods of evaluation you are currently using. Please include any examples or documentation.

ASSESSING FUNDING FOR TRAINING

1. What was the overall annual budget (fiscal resources) allotted to training for the last fiscal year?
2. Does this amount represent an increase or decrease in funding from the previous year?
3. Of that amount, how much is allocated for the evaluation of training?
4. Does this amount represent an increase or decrease in funding for evaluation from the previous year?

COMMENTS

Please provide any additional comments or documentation that may help us understand your department's training evaluation process.

APPENDIX B

List of Responding Transportation Agencies

Alaska
Arizona
Arkansas
California
Connecticut
Delaware
District of Columbia
Florida
Georgia
Illinois
Indiana
Kansas
Kentucky

Louisiana
Maine
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nevada
New York
North Carolina
North Dakota
Ohio

Pennsylvania
Rhode Island
South Dakota
Tennessee
Texas
Vermont
Virginia
Wisconsin
Wyoming
Alberta (Canada)
Federal Highway Administration

THE TRANSPORTATION RESEARCH BOARD is a unit of the National Research Council, a private, nonprofit institution that provides independent advice on scientific and technical issues under a congressional charter. The Research Council is the principal operating arm of the National Academy of Sciences and the National Academy of Engineering.

The mission of the Transportation Research Board is to promote innovation and progress in transportation by stimulating and conducting research, facilitating the dissemination of information, and encouraging the implementation of research findings. The Board's varied activities annually draw on approximately 4,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation.

The National Academy of Sciences is a nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce Alberts is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encouraging education and research, and recognizes the superior achievements of engineers. Dr. William A. Wulf is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences, by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce Alberts and Dr. William A. Wulf are chairman and vice chairman, respectively, of the National Research Council.

**Transportation Research Board
National Research Council
2101 Constitution Avenue, N.W.
Washington, D.C. 20418**

ADDRESS CORRECTION REQUESTED

**NON-PROFIT ORG.
U.S. POSTAGE
PAID
WASHINGTON, D.C.
PERMIT NO. 8970**