

1997

Technical Training Program Evaluation: Present Practices in United States' Business and Industry.

Skip Twitchell

Louisiana State University and Agricultural & Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_disstheses

Recommended Citation

Twitchell, Skip, "Technical Training Program Evaluation: Present Practices in United States' Business and Industry." (1997). *LSU Historical Dissertations and Theses*. 6552.
https://digitalcommons.lsu.edu/gradschool_disstheses/6552

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Historical Dissertations and Theses by an authorized administrator of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

**A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600**

**TECHNICAL TRAINING PROGRAM EVALUATION: PRESENT PRACTICES IN
UNITED STATES' BUSINESS AND INDUSTRY**

A Dissertation

**Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy**

in

The School of Vocational Education

**by
Skip Twitchell
B.S., Auburn University, 1979
December, 1997**

UMI Number: 9810840

UMI Microform 9810840

Copyright 1997, by UMI Company. All rights reserved.

**This microform edition is protected against unauthorized
copying under Title 17, United States Code.**

UMI

**300 North Zeeb Road
Ann Arbor, MI 48103**

DEDICATION

This work would not have been possible without the sacrifices of my wife Diane and daughter Christine, the continuing encouragement and support of my major professor Dr. James Trott over many years, and the understanding of my graduate committee through the difficult and halting process of this research.

ACKNOWLEDGMENTS

The author would like to acknowledge the support of the American Society for Training and Development in supplying the population data and some mailing supplies used in this study. Dr. Elwood F. Holton III and Dr. Jack J. Phillips as second and third authors of the survey instrument increased the breadth of the study. Thanks to Anne Brown for her diligent proofreading which always finds more errors than any other reader. The grammatical, typographical, and spelling errors contained in the original work have been greatly reduced by her efforts. Finally, the author would like to publicly thank the respondents for their time and understanding in providing sensitive data for this study. The following organizations were willing to be identified as participants in this research:

Acuity Imaging, Inc.	Comprehensive Technologies
ADP Hollander	Diane Thomas Operations Officer
Alcoa Quality Education and Training	Don Dimmick and Assoc.
Autologic, Inc.	Florida First Federal Savings Bank
Basic Software, Inc.	Gulfstream Aerospace Corporation
Bayer Diagnostics Division	Health Images, Inc.
Bush Ind.	Henry L. Berman Solutions By Design
Cerberus Pyrotronics	IDX Systems Corporation
City of Phoenix Transit System	Ingersoll - Rand Company
Communications Workers of America Company-TTS	Insta-Foam Products, Inc.

International, Inc.	Security-Connecticut Life Insurance Co.
James E. Hamiter	Southwestern Public Service Company
Lehecka Pratt Associates, Inc.	Sovereign Bank
Management Technologies Group, Inc.	System One Corp.
MSAS Cargo International Inc.	The Raymond Corporation
North American Security Life	The George F. Cram Company, Inc.
Novations Group Inc.	The Peoples Gas Light and Coke
Numatics Inc.	United Innovations
Okidata	Unocal
Oncology Nursing Society	US West Communications
Pacific Gas + Electric Co Humbolt Bay	Vulcan Materials Co.
Packard Electric Division of GM	
Pepsi- Cola	
Planning Masters	
Power Plant	
Rite Aide Corporation	
Science Applications International Corp.	
Scoville Press Inc.	

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGMENTS	iii
LIST OF TABLES	xi
LIST OF FIGURES	xiii
ABSTRACT	xiv
CHAPTER 1: INTRODUCTION	1
<u>Importance of Evaluation</u>	1
<u>Training costs</u>	1
<u>Evaluating for improvement</u>	2
<u>Evaluating to prove value</u>	3
<u>Lack of Training Evaluation</u>	3
<u>Problems with Available Data</u>	6
<u>Importance of Training</u>	8
<u>Competing globally</u>	8
<u>Upgrading skills</u>	9
<u>Worker benefits</u>	10
<u>Technical Training</u>	10
<u>Summary</u>	11
<u>Problem Statement</u>	12
<u>Objectives</u>	12
<u>Significance of Study</u>	14
<u>Assumptions</u>	16
<u>Limitations</u>	18
<u>Definitions</u>	19
CHAPTER 2: REVIEW OF LITERATURE	22
<u>Introduction</u>	22
<u>Introduction to Kirkpatrick's Model</u>	23
<u>Elaborations on Kirkpatrick's Model</u>	24
<u>Participant reaction</u>	24
<u>Learning outcomes</u>	26
<u>Behavior changes</u>	27
<u>Organizational change due to training</u>	32
<u>Evaluation Use</u>	35

<u>Level 1</u>	35
<u>Level 2</u>	36
<u>Level 3</u>	37
<u>Level 4</u>	38
CHAPTER 3: METHODOLOGY	40
<u>Purpose of the Study</u>	40
<u>Population and Sample</u>	40
<u>Instrumentation</u>	44
<u>Data Collection Procedure</u>	45
<u>Data Analysis</u>	47
CHAPTER 4: DATA ANALYSIS	51
<u>Sample Size and Response Rate</u>	51
<u>Comparison of Respondents to Non-respondents</u>	52
<u>Demographics</u>	53
<u>The Extent to Which Business and Industry Are Using Evaluation Levels</u> <u>1, 2, 3, or 4 to Evaluate Technical Training</u>	57
<u>Evaluation methods used at Level 2</u>	63
<u>Evaluation methods used at Level 3</u>	68
<u>Evaluation methods used at Level 4</u>	73
<u>Various Reasons for Not Evaluating</u>	76
<u>Exploratory Correlations Between Business and Industry Variables</u> <u>and Evaluation Variables</u>	82
<u>Exploratory correlations between organizational function and the</u> <u>Level of evaluation used</u>	82
<u>Exploratory correlations between organizational function and the</u> <u>method of evaluation used</u>	83
<u>Exploratory correlations between an organization's size and the</u> <u>Level of evaluation used</u>	83
<u>Exploratory Correlations Between Training Managers' Perceptions of the</u> <u>Importance of a Level and the Level of Evaluation Used</u>	84
<u>Exploratory Correlations Between Training Managers' Experience and the</u> <u>Level of Evaluation Used</u>	85
<u>Exploratory Correlations Between Age of the Training Program and the</u> <u>Level of Evaluation Used</u>	85
<u>Exploratory Correlations Between Various Organizational Training</u> <u>Practices and the use of each Level or Method of Evaluation</u>	85
<u>Training methods and the percentage of programs using each</u> <u>evaluation Level 1 through 4</u>	86
<u>Integration of evaluation in programs and the percentage of</u> <u>programs using each evaluation Level 1 through 4</u>	88

<u>Perceived importance of evaluation in proving value to upper management and the percentage of programs reporting evaluation results to management</u>	89
<u>Perceived importance of evaluation in improving programs and the percentage of programs reporting outcomes to participants</u> .	90
<u>Programs delivered to all the members of a target audience and the Level of evaluation used</u>	91
<u>Programs that are delivered to change performance or organizational outcomes and the Level of evaluation used</u>	91
<u>Percent of staff involved in evaluation and the Level of evaluation used</u>	92
<u>Percent of staff with formal training in evaluation and the Level of evaluation used</u>	93
<u>Scientifically accepted research techniques and reasons for delivering training</u>	93
<u>Percentage of programs dependent on evaluation for funding and the Levels of evaluation used</u>	94
<u>Percentage of budget dependent on evaluation for funding and the Levels of evaluation used</u>	95
<u>Funding methods for training and the Level of evaluation used</u>	96
 CHAPTER 5: SUMMARY	98
<u>Introduction</u>	98
<u>Objectives</u>	100
<u>Methods</u>	101
<u>Results</u>	103
<u>The extent to which business and industry are using evaluation</u>	
<u>Levels 1, 2, 3, or 4 to evaluate technical training</u>	103
<u>Evaluation methods used at Levels 2, 3, and 4</u>	103
<u>Reasons for not evaluating</u>	105
<u>Exploratory correlations between various organizational variables and the 4 Levels of evaluation</u>	105
<u>Exploratory correlations between business and industry function and Level or method of evaluation used.</u> ..	106
<u>Exploratory correlations between business and industry size and Level of evaluation used.</u>	106
<u>Exploratory correlations between training managers' perceptions of the importance of a Level and its use.</u>	106
<u>Exploratory correlations between training managers' experience and the use of each Level of evaluation.</u>	107

<u>Exploratory correlations between age of the training program and the use of each Level of evaluation.</u>	107
<u>Exploratory correlations between various organizational training practices and the use of each Level or method of evaluation</u>	108
<u>Training methods and the percentage of programs using each evaluation Level 1 through 4.</u>	108
<u>Correlation between integration of evaluation in programs and the use of evaluation at each Level.</u>	108
<u>Correlation between perceived importance in demonstrating value to management and frequency of reporting to management.</u>	109
<u>Correlation between perceived importance in improving programs and frequency of reporting to participants.</u>	109
<u>Programs that are delivered to change performance or organizational outcomes and the Level of evaluation used.</u>	109
<u>Correlation between percent of staff involved in evaluation and the Level of evaluation used.</u>	110
<u>Percent of staff with formal training in evaluation and the Level of evaluation used.</u>	110
<u>Correlation between scientifically accepted evaluation techniques and reason for delivering training.</u>	110
<u>Percentage of programs dependent on evaluation for funding.</u>	110
<u>Training managers' perceived values for each evaluation Level and the Level of evaluation used.</u>	111
<u>Conclusions</u>	111
<u>The extent to which business and industry are using evaluation Levels 1, 2, 3, or 4 to evaluate technical training</u>	111
<u>Evaluation methods used</u>	115
<u>Reasons for not evaluating</u>	115
<u>Exploratory correlations between business and industry variables and evaluation</u>	116
<u>Exploratory correlations between training managers' perceptions of the importance of a Level and the Level of evaluation used</u>	116
<u>Exploratory correlations between training managers' experience and the Level of evaluation used</u>	117

<u>Exploratory correlations between various organizational training practices and the use of each Level or method of evaluation</u>	117
<u>Training methods and the percentage of programs using each evaluation Level 1 through 4.</u>	117
<u>Integration of evaluation in programs and the percentage of programs using each evaluation Level 1 through 4.</u>	118
<u>Perceived importance of evaluation and the percentage of programs reporting outcomes.</u>	118
<u>Training staff evaluation variables and the Level of evaluation used.</u>	119
<u>Scientifically accepted research techniques and reasons for delivering training.</u>	119
<u>Funding variables and the percentage of programs using each Level of evaluation.</u>	119
<u>Recommendations for Practice</u>	120
<u>Suggestions for Further Research</u>	120
<u>Research questions based on Kirkpatrick's Model</u>	121
<u>Research questions based on identifying what other methods are being used to evaluate training</u>	121
<u>Implications</u>	123
<u>The lack of evaluation</u>	123
<u>Evaluating for an audience</u>	124
 BIBLIOGRAPHY	 127
 APPENDIX A: SURVEY	 131
 APPENDIX B: LETTERS AND POST CARDS	 138
<u>Cover letter sent with survey</u>	138
<u>Reminder post card</u>	139
<u>First follow-up letter</u>	140
<u>Second follow-up letter.</u>	141
 APPENDIX C: COMPARISON OF RESPONDENTS AND NON-RESPONDENTS ON SELECTED VARIABLES	 142
 APPENDIX D: TABLES FOR THE WRITTEN RESPONSES TO THE METHODS USED TO EVALUATE AT EACH LEVEL	 143

APPENDIX E: REASONS FOR NOT EVALUATING WRITTEN IN BY THE	
RESPONDENTS	146
<u>Level 1</u>	146
<u>Level 2</u>	146
<u>Level 3</u>	148
<u>Level 4</u>	149
VITA	151

LIST OF TABLES

1.	<u>Non-respondent Sample</u>	52
2.	<u>Descriptive Statistics for Demographic Data</u>	54
3.	<u>Percent of Programs using Each Level of Evaluation</u>	59
4.	<u>Frequencies and Percentages of Programs Using Each of Kirkpatrick's Four Levels of Evaluation for Selected Percentage Intervals</u>	59
5.	<u>Percentages and Frequencies for Organizations Using Each Level of Evaluation in Some, More than Half, and Most of Their Programs</u> .	61
6.	<u>Frequencies for the "Percentage of Programs Using Each Level 2 Evaluation Method" for All Respondents</u>	64
7.	<u>Percentages for the "Percentage of Programs Using Each Level 2 Evaluation Method" for All Respondents</u>	65
8.	<u>Frequencies for the "Percentage of Programs Using Each Level 3 Evaluation Method" for All Respondents</u>	69
9.	<u>Percentages for the "Percentage of Programs Using Each Level 3 Evaluation Method" for All Respondents</u>	70
10.	<u>Frequencies for the "Percentage of Programs Using Each Level 4 Evaluation Method" for All Respondents</u>	73
11.	<u>Percentages for the "Percentage of Programs Using Each Level 4 Evaluation Method" for All Respondents</u>	75
12.	<u>Reasons for Not Evaluating at Each Level 1 through 4</u>	78
13.	<u>Correlations Between Training Methods and the Percentage of Programs Using Each Level of Evaluation</u>	87
14.	<u>Correlations Between Integration of Evaluation in Programs and the percentage of programs using each Level of evaluation</u>	89
15.	<u>Correlations Between Evaluation Techniques and the Reasons for Training</u>	95

16.	<u>Correlation Between Funding and the Use of Each Evaluation Level</u>	97
D1.	<u>Response and "Percentage of Programs Using Each Level 2 Evaluation Method" for Respondents Answering Other</u>	143
D2.	<u>Response and "Percentage of Programs Using Each Level 3 Evaluation Method" for Respondents Answering Other</u>	144
D3.	<u>Response and "Percentage of Programs Using Each Level 4 Evaluation Method" for Respondents Answering Other</u>	145

LIST OF FIGURES

1.	Frequencies by size range for the 106 organizations that provided size datum	55
2.	Frequencies for the 93 respondents who self-selected into the industry groups provided on the survey	56
3.	Frequencies for Level 2 evaluation methods in all use categories other than zero	67
4.	Frequencies for Level 3 evaluation methods in all use categories other than zero	72
5.	Frequencies for level 4 evaluation methods in all use categories other than zero	76
6.	Reasons for Not Evaluating at Each Level	80

ABSTRACT

There is a lack of information on training evaluation. For those studies addressing the amount of evaluation that occurs and in which samples can be identified, the samples are non-random. There have been very few reports on the methods used for evaluation or the reasons why there is so little evaluation.

This study used a random sample selected from members of the ASTD professional practice area titled, Technical and Skills Trainers. It can be argued that this is an informed sample involved in a training area that produces objective outcomes and should produce a favorable picture of training evaluation.

Questions on the survey were based on Kirkpatrick's four Levels of evaluation and gathered information on amount of evaluation, methods used, reasons for not evaluating, organizational training practices, respondents thoughts about the value of evaluation, and demographics. A total of 146 surveys were returned for an overall response rate of 42 %, a higher than normal response rate (20-30%) for this type of survey.

Survey results supported the lack of training evaluation cited in other studies. This study found that technical training managers reported using each of Kirkpatrick's four Levels of evaluation in the following percentage of their courses: Level 1 - 72.74%, Level 2 - 47.05%, Level 3 - 33.73%, and Level 4 - 20.82%. Level 1 evaluation methods were not surveyed. The most commonly used methods at Level 2 were skill demonstrations and posttest with no pretest, at Level 3 observation and

performance appraisals, and at Level 4 productivity estimates, productivity measures, and regulation compliance.

The results of this study indicate that organizations seldom require training departments to evaluate, training departments do not have the knowledge and skills required to perform evaluations, and the cost of evaluations are seen to outweigh the benefits.

The only significant correlation between a training manager's perception of the importance of a level of evaluation in demonstrating value to management and the frequency of reporting evaluation outcomes to managers was at Level 1. No significant correlations were found between any of the four Levels and dependency on evaluation for funding.

CHAPTER 1

INTRODUCTION

The literature commonly starts any general discussion of evaluation with a paragraph or more on the importance of evaluation. The points most frequently made relate to the justification of the investment in time, the justification of the capital required to train employees, and the need for information with which to decide what programs to develop, implement, and retain. This information should also be the basis for decisions about the role of training in an organization (Carnevale & Shultz, 1990; Dixon, 1990; Gordon, 1991; Phillips, 1991; Robinson & Robinson, 1989). A second common thread in the literature on evaluation is the generally accepted belief that only a small portion of training is evaluated (Dixon, 1990; Carnevale & Shultz 1990; Gordon, 1991; Phillips, 1991; Robinson & Robinson, 1989; Survey #11, 1989). This raises two questions. Is there a lack of training evaluation? If training evaluation is seldom carried out, why is this important function not an integral part of training programs?

Importance of Evaluation

Training costs

U. S. businesses and industries spend a tremendous amount of time and money providing training to their employees. Production workers and administrative employees alone received an estimated 751 million hours of formal training in 1996, an increase of 100 % in eight years ("Industry report", 1988; "Industry report", 1996).

Organizations with more than 100 employees budgeted \$60 billion for formal training in 1996, a 50 % increase over expenditures in 1988. ("Industry report", 1988; "Industry report", 1996). Costs are both in time and in dollars. The United States, its industries, and its workers all have a tremendous investment in training (Carnevale, 1991).

To retain the United States' position as the most productive country in the world, new workers must be appropriately prepared for positions in a workplace that demands more skills. Those already in the workforce, who are not prepared or are falling behind in keeping up with the changing technological and organizational skills required by an ever more demanding work environment, must be retrained (Carnevale, 1991). The demand for more skilled workers is rising and the trend toward increasing training costs will continue, barring an unforeseen change in the business environment.

Evaluating for improvement

"The reason for evaluating is to determine the effectiveness of a training program" (Kirkpatrick, 1994). Evaluation can provide information about such factors as how much learning takes place, the use of what is learned on the job, student variables, and delivery variables. This information can be used to identify those factors that have a positive or a negative effect. Once identified, those factors that positively affect training can be maintained and those factors that negatively affect training can be changed or addressed (Dixon, 1990; Kirkpatrick, 1994). The data

gathered through evaluation can be used to compare actual training outcomes with the predicted or required outcomes so that programs may be modified to match needs.

Evaluation can be used "to improve the design or delivery of learning events" (Dixon, 1990, p. 2). Concepts such as experiential learning, learning styles, and cognitive aging differences can be applied to training programs and their ability to change the effectiveness and efficiency of training tested using evaluation data. New instructional technologies such as multimedia delivery can be compared with other means of delivering training on cost, effectiveness, and acceptance based on evaluation results. Evaluation can be used to identify the most effective types of learning events (Dixon, 1990).

Evaluating to prove value

The literature on program development, whether in the field of training or education, includes evaluation as a necessary part of program development. The literature discusses the need for such evaluations to prove the value of training, to maintain funding, and to give management the information on which to base decisions concerning the development, modification, and continuation of training programs.

Lack of Training Evaluation

Information about how much training evaluation is conducted in business and industry is usually based on Donald Kirkpatrick's four Levels:

- Participant reaction
- Attainment of learning objectives
- Actual changes in on-the-job performance
- The effect of training on the organization

(Medsker & Roberts, 1992).

Participant evaluation (Level 1 or reaction) addresses subjective issues such as the trainee's feeling about the value of the program, the quality of the instructor, how the program may be improved, and other variables included on the reaction form. Participant evaluations are usually accomplished using reaction forms. These forms are administered during, at the end of, or immediately after a training program and provide the trainee an opportunity to evaluate training subjectively.

Data on the participant's feelings about various aspects of a training program, such as presentation, content, and appropriateness, are determined using Level 1 evaluation. This Level of evaluation collects data that may be used to judge a program's acceptance.

Level 2 evaluation (measures of learning) consists of posttests (either pen and paper or skills tests) to find out if the objectives of the training were accomplished. Did the employee gain the skills and knowledge that the training was designed to deliver? A set of well-developed objectives based on Mager's seminal work in this area, Preparing Instructional Objectives (1984b), makes this Level of evaluation simple and straightforward. The testing criteria are written into each objective.

Level 2 evaluation verifies that training achieved its objectives, not that it achieved training goals (changes in job performance). At this Level, the skills and knowledge gained by the participant during the learning event are measured. The skills and knowledge gained may or may not be used in the workplace (Kirkpatrick, 1975c).

Kirkpatrick's Level 3 evaluation attempts to determine if changes in job performance occurred as a result of training. To evaluate any new training concept, method, or technique, simply comparing outcomes using Level 2 evaluations is not sufficient (Dixon, 1990). "There may be a big difference between knowing principles and techniques and using them on the job" (Kirkpatrick, 1975c, p. 10).

Level 4 evaluation evaluates the impact of training on the organization. The evaluation is summative, placing a value on the outcomes of training. This is information on which to decide if training is an effective solution to an organizational problem, not if a training program is effective within itself (meets its own objectives) (Cascio, 1982 ; Dixon, 1990; Phillips, 1991).

Most of the available literature reports that Level 1 evaluation is common across business and industry and that each Level becomes less common moving from 1 to 4. "As recently as 1988, a report on forty-five Fortune 500 companies showed that although 100 percent used some form of participant reaction form, only 30 percent used measures of learning and only 15 percent used measures of behavior" (Dixon, 1990, p. 1). "It is probably safe to say that the bulk of training programs conducted in

the United States are evaluated only at Level 1, if at all. Of the rest, the majority are measured only at Level 2" (Gordon, 1991, p.21).

In a study reported by the American Society of Training and Development, "Behavioral change on the job was the least measured: among companies surveyed, only about 10 percent evaluated training at this Level. Employee training was only evaluated at the results Level about 25 percent of the time" (Carnevale & Schulz, p. s-24).

Robinson and Robinson's report (1989, p. 170-171) breaks evaluation out by the percentage of courses using each Level of evaluation in relation to the percentage of training managers using that Level of evaluation. In this report, only 22 % of training managers use Level 2 in more than 80 % of their courses and only 9 % use Level 3 in more than 80 % of their courses. Only 10 % of the managers fail to use Level 2 at all, and 31 % fail to use Level 3 in any of their courses. All three surveys, discussed above, support the idea that how much evaluation is done at each of Kirkpatrick's four Levels of evaluation (1975a,b,c,d) decreases, moving from the most at Level 1 to the least at Levels 3 and 4.

Problems with Available Data

Not all the literature agrees with the surveys discussed above. The Corporate HRD Executive Survey of the American Society of Training and Development in their Survey #11 Report (1989) said that for technical training, only 57 % of the companies surveyed used participant reaction forms. "This report is based on 106 responses from

Fortune 500 companies and private companies with sales of \$500 million or more in sales" (1989, p. cover). This report listed Level 3 evaluation not at the 10 to 15 % found in other literature, but 31 % using performance records and 29 % using supervisor feedback (Survey #11, 1989). Phillips (1991) discussed a study in which only 52 % of the companies measured participant satisfaction, 5 % measured the skills acquired after a learning experience, 17 % measured application of skills on the job, 13 % measured changes in the organization's functioning, and 13 % did no systematic evaluation. The American Society of Training and Development Survey #11 Report (1989) and the study discussed by Phillips (1991) are numerically different from the other studies discussed. However, all the studies report less than 50 % of the companies perform evaluation above Level 1 and most studies place the use of Levels 2, 3, and 4 at approximately 25 %.

The information available reports consistently low rates of evaluation but does not establish how much evaluation is being done in business and industry in a form such that additional research can be based on the reported data. The indications gleaned from references to unavailable surveys are that very little evaluation beyond Level 2 is done. In a phone discussion, Kristey L. Husband, the Project Assistant in charge of the research for Survey #11 (1989), said that a sample of convenience was used (personal communication, October 13, 1992). Robinson and Robinson's report (1989) is also based on a sample of convenience. The sample was 150 HRD managers and directors at the Training Director's Forum, sponsored by Lakewood Publications'

Training Director's Forum Newsletter and Training: the Magazine of Human Resources Development. This may make the information in these surveys generalizable only to the drawn samples. Robinson and Robinson (1989) discuss the problem of representativeness in their report. Information on the survey methods used in the other studies discussed was not available. One set of data was traced to one of the two original presenters. However, neither the presenter nor the organization for whom the presentation was made had copies of the information presented. The search for the other presenter failed. The evidence available suggests a significant lack of evaluation within business and industry. However, no study using a random sample of training or specific area of training was found.

Importance of Training

The importance of evaluation is based on the importance of training. Training is an important tool in making a company competitive, for upgrading the skills required for new technologies, and for keeping the workforce employable.

Competing globally

Training will allow us to hold our lead over the rest of the world in productivity. This lead makes it possible for us to maintain a standard of living that is at least equal to that of the other major industrial nations of the world. Global competition is increasing. To maintain the highest standard of living possible, America must compete ("Skills", 1990).

Upgrading skills

High paying jobs within the technical sector require the constant upgrading of employee skills. Job skills are becoming obsolete at an increasing rate. The future described in Alvin Toffler's Future Shock (1970) is here today. Workforce 2000 (Hudson Institute, 1987) describes this phenomenon in detail. Today's industry is constantly increasing technical knowledge required by its workers and the number of positions requiring technological skills is rapidly increasing. Skilled workers, a fifteen million member segment of the population (Carnevale, Gainer, Schulz, 1990), must be either trained or retrained to address the changing technological needs of industry and construction.

The technologies presently used in these areas are changing rapidly or being replaced by new and different materials, processes, and equipment. These changes place new requirements on skilled workers employed in all areas. "Technology will introduce change and turbulence into every industry and every job. In particular, the necessity for constant learning and constant adaptation by workers will be a certain outgrowth of technological innovation" (Workforce 2000, p. 37). Without additional training, today's workers will no longer be employable except in low paying, low skill jobs ("Skills", 1990). No matter what agency or method is used, workers must learn new skills, accumulate the necessary knowledge, and apply the skills and knowledge gained in a new work environment or lose their jobs.

Worker benefits

Training also provides direct benefits to the worker. Not only do America and its industries benefit from training which upgrades knowledge and skills, but employees also benefit personally. "Ultimately because of the growing importance of skill and its general applicability across institutions, workers who pay attention to education, training and work experience can increase their control over their working lives" (Carnevale, 1991, p. 140).

Workers gain independence based on the skills provided by training. They gain not only financial independence based on their increased value to an employer, but as the quantity of training increases, their ability to work at various tasks also increases. This allows the worker a choice of positions (Carnevale, 1991).

Technical Training

Evaluating training for those whose work produces objectively measurable outcomes, such as technical or clerical employees, lends itself to quantitative evaluation. Technical workers produce goods, construct physical works (public or private), repair machines, operate equipment, and develop computer programs. All of this work produces objectively measurable outcomes. The ability to do each of these jobs can be evaluated using objective outcomes. Outcomes are usually a product such as an automotive part or a typewritten page. However, the outcome could be a machine returned to service after a failure. This product can still be effectively measured by measuring the operation of the machine.

Many clerical workers such as secretaries, bookkeepers, data entry personnel, bank tellers, etc., receive technical training in the use of computer software and various office machines. The ability to use software and machines on the job can be objectively measured.

Technical training is designed to produce objectively measurable changes in skills and knowledge. The costs and benefits of this training are important to the nation, industry, and the individual participants. To ensure that costs are minimized and benefits maximized, evaluating the outcomes of training is necessary. Technical training is well suited for evaluation because outcomes can easily be objectively measured.

Summary

Training is an important tool in keeping companies competitive. Evaluation is an important tool for developing and maintaining effective and efficient training programs. Evaluation can help justify training expenditures and provides the information required to decide what type and how much training is required to maintain company functions. The available literature says that very few organizations use all four Levels of the Kirkpatrick model for evaluation. Finally, evaluation of technical training should be straightforward because of its objective outcomes. This poses a two-part question. Is there a lack of evaluation in technical training where the process should be straightforward and if there is a lack of evaluation what are some barriers to evaluation?

Problem Statement

To what extent are each of the four Levels of evaluation, as defined by Kirkpatrick (1975a,b,c,d), used to evaluate technical training in business and industry? What methods of evaluation are used at Levels 2, 3, and 4? Additionally, is there a relationship between selected demographics, training practices, and the value placed on training by training managers and what Levels of evaluation are used? Finally, what are some impediments to the use of evaluation at all Levels?

Objectives

1. Determine the extent to which business and industry are using evaluation Levels 1, 2, 3, or 4 to evaluate technical training, based on the reported percent of the respondent's programs using each Level of evaluation.
2. In instances where Level 2, 3, or 4 evaluations are used, determine the percent of programs using each evaluation method commonly described in the training literature and what other methods are used, based on the percent of the respondent's programs using each commonly described method of evaluation and any additional methods provided by the respondents.
3. Identify possible reasons for not using a given Level of evaluation based on the percent of respondents reporting each reason.
4. Determine if relationships exist between selected industry demographics and the percent of programs using each evaluation Level and determine if a relationship exists between selected industry demographics and the percent of

programs using each evaluation method. The selected demographics are these three variables: a business' or industry's function, the number of people employed, and the number of individuals trained per year.

5. Determine if a relationship exists between the training manager's reported perception of the importance of a Level of evaluation to selected organizational functions and the percent of programs evaluated at that Level. The functions to be studied are: improving training, gaining upper management's support for training, and reaching organizational goals.
6. Determine if a relationship exists between a training manager's experience in training and the percentage of programs using each Level and method of evaluation.
7. Determine if a relationship exists between the length of time the technical training program has been in existence and the percentage of programs using each Level and method of evaluation.
8. Determine if a relationship exists between organizational training practices:
 - a. types of training programs
 - b. when training evaluation is planned
 - c. to whom evaluation results are reported
 - d. why training is done
 - e. the percent of training personnel involved in evaluation
 - f. the percent of training staff trained in evaluation

- g. evaluation techniques used that match standard research techniques
- h. the relationship between funding and program evaluation
- i. the overall funding of the training function and the following variables
- j. the percentage of use for each Level of evaluation
- k. the methods of evaluation used
- l. the reasons for not evaluating
- m. the perceived value of each Level of evaluation.

Significance of Study

The effectiveness of training is a major issue. Donald Kirkpatrick (1994) provides three basic reasons for the importance of evaluation: to justify the existence of the training department by showing how it contributes to the organization's objectives and goals, decide whether to continue or end training programs, and gain information on how to improve future training programs. Each of these reasons focuses on the needs of the training department and the organization that it supports. Evaluating possible methods of addressing a performance problem requires a selection based on the ability of each possible intervention to address that problem. When choosing the best method to address a performance problem, the effectiveness of training must be compared with the effectiveness of other solutions. During tests of new methods or techniques for training to address a performance discrepancy, their value can only be assessed based on changes in job performance (Mager & Pipe, 1970). Specifically, training is used to correct a performance discrepancy; the

outcome of training should be a change in performance. Without Level 2 and Level 3 evaluations, there is no objective basis for choosing interventions to address performance discrepancies, whether they be instructional methods or other techniques. Without Level 4 evaluation, an organization cannot make informed decisions concerning the value of training to the organization's function or profitability.

The literature on the use of training evaluations does not provide a clear picture of how the available models for evaluation are being used in technical training. Much of the data quoted in articles on this subject are from research surveys that are unpublished and generally unavailable. Many articles quote figures from sources that are at least once removed from the actual data. The original figures come from sources such as presentations to professional organizations where only the results were discussed. Most of this data and the methods under which it was collected are unavailable for professional review. This researcher has followed several references to their source only to find that the data, notes, and records of that meeting are not available from either the presenters or the supporting organization.

The training literature is replete with books and articles discussing the value of evaluation, the need to evaluate, and the need to prove the value of training. As noted above, these articles refer to studies showing a general lack of evaluation. This study attempts to identify the degree of usage of each of Kirkpatrick's four Levels of evaluation (1975a,b,c,d) in technical training, identify impediments to implementing Level 2, Level 3, and Level 4 evaluations, and describe the organizational

environment for evaluation. Additionally, the study will make exploratory correlations between the percentage of programs using each Level or method of evaluation and the organizational evaluation climate to generate a base of information to support further study.

The data collected in this study provides a base for the further study of evaluation in technical training and in other subject areas. This is important to the country, business and industry, training organizations, and participants in training programs. What is important to the country is the effective training of the workforce to make it globally competitive. The standard of living in the United States could decline without effective workforce training ("Skills", 1990). What is important to business and industry is evidenced by the ever increasing call to show the effect of training on organizational goals (Carnevale and Schulz, 1990; Dixon, 1990; Kirkpatrick, 1994; Phillips, 1991). What is important to training organizations is program improvement, program evaluation, and justifying the capital investment in training (Cascio, 1982; Carnevale and Schulz, 1990; Dixon, 1990; Kirkpatrick, 1994; Phillips, 1991). Finally, the importance to participants in training programs is that their personal value to the organization and their ability to earn are tied to the knowledge and skills they acquire ("Skills", 1990).

Assumptions

These assumptions form the basis for the idea that training is important enough to be measured and the selection of the accessible population for this study.

1. The training function is used to improve or maintain performance.
2. The training function is provided at a cost to the organization and that organization expects a return on its investment.
3. The training is provided at a substantial expense and capital expended makes the results of training a significant issue for any organization that provides training to its members.
4. The competitiveness of today's business environment makes training a necessary function.
5. The population surveyed will be better informed and more actively evaluate than the total population of organizations involved in technical training since the American Society for Training and Development's journal has presented more articles on the need for evaluation than any other over the last twenty years (based on a literature search using the ERIC and Psylit data bases and on actual articles located by the researcher).
6. The organizational evaluation climate in the organizations who are members of the selected population may be more positive toward evaluation than for all organizations providing technical training since the ASTD's journal has presented more articles on the need for evaluation than any other over the last twenty years (based on a literature search using the ERIC and Psylit data bases and on actual articles located by the researcher).

Limitations

1. The accuracy of this study was limited to the accuracy of the data supplied by the respondents.
2. The survey questions assumed a basic knowledge of industry evaluation practices.
3. The effort and cost of becoming a member of ASTD would show an interest in the training function. Therefore, the results of this study may overstate the actual amount of evaluation carried out by nonmember organizations providing technical training. Comparison data on the reasons for not evaluating courses or programs is unavailable. This study described the respondents' reported reasons for not evaluating but was unable to draw any comparisons with other groups. The data gathered provides only initial information in this area and additional study will be required to describe the reasons for not evaluating further and those reasons' relationship to the technical training function.
4. Information on the relationship between organizational evaluation environment and evaluation practices is nonexistent within the literature. Jack Phillips (1991) in his Handbook of Training Evaluation and Measurement Methods produced a test asking multiple choice questions about evaluation practices. The results of this test are supposed to score a company on how well they evaluate and how well evaluation is linked to results. No available data or

research theory supports the connection between these variables and the type or amount of evaluation done. Therefore, only non-directional exploratory correlations can be drawn between actual evaluation practices and measures of an organizational evaluation climate.

5. There is subjective value to technical training and in training that has specifically stated subjective goals. Technical training was chosen as the focus of this study because it produces objective outcomes that can be readily measured. The value of subjective outcomes is recognized but was not addressed in an effort to study evaluation of training outcomes that were more easily measured.

Definitions

The following working definitions are for the specific use of the words and abbreviations used in this study.

ASTD (The American Society for Training and Development) is an association of training professionals with a membership of 26,344 at the time when the population for this study was selected. This organization publishes both a magazine (Training and Development) and a journal (Human Resource Development Quarterly).

Effective programs produce the desired change in behavior in its participants.

Efficient programs produce the desired change in behavior in its participants at the least cost in capital and time.

HRD (human resource development) "The discipline charged with the development of people, processes, and organizations so that all three may contribute to improved organizational effectiveness and success" (Wimbiscus, 1995).

Kirkpatrick's four Levels is the basic model for evaluation in business and industry based on four articles written in 1959 for Training and Development Journal. The four Levels are reaction, results, on-the-job performance, and organizational outcomes.

ROI (return on investment) is the ratio between the cost of a program and the value of its outcomes (Brinkerhoff, 1991).

Technical employees are employees that use "principles from mathematical, physical, or natural sciences in their work" (Carnevale, Gainer, & Schulz, 1990, p. 2). Technical workers produce objectively measurable outcomes based on numbers or things. Technical workers include skilled craftspeople, computer programmers, journeymen of all types, production workers, and others.

Technical training is any training producing changes in knowledge or skills or both which are required to design, build, operate, maintain, or modify the software and hardware used in business and industry. The term hardware as used above is not restricted to computer hardware but includes such items as manufacturing machinery, construction or transportation equipment, testing devices (devices for measuring operating variables on machines such as voltage, pressure, or flow), office machines, and technical medical equipment (Carnevale, Gainer, & Schulz, 1990).

Training consists of formal or informal activities that produce changes in a participant's skills, knowledge, or attitudes that directly impact on present job performance or job performance required to enter a new position.

CHAPTER 2

REVIEW OF LITERATURE

Introduction

This review of literature is restricted to evaluation used in business and industry. This narrow approach to reviewing the literature was taken because the focus on evaluation in education and for publicly funded education programs is different from that of training evaluation in business and industry. Business and industry generally focus on summative measures and neglect formative measures.

These areas do have the same foundations. "Based on his experience in the Eight Year Study of the 1930's, Ralph W. Tyler proposed that educators should carefully define their objectives and gather data needed to determine whether they had been achieved" (The Joint Committee, 1981). This is still common across learning situations. However, the large number of new programs and materials developed in the 1960's were often accompanied by requirements for evaluation. This was especially true in the case of congressional funding. These programs were evaluated not only for results, but program content, management, and the types of students served by the programs (The Joint Committee, 1981).

This broad-based view of evaluation is echoed by Finch and Crunkilton (1989, p. 273) "evaluation may be defined as *the determination of the merit or worth of a curriculum (or portion of that curriculum). It includes gathering information for use in judging the merit of the curriculum, program, or curriculum materials.*" Business

and industry also evaluate programs and materials. However, their focus in most evaluations is on measuring a program's effect on (a) the participants, (b) the participant's work, and (c) the organization (Brinkerhoff, R. O. 1991; Broad & Newstrom, 1992; Dixon, 1990; Kirkpatrick, 1994; Phillips, 1991).

Introduction to Kirkpatrick's Model

The model most cited in the literature on evaluating industrial training is Kirkpatrick's (1975a,b,c,d). "Almost every discussion of training and development evaluation begins by mentioning Donald Kirkpatrick's well-known four-levels of evaluation" (Medsker & Roberts, 1992, p. 1). Kirkpatrick's model divides training evaluation into four Levels or steps. The first Level "reaction may well be defined as how well the trainees liked the program" (Kirkpatrick, 1975a, p. 1). This Level is commonly evaluated using participant reaction surveys at the end of a program.

Level 2 evaluates the learning that took place during training. "Learning is defined in a rather limited way as follows: What principles, facts, and techniques were understood and absorbed by the conferees? In other words, we are not concerned with the on-the-job use of these principles, facts, and techniques" (Kirkpatrick, 1975b, p. 6). Second Level evaluation consists of various types and styles of posttests.

The third Level of evaluation measures "changes in behavior on the job" (Kirkpatrick, 1975c, p. 10). Kirkpatrick takes pains to explain that "evaluation of training programs in terms of behavior is more difficult than the reaction and learning evaluations" (Kirkpatrick, 1975c, p. 10). He goes on to abstract nine studies as

examples of methods that may be used at this Level of evaluation (Kirkpatrick, 1975c).

The fourth Level of evaluation is the organizational results Level. This is an attempt to measure actual organizational changes due to training and place a dollar value on those changes. Kirkpatrick offered little as to methodology for this Level of evaluation. He presented abstracts discussing six cases where efforts toward this Level of evaluation had been made. He was prophetic in his discussion of future attempts to evaluate at this Level: "In years to come, we will see more efforts along this direction and eventually we may be able to measure human relations training, for example, in dollars and cents" (Kirkpatrick, 1975d, p. 17).

Elaborations on Kirkpatrick's Model

Participant reaction

Kirkpatrick saw participant reaction evaluation as important for three reasons: (a) management decisions on whether to continue funding training programs are often made based on comments from the participants, (b) participants can provide information that would help to improve programs (Kirkpatrick, 1975a,b), and (c) participants "must like training to receive the maximum benefit from it" (Kirkpatrick, 1975a, p. 4). The literature supports the first reason: managers make decisions based on participant comment. "If the true purpose of a training program is to reward good performers or renew sagging spirits at company expense, an extensive performance

based training evaluation is misguided. A simple reactions measure, or 'smile sheet,' may be all that is really necessary" (McEvoy & Buller, 1990, p. 40).

The second reason, improving programs, may be viable only in the sense that it supports the first. If increasing the participants' enjoyment of the program does not negatively affect the program's effectiveness or efficiency, such changes can be seen as improvements.

Jones in his "list of 26 limitations of end-of-course ratings" (1990, p. 20) lists as number one: "ratings don't correlate with transfer of training. No available research shows a clear relationship between end-of-course ratings and the extent to which participants apply training on the job" (Jones, 1990, p. 20). This is supported by a 1990 Dixon study where

... over 1,400 employees in a large manufacturing organization participated in one of three courses related to the implementation of a new manufacturing process . . . Results of the performance measures were correlated with each participants perceptions. For the first variable, 'amount of new information learned,' correlations ranged from -.07 to -.18, indicating no significant relationship between how much participants said they learned and how well they actually did on the performance measure. (Dixon, 1990, p. 29)

Participant reaction forms provide information that may be used to make the learning process more enjoyable and fulfilling. They do not evaluate training as to its effectiveness. "Studies of the relationship between actual learning achieved in a course and how participants complete reaction forms indicate that such a relationship is either very small or nonexistent." (Dixon, 1990, p. 28)

Participant reaction forms provide information required to evaluate the participant's feelings about training; they provide information that is nice to know. Most training managers would like to know that the participants enjoyed a particular program. "What you're measuring with a happiness sheet, he . . . (Kirkpatrick) . . . says, is initial customer satisfaction with the training experience. . . . The sheet only becomes sneerworthy if you pretend it's telling you what is happening at higher levels of evaluation" (Gordon, 1991, p. 21). Participant reaction forms provide information on customer satisfaction with the training experience, not information as to the effectiveness or efficiency of the training. Level 1 evaluation does not garner information that can be used to evaluate training against its main goals: changing on-the-job performance and organizational outcomes.

Learning outcomes

Jack Phillips, in his chapter on evaluation design, discusses the most common form of learning outcome evaluation, the posttest (1991). Phillips discusses pretest-posttest designs and discusses validity issues based on testing effects and threats to internal validity. Even if this form of evaluation is well designed and addresses validity issues, the best an evaluation at this Level can hope to do is learn whether the direct objectives of the training program were reached.

Mager's book, Measuring Instructional Results, goes further than pencil and paper posttests. He states that measuring result requires that the objectives of the training must be measured directly. Skills must be demonstrated and knowledge

applied (1984a). If well-written objectives are the basis for the training program, this method will detect whether the training participant learned the skills and knowledge addressed by the training.

Attaining training goals is necessary but not sufficient to guarantee that the goals of a program are met. "Instructors tend to think that if participants have mastered a skill during the learning event, they are adequately prepared to implement it on the job. . . . However, research on the transfer of training does not support the view that the training adequately prepares participants to transfer the skills to the work place." (Dixon, 1990, p. 90-91)

Behavior changes

The direct goal of training is a behavior change. Level 2 evaluation, at its best, assures that the worker has the skills and knowledge to perform a behavior on the job. Level 2 can only assure that the skills and knowledge to perform a behavior on the job have been learned. It cannot assure that the worker: (a) will have an opportunity to perform a behavior, (b) know when to use the learned behavior, or (c) will use the behavior even if the opportunity is recognized. Level 3 evaluation determines if the learning that took place is applied in the workplace.

Once it has been determined that the purpose of training is truly work related and that the desired outcomes are more substantive than symbolic and more external than internal, the question becomes how to measure the changes back on the job. (McEvoy & Buller, 1990, p. 41)

The question is how can on-the-job behavior be measured? The answer is that objective changes in behavior can be measured.

One of the most straightforward ways to measure changes in performance due to training is to use existing documentation. Documentation of output, quality of output, waste, time to complete a specific job, uptime of machinery, and other like measures provide a source of information to detect changes in performance. They each can be used as a direct measure of change in job performance (Phillips, 1991; McEvoy & Buller, 1990; Bernardin & Beatty, 1984). If this data is already being collected, the cost of evaluation is only the cost of reducing existing data to a usable form (Phillips, 1991).

The types of measures listed above were used to report "the following results:

- ▼ Customer service increased by up to 98 percent.
- ▼ Scrap fell by 58 percent.
- ▼ Cost of quality decreased by 54 percent.
- ▼ Total inventory decreased by 73 percent.
- ▼ Customer returns decreased by 91 percent" (Vanpelt, 1992, p. 11).

The expense of gathering evaluation data shown above was not directly attributable to the training effort itself. The data was gathered as part of an overall quality improvement program. Therefore, the expense of data collection for training evaluation can be reduced using existing data.

As part of an overall training effort, evaluation based on objective standards can be added as part of maintaining quality. As part of a total plan to improve production, "Conner also includes skills verification as part of the regular production process audit procedure - a procedure in which quality assurance auditors monitor the production process and check work performance." (Cocheu, 1990, p. 26-27) Here objective data are collected to measure the application of skills and knowledge gained directly from training.

Objective data on performance changes are susceptible to at least two sources of error. The first is opportunity bias. Was the worker provided with an opportunity to use newly gained skills and knowledge? The second is the possibility of erroneous data. Data of this type are subject to errors in both collection and compilation (Bernardin & Beatty, 1984).

The first type of error, opportunity bias, can be addressed by questionnaires that assess the extent to which the employee was given an opportunity to change their performance. This information can then be used to develop a weighting factor applied when analyzing data. Weighting is used to balance differences in opportunity (Bernardin & Beatty, 1984).

All data should be checked for accuracy. Methods by which such data are collected should be scrutinized. Besides data collection techniques, data processing should be examined for its potential to produce errors. Objective data should not be accepted at face value. It should be examined for bias and errors.

Interviews (formal and informal), focus groups, and surveys can be used to gather information on the effects of training (Dixon, 1990). Interviews and focus groups are discussed by several authors as viable methods for collecting data on the use of skills and knowledge gained in training (Darraugh, 1991; Bushnell, 1990; Dixon, 1990; Giusti, 1990; McEvoy & Buller, 1990).

Anecdotal information can be used to study the effect of training on performance. Supervisors or participants can collect anecdotal evidence or this evidence can be gathered using interviews and open-ended questions on surveys (Dixon, 1990). Anecdotal information provides qualitative information showing that the program is working. Often this is all the evaluation that the course or program gets. "We get some anecdotal evidence at Level 3. That's about it" (Gordon, 1991, p.21).

Questionnaires and structured interviews provide information on the use of skills and knowledge learned during training (Dixon, 1990). The questionnaires may be sent to managers, participants, and subordinates (Russ-Eft, 1992). Using two sources in gathering evaluation information through questionnaires gives a measure of concurrent validity.

Dixon describes the types of questions that need to be asked for the results of a survey to provide usable data. "Questions need to be pointed toward the specific intellectual skills involved, such as, 'in the last three months how many histograms have you constructed? Please state the title of each.'" (Dixon, 1990. p. 100). This is

supported by Connolly: "trainers must skillfully design the questions to elicit specific examples of behaviors that respondents perceive to be a direct result of the training." (1991, p. 45).

A more time consuming but direct method of evaluating the use of training on the job is observation. Where it is ethically feasible, the evaluator can observe, either overtly or covertly, the behavior of employees who have been through a training program (Darraugh, 1991; Dixon, 1990; Callahan, 1986). The observations " . . . require skilled observers who can rate behaviors identically" (Darraugh, 1991, p. 8). This data can only be considered objective with the broadest of definitions and specially trained observers (Bernardin and Beatty, 1984).

Performance appraisals are another method of evaluating the effect of training on performance. Performance appraisals are structured rating surveys of various types used by managers to rate employees (Bernardin & Beatty, 1984). Since these data are often collected independent of the training function, they provide a cost-effective method of determining training effects.

Bernardin and Beatty (1984) discuss legal, ethical, and methodological issues involved with the results from performance appraisals at length. "Even the toughest critics of performance appraisal stop short of recommending abolition . . . a strong minority voice holds that the faults are systemic but human." (Zemke, 1991, p. 37-38) "Donald Kirkpatrick, author of the 1982 book How to Improve Performance Through Appraisal and Coaching, blames unhappy appraisal experiences on people who ask the

process to handle two contradictory functions, salary administration and performance improvement" (Zemke, 1991, p. 38). Performance appraisals can provide information on the transfer of training but should be used carefully, recognizing their limitations. The data from performance appraisals is dependent (a) on the quality of the rating surveys used, (b) the ability of the raters, and (c) the problems with using subjectively gathered data -- reported quantitatively (Bernardin & Beatty, 1984).

Another method of evaluating training is a work simulation. The method can be used as a Level 2 evaluation when applied at the end of training. Work simulation tests, administered after the employee has had a chance to use knowledge and skills on the job, is Level 3 evaluation (Phillips, 1991). Simulations provide a controlled environment in which a contrived problem addresses the material from the training course directly. Simulation can filter out extraneous variables (Ostroff, 1991).

All behavioral change measures are subject to one major problem. They can be contaminated by extraneous variables. This fact should be recognized and adjusted for where possible (Phillips, 1991). "Although measurements may not be precise, they are better than no measurement of change at all." (Phillips, 1991, p. 4)

Organizational change due to training

"From an evaluation standpoint, it would be best to evaluate training programs in terms of the results desired . . . The results would be classified as: reduction of costs; reduction of turnover and absenteeism; reduction of grievances; increase in quality and quantity of production; or improved morale which, it is hoped, will lead to

some of the previously stated results"(Kirkpatrick, 1975d, p. 14). "Accounting for the positive economic influence of training and development is the most critical issue in the training profession today" (Carnevale & Schulz, 1990, p. s-2). The preceding statements were originally made 30 years apart, but both reflect the interest in the training establishment to be able to back claims about the value of training with hard monetary figures.

When Kirkpatrick was making his first contributions to the field of evaluation, there were few methodologies in place to measure the value of training at this Level. "At the present time, however, our research techniques are not adequate" (Kirkpatrick, 1975d, p. 17). The newest methods of calculating results, ROI, or utility (measures of the value of training to an organization) involve complex estimates of multiple values. The estimated values are then used in a utility equation to provide a range of possible values for training. How well a utility formula reflects the actual value of training is dependent on the accuracy of the estimated values (Cascio, 1982). The process of calculating results has evolved since Kirkpatrick's article in 1960, but it has not reached the point where calculating the value of training in dollars is easy.

Cascio offers one method for calculating the total benefit of training illustrated by the equation $B = (N)(T)(dt)(SDy)$, where B is the benefit, N is the number of employees to be trained, T is the duration of the benefit, dt is the change due to training in standard deviations, and SDy is the value of the change for each standard deviation from the norm (1982). The number of employees to be trained is

the only value in the formula that is simple to acquire. The duration of the benefit is confounded by the estimating length of time a procedure will be in place and the length of time an employee will be retained by the company. The difference in employee output due to training is relatively simple to find in circumstances where the output can be measured objectively, such as in units produced, inspected, or repaired. Difference due to training (dt) is based on the outcomes of Level 3 evaluation. The value of a standard unit of deviation (SDy) is at least as difficult to quantify as dt , if not more difficult (Cascio, 1982).

The approach that is often taken is to quantify all the values except SDy and then work the equation backwards to find the minimum value of SDy that will produce an acceptable ROI based on benefit minus cost. Several different methods are used to estimate SDy and provide a range of possible values for SDy . The range can then be compared to the minimum acceptable SDy , and decisions may be made based on the relationship of the range of estimates to the minimum acceptable value of B (Cascio, 1982).

Evaluation at the organization results Level has the same problem as it does at the behavioral Level: the problem of extraneous variables. This is not surprising since benefit (B) is calculated using increments of behavioral change times the estimated incremental value of that change. Kirkpatrick (1975d, p. 17) recognized these problems in his early work. In a few of them (ROI studies), the researchers have attempted to segregate factors other than training which might have had an effect.

Usually, the measure on a before and after basis has been directly attributed to training although other factors might have been influential.

Evaluation Use

Level 1

"It's probably fair to say that the bulk of all employee training programs conducted in the United States are evaluated only at Level 1, if at all. Of the rest, the majority are measured only at Level 2" (Gordon, 1991, p. 21). This quotation sums up what can be found in the literature. Very little thorough training evaluation is accomplished across American industry. (Dixon, 1987; Phillips, 1991; Robinson & Robinson, 1989)

Of the evaluation done, "the vehicle that is most frequently used to accomplish this evaluation is the participant reaction form" (Dixon, 1987, p. 108). "Most training courses include end-of-course questionnaires that ask participants to rate various aspects of the experience" (Jones, 1990, p.19).

In a 1987 survey of 150 training directors, selected because of their attendance at a training directors' forum, 77 % of the training directors reported using Level 1 evaluation in 81 to 100 % of their courses and 97 % reported using Level 1 evaluation to some extent in their training courses (Robinson & Robinson, 1989). This Level of use is far above that of the other three Levels of training evaluation.

In technical training, even Level 1 evaluation may be neglected. Survey #11 (1989), the Corporate HRD Executive Survey of the American Society of Training

and Development, reported that in technical training only 57 % used participant reaction forms. This survey reports a much lower use of participant feedback in technical skills training than in management training where the rate of participant feedback evaluation is 92 %.

Participant reaction to learning experiences is reported by all the surveys in the literature as the most common form of evaluation. Articles on the subject also agree that this is the most common form of evaluation.

Level 2

Robinson and Robinson (1989) report that only 36 % of the training directors surveyed use Level 2 evaluation in 60 % or more of their courses. The American Society for Training and Development's Survey Report #11 (1989) shows this type of evaluation in a very low percentage of programs for nontechnical training. However, for technical skills training it reports 45 % of the companies used this form of evaluation. In technical training where it may be assumed that objective measures of training would be most applicable, Survey Report #11 shows a greater amount of evaluation at Level 2 than do other surveys. Based on the surveys discussed above and references to other surveys in the literature (Carnevale & Schulz, 1990; Dixon, 1987; Phillips, 1991), Level 2 evaluation is not used as regularly as is Level 1.

Mager's Measuring Instructional Results (1984a) covers nothing but Level 2 evaluation and Phillips' Handbook of Training Evaluation and Measurement Methods, (1991) contains information on this Level of training evaluation in several chapters.

Both Robinson and Robinson (1989) and Brinkerhoff (1991) devote a whole chapter to Level 2 evaluation techniques. Information on developing tests of various types at this Level is readily available.

Articles on programs or courses using Level 2 evaluation are common. Five of 18 of the articles in the ASTD's The Evaluation of Training (American, 1991), a collection of articles from Training and Development and Technical & Skills Training covering four years, directly address Level 2 evaluation.

Level 3

Carnevale & Schulz (1990) describe the use of Level 3 evaluation as between 10 and 15 % based on two different surveys. A third survey reports that six to 13 % of training managers use Level 3 evaluation for 80 % of the courses taught. In the same report, 20 % of courses were evaluated at Level 3 by 31 % of training managers (Robinson and Robinson, 1989). Survey #11 (1989), the Corporate HRD Executive Survey of the American Society for Training and Development, reported that for technical training, Level 3 was performed using three different methods: post course tests, performance records, and supervisor feedback. The percentage of companies using each method was 37, 31, and 29 respectively.

Discussions in the literature cover at least seven distinct methodologies for evaluation at Level 3. (Dixon, 1987; Phillips, 1991) With this degree of knowledge about various techniques for evaluating, it seems that the literature should contain case studies of evaluations carried out at Level 3. No case studies of this sort were found.

Level 4

Based on the four surveys found in the literature (Carnevale & Schulz, 1990; Dixon, 1987; Phillips, 1991; Robinson and Robinson, 1989; Survey #11, 1989) covering the use of evaluation in industry, only two reports discuss evaluation at Level 4. Carnevale and Schulz (1990) report that "Employee training was only evaluated at the organizational result level about 25 % of the time, despite new pressures on training practitioners to assess the economic worth of HRD activities." Robinson and Robinson (1989, p. 171) report that for 80 % of the courses taught, less than 6 % of training managers use Level 4 evaluation and 59 % of managers never use Level 4 evaluation.

Most texts on evaluation discuss cost-benefit analysis or return on investment (ROI) calculations. These texts spend more time discussing methods of cost accounting for training than they do on estimating the dollar benefits of training. The same can be said for articles on cost benefit analysis. The most often referenced text concerning placing dollar values on the benefits of training is Cascio's Costing Human Resources: The Financial Impact of Behavior in Organizations (1982). This text is not a text on evaluating training but a text on placing dollar values on changes in behavior. Chapter 9, "Estimating the Economic Value of Job Performance ," explains in detail a method for placing an estimated range of value for changes caused by and impacting on an organization.

Cascio's Costing Human Resources: The Financial Impact of Behavior in Organizations (1982) describes an involved evaluation of the outcomes of two different methods for training operators of a plastics extruder. This is the only instance of the use of Level 4 evaluation in a technical training situation that could be found in the literature. This was an experimental study set up to evaluate two different methods of training. It was not an evaluation of ongoing training or an evaluation process developed for a training program. Reported instances of Level 4 evaluation in the literature are lacking.

CHAPTER 3

METHODOLOGY

Purpose of the Study

The purpose of this study was to examine business and industry's use of evaluation in technical training. Specifically, the purpose of this study was to describe the percent of technical programs using each of Kirkpatrick's four Levels of evaluation, identify the various evaluation methods used at each Level, and describe the percent of programs using each method. The literature suggests that Level 1 evaluation is commonly conducted and that the methods used to collect this data are well known. Therefore, data was not gathered on Level 1 methods. For Levels 2, 3, and 4, the methods used to collect data and the percent of programs using each method were identified. Selected relationships among industry demographics, organizational training practices, perceived values of evaluation, and evaluation practices were explored. Additionally, reasons for not evaluating at Levels 2, 3, and 4 were identified.

Population and Sample

The population for this study consisted of all U.S. organizations providing technical training. The American Society for Training and Development (ASTD) is the world's largest association of training professionals. Within the ASTD, each member joins a professional practice area. Technical and Skills Training is a professional practice area. Thus the accessible population was the 2,569 U.S.

organizations with ASTD members who were in the Technical and Skills Training professional practice.

To identify this population, all consultants, schools, and government organizations were removed from the ASTD mailing list for the Technical and Skills Training professional practice area. Schools and government organizations were removed because they are not members of business or industry. Consultants were removed because they are both businesses that train their own personnel and the personnel from other organizations. It was felt that they would provide a mixture of data from the two groups and would not be representative of business and industry. Population errors included some consultants among those surveyed because the business names and person's title did not provide information about the business' function. The data from these surveys were not used, but notes returned with the surveys supported the assumption that consultants were a special case and that the data collected from this group would include a mixture of internal and external training data.

When multiple members existed within an organization, the survey was sent to a single member. The first choice was the member listed as managing or directing the technical training function. Second choice was the highest position in the training function based on the following six positions in the order listed: (a) Vice-president for training, (b) director of training (and development), (c) manager of training, (d) administrator of training, (e) coordinator of training, (f) supervisor of training. Third

choice was the highest position within the HRD or HRM function based on the following six positions in the order listed: (a) Vice-president for HRD or HRM, (b) director of HRD or HRM, (c) manager of HRD or HRM, (d) administrator of HRD or HRM (e) coordinator of HRD or HRM, (f) supervisor of HRD or HRM. Fourth choice was the highest position within the organization based on the following five positions in the order listed: (a) President, (b) Vice President, (c) Director, (d) Manager, (e) Administrator. Fifth choice was position deemed most likely to be connected to technical training. For identical titles, the choice was made using a table of random numbers and the order in which the members are listed. Those with no title were used only if no other alternative was available. The organizations with no member's name listed were excluded from the sample, since there was no way to assure that each mailing went to the same person.

This accessible population was chosen because it should represent organizations in which those persons responsible for the technical training function are well informed as to the current practices in industry. As such, the population should report higher uses of evaluation than the total population. As low levels of use were expected based on the literature, this scheme provided a best case scenario of the use of evaluation in the technical training area.

A random sample was drawn from the accessible population. Sample size was 384 minus the small population factor to provide for a $p = .05$ based on Cochran's

formula $n_o = \frac{t^2(pq)}{d^2}$. This formula was used, where n_o = sample size, t = risk, d = acceptable margin of error, p = the proportion of the sample estimated to have the characteristic, and q = the proportion of the sample estimated not to have the characteristic. Since the variance of the population was unknown, p and q were set at .5 (the largest possible variation). This produced the largest n_o or sample size based on variance. If the sample has the largest possible variance and half have the characteristic being sampled and half are missing the characteristic, the results will still be accurate to $\pm 5\%$ with t set at 1.96 and d set at .5.

$$n_o = \frac{(1.96)^2(0.5)(0.5)}{(0.05)^2} = \frac{(3.846)(0.25)}{0.0025} = 384$$

Cochran's adjustment formula for small populations, as shown below, was used to calculate the adjusted sample size for this small population. The accessible population was 2,569, the total number of members in the technical and skills area. The required sample size for 95 % reliability was 334 as shown below.

$$n = \frac{n_o}{1 + \frac{n_o}{N}} = \frac{384}{1 + \frac{384}{2,569}} = \frac{384}{1.15} = 334$$

Since this was expected to be a reluctant population, the entire population was randomized which allowed for replacement of those unwilling to participate and for frame errors. In a similar study, Gutek's original response rate was 18 % and after phone contact with all the non-respondents the final sample consisted of 46 % of the

drawn sample (1988, p. 55-56). Schilling's original response rate from an unsolicited set of mailed surveys was 3 %. In that study, a non-random sample was used to increase the number of respondents (1991, p. 38-40). The drawn sample in this study had an n set using Cochran's adjustment formula for small populations.

Instrumentation

The data for this study were collected using a survey. This method was chosen as a practical method for acquiring the required data from a sample of this size distributed across the entire United States. Additionally, it allowed the respondents much more time flexibility in answering the questions than would telephone interviews. The wording of the questions in the survey is based on the training evaluation literature. The survey was authored in a joint effort including the researcher, Dr. Jack Phillips (a nationally known expert in the field of evaluation) and Dr. Ed Holton III (Associate Professor, Human Resource Development with published articles on evaluation). An effort was made to use the most common terms in the literature to insure that the terminology used in the questions would be as clear as possible to the respondents. The survey was subsequently reviewed by a group of experts including the members of a graduate class in research, academic researchers, training managers, training specialists, and two business and industry experts on training evaluation.

The survey consisted of seven sections. Sections 1 through 4 included questions on the percent of programs evaluated at each Level, types of evaluations

used, and reasons for not evaluating at each Level. Section five consisted of questions used to determine the training practices of the organization. The questions asked what methods of program delivery were used, when evaluations were planned, to whom evaluation results were reported, the purpose of training, number of personnel involved in evaluation and their training in evaluation practices, research practices used, and the relation of evaluation to budgeting. Section six gathered information on the respondents' thoughts on the value of evaluation to improve training, demonstrating the value of training in improving job performance, demonstrating the value of training to upper management, and demonstrating the value of training in attaining organizational goals. This section was based on earlier research that showed a relationship between training manager's thoughts on training evaluation and how much evaluation was done. Section seven gathered demographic data. Questions in sections five and six were used to gather information on evaluation practices that were not a part of Kirkpatrick's model but discussed by Gutek (1988) and Phillips (1991) as having an effect on the amount and types of evaluation carried out. (see appendix A.)

Data Collection Procedure

The following procedure was based on the portion of Dillman's TDM (Total Design Method) dealing with survey mailings (1978). A cover letter using ASTD letterhead, the survey, and a stamped reply envelope was mailed to each respondent. (see appendix B.) The survey was followed by a mailing of reminder cards to the entire sample one week after the initial survey was mailed. This postcard contained

three statements: notifying the respondent that a survey was sent, thanking them for responding, and asking them for a collect call if they had not received the survey. (see appendix B.)

Three weeks after the original mailing, a second mailing of surveys to the non-respondents was sent with a new cover letter explaining the importance of their reply to the quality of the results of this study. (see appendix B.) Seven weeks after the first mailing, a second follow-up letter was sent to all non-respondents. (see appendix B.) There were population errors and surveys returned as undeliverable. The population errors were misidentification of consultants. Because of these errors in the original mailing, additional surveys were sent in two mailing sequences matching that of the first mailing.

A random sample of 20 non-respondents was contacted by phone to have them respond to a selected sample of the questions from the original survey or to find out the reason for non-response. The data gathered were used to compare the non-respondents with the respondents to decide if the conclusions drawn about the sample could be extended to the accessible population.

A sample of 22 questions was drawn from the original questionnaire and used as the telephone interview guide for the sample of non-respondents. The questions were selected based on their importance to the objectives of the study. Members of the original sample of 20 were replaced only if they could not be contacted across a two-month period or if they refused to respond to the questionnaire.

The telephone interview guide consisted of 22 questions, 21 of which fell into four distinct groups: use of the four Levels of evaluation, training methods, evaluation implementation, and demographics. The question that fell outside the four groups described above asked about the relationship between program continuation and evaluation of that program.

Multiple t tests were used to compare the respondent responses with those of the non-respondents. Although some responses were ordinal in nature, it was felt that for the purposes of comparing respondents with non-respondents, the t test was robust enough to be used on this data. An alpha level of .05 was set a priori for significant difference between the respondents and non-respondents on each variable. Since it would be expected that at least some differences would be found as an artifact of a large number of comparisons, the respondents and non-respondents were considered different only if they differed on more than 10% of the tests or, in this case, if more than two significant differences were found. Appendix C contains the results of the comparisons of the respondents and non-respondents on the selected questions.

Data Analysis

The survey data were analyzed using frequencies, measures of central tendency, and variability, as appropriate. Exploratory correlations were drawn between selected demographic data, organizational training practices, and the use of various Levels of evaluation to describe relationships. Davis' (1971) scale for interpreting correlation coefficients was used as the basis for discussing correlations.

Davis' scale was chosen over the Hinkle, Wiersma, and Jurs scale (1988) because it is less restrictive and these were exploratory correlations. Using Davis' (1971) scale reduced the chance of undervaluing any relationships that might be found. The statistical functions provided in Microsoft Excel were used in the analysis.

A list of the specific statistical procedures used for each research objective follows on the list below. The number leading each procedure corresponds to the objective it addresses:

1. The extent to which business and industry are using evaluation Levels 1, 2, 3, or 4 to evaluate technical training was measured by percentages of programs using each Level. Means and standard deviations for each of the four Levels were reported.
2. The types of Level 2, 3, or 4 evaluation used were measured using percentage ranges for the number of programs in which each method was used. This information was reported using relative frequencies.
3. Where a Level of evaluation was not used, the reasons for not using that Level of evaluation were reported using the percentage of organizations reporting each reason.
- 4a. To determine if a relationship existed between the function of a business or industry and the Level of evaluation used, correlations were drawn between the functions and the use of each Level and method of evaluation as measured by the percentage of use. If the correlation coefficient was .30 or greater, the strength of the relationship was discussed using Davis' (1971) scale for interpreting correlation coefficients.

4b. To determine if a relationship existed between the size of a business or industry (as determined by the number of employees) or the number of employees trained each year and the evaluation used, correlations were drawn between "number of employees" and "number of employees trained" and the percentage of use for each Level of evaluation. If the correlation coefficient was .30 or greater the strength of the relationship will be discussed using Davis' (1971) scale for interpreting correlation coefficients.

5. To determine if a relationship existed between the training manager's perception of the importance of a Level of evaluation in improving training, gaining upper management's support for training, or reaching organizational goals, and the percentage of use of each Level, the correlation between the variables (perception and percentage) of use was calculated. The strength of the relationship was discussed using Davis' (1971) scale for interpreting correlation coefficients.

6. To determine if a relationship existed between a training manager's experience in training (based on years in the training function) and the percentage of programs using each Level, the correlation between these two interval variables (experience and percentage of use) was calculated using Pearson's r . The strength of the relationship was discussed using Davis' (1971) scale for interpreting correlation coefficients.

7. To determine if a relationship existed between the length of time the technical training program has been in existence and the percentage of use of each Level, the correlation between the variables (experience and percentage of use) was calculated

using Pearson's r . The strength of the relationship was discussed using Davis' (1971) scale for interpreting correlation coefficients.

8. To determine if a relationship exists between organizational training practices and how much each Level or method of evaluation is used. The organizational training practices were (a) types of training programs, (b) when training evaluation is planned, (c) to whom evaluation results are reported, (d) why training is done, (e) the percent of training personal involved in evaluation, (f) the percent of training staff trained in evaluation, (g) evaluation techniques used that match research techniques, (h) program funding, and (i) the overall funding of the training function. Correlations were calculated between organizational training practices and (a) the percent of programs using Levels 1 through 4 or (b) the percent of programs using various methods of evaluation. Correlations were drawn only where relationships could logically be expected. The correlations between the variables were calculated using Pearson's r . The strength of the relationship was discussed using Davis' (1971) scale for interpreting correlation coefficients.

CHAPTER 4

DATA ANALYSIS

Sample Size and Response Rate

Survey forms were sent to a total of 348 organizations with members in the ASTD professional practice area, Technical and Skills Trainers. Three hundred thirty-five were in the first mailing. Frame errors were found in the returns from the original mailing. The frame errors were consulting firms that could not be identified by their company name or any other information available at the time of the original mailing. The organizations' reported function on the returned surveys was used to identify respondents who were not part of the accessible population. Two subsequent mailings were sent to a total of 13 additional organizations. These organizations were selected to replace those in the original sample that were not part of the accessible population.

A total of 146 surveys was returned for an overall response rate of 42%. Three hundred twenty-two surveys were sent to members of the defined population (Due to frame errors 26 respondents were not part of the population as defined; thirteen were replaced by two additional mailings to maintain sample size; the other thirteen that were not within the population were identified after the second and third mailings.) and one hundred twenty-four were returned for a response rate of 39%. Of those responding, 12 could not provide the requested information or refused to participate. One hundred twelve usable surveys were returned making the response rate for usable surveys 35%.

Comparison of Respondents to Non-respondents

A sample of 22 questions was drawn from the original questionnaire and used as the telephone interview guide for the non-respondents. Sample questions were selected based on their relevance to the overall objectives of the study. Members of the original sample of 20 non-respondents were replaced only if they could not be contacted across a two-month period or if they refused to respond to the questionnaire. The total drawn sample of non-respondents was 80. Nineteen could not be contacted based on the population information available. Almost 38% could not be contacted after four calls spread across two months. The sample of 20 respondents that answered the questions during the phone survey made up 25% of the total attempts to contact non-respondents. (see Table 1)

Table 1

Non-respondent Sample

Number of respondents	Status
9	No phone listing could be found for the respondent or their organization
1	Refused to respond for lack of time
10	No longer with the organization
2	No longer involved in training
38	Not contacted after four calls
20	Responded to telephone survey

Note: A total of 80 attempts resulted in 20 responses.

The groups differed significantly on only two items (a) the percentage of programs in which evaluation was the first step in program development and (b) the percentage of programs dependent on evaluation for continued funding. Seventy-five of the 111 respondents reported that they never started the evaluation process as the first step in program development. Of the respondents that started evaluation as the first step, the median was for 1-19 % of their programs. The total non-respondent sample reported the evaluation process as the first step in program development for median of 1-19 % of their programs. Four percent of the respondent's programs were dependent on evaluation for funding and 35 % of the non-respondent's programs were dependent on evaluation for funding. Appendix C contains the results of the comparisons of the respondents and non-respondents on the selected questions.

Demographics

To determine the size, age, and type of organizations within the sample, the respondents were asked the following demographic questions about their organizations and themselves on the last page of the survey form:

- ▶ Number of employees working in the United States
- ▶ Number of employees working in the United States that participated in technical training last year
- ▶ Number of years your organization has been providing technical training
- ▶ Number of years you have been performing the training function in this or any other organization

- ▶ Gender
- ▶ Please check the industry group that best describes your organization

The statistics for the first four questions are shown in Table 2. Responses to all four questions had large deviations and small standard errors.

In response to the fifth question that asked the respondents gender, 57 of the respondents were female and 51 were male. Five respondents either left the answer to the gender question blank or entered a note refusing to answer.

Table 2

Descriptive Statistics for Demographic Data

Demographic	Statistic					
	<u>n</u>	<u>M</u>	<u>SE</u>	<u>SD</u>	Min.	Max.
Total Number of Employees	107	4420.93	797.04	8244.67	5	50,000
Number of Employees Receiving Technical Training	103	2533.61	556.79	5650.81	1	30,000
Years the Training Program Been Operating	105	21.89	2.37	24.30	1	150
Respondents' Training Experience in Years	109	11.00	.66	6.93	1	33

Size was based on the number of total employees and ranged from five employees to 50,000 employees. Figure 1 reports frequencies using size ranges. The ranges used were selected so that each group would represent approximately 10 to 20 % of the sample. The group size ranges were increased following a pattern to



Figure 1. Frequencies by size range for the 106 organizations that provided size datum

prevent odd increments such as seven to 137 and 138 to 703. When this was done, the first two groups were (a) less than 5 % of the sample, (b) the third group was 25 % of the sample, and (c) the fifth group was 22 % of the sample. The three groups with the highest frequencies covered the range from more than one hundred to less than five thousand employees and comprised 69 % of the sample. Typically these organizations should be large enough to have a well defined human resource development function.

Figure 2 presents the data on the membership of each responding company in their industry group. Figure 2 is based on an n of 93. However, there were 97 responses. Four respondents chose two groups. Three respondents chose

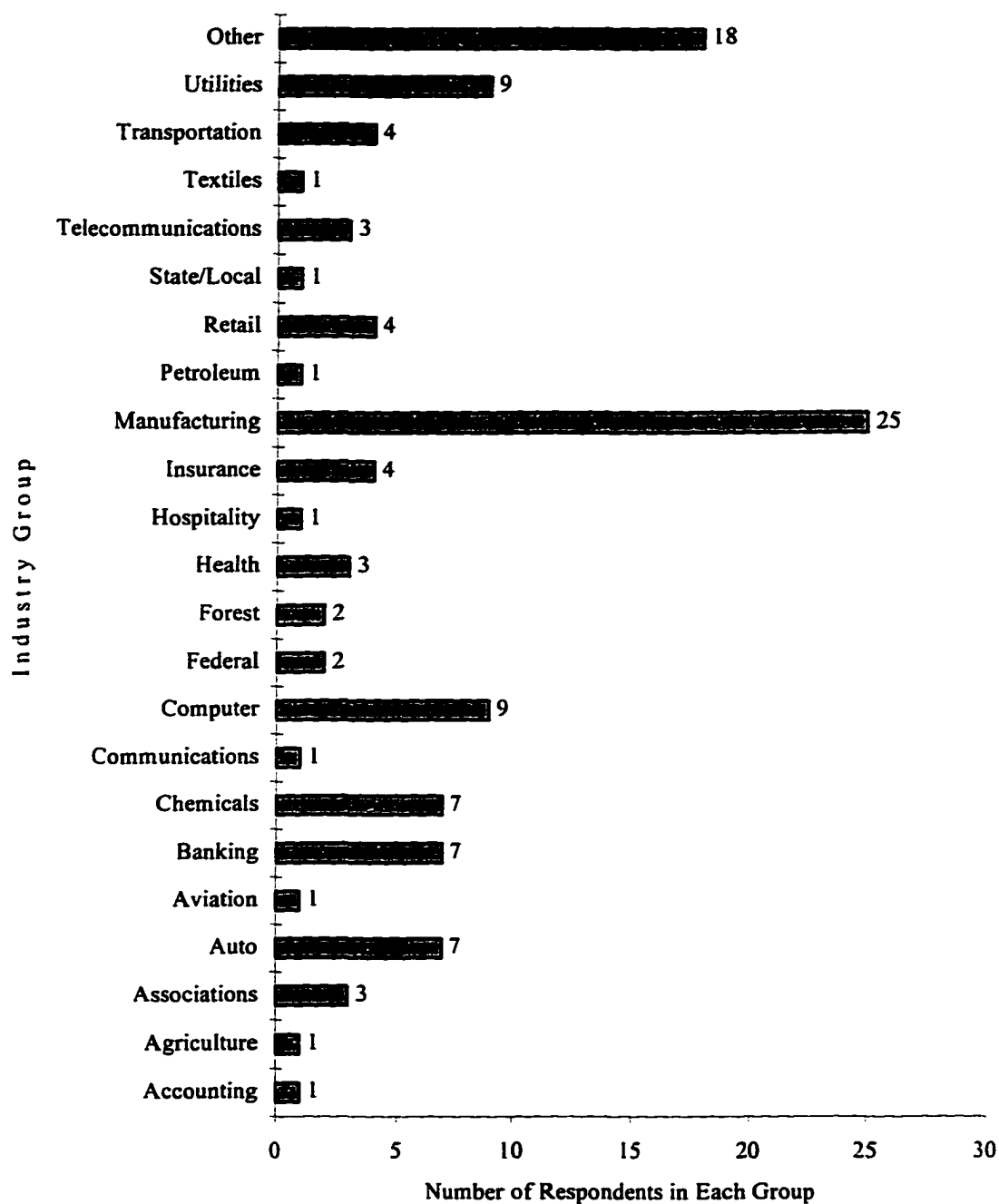


Figure 2. Frequencies for the 93 respondents who self-selected into the industry groups provided on the survey

manufacturing plus the function that further defined their organization's manufacturing. Those three functions were telecommunications, health, and agricultural. One respondent chose retail and automotive. The chart is based on both responses because the two responses provided more information than would have been available if only retail or manufacturing were included in the data. Manufacturing was the largest reporting group with 25 respondents. Six industry groups were self-selected by six (5 %) or more of the respondents: automotive, banking, chemicals and pharmaceuticals, computer - data processing, manufacturing, and utilities. Eighteen respondents chose other. Their responses included law firm, financial services, mining, software development, temporary help, document outsourcing company, copying, private and public rehabilitation, dry-cleaning, business repair services, electronic distribution, ready to assemble furniture, service, machine vision, construction, fire protection, and fluid power.

The Extent to Which Business and Industry Are Using Evaluation Levels 1, 2, 3, or 4 to Evaluate Technical Training

Objective one was to determine the extent to which business and industry were using evaluation Levels 1, 2, 3, and 4 to evaluate technical training. For each Level, the respondents were asked to estimate what percentage of currently active training programs were using Level 1, Level 2, Level 3, or Level 4 evaluation. They were provided with a space to enter the number. The questions were asked at the beginning of each of the first four sections of the survey instrument. Percentage of programs is a common measure found in the literature (Carnevale & Schulz, 1990; Robinson &

Robinson, 1989). The percent of organizations that use a Level of evaluation can also be found as a common measure in the literature (Dixon, 1987; Phillips, 1991; Survey #11, 1989). Percentage of programs was used as a measure because it could be reported as either percentage of programs or percentage of organizations using a Level. Using a Level in any percentage of programs greater than zero means the organization uses that Level. Responding organizations can be separated into ranges based on the percentage of programs that use a Level. These ranges allow data collected as percentage of programs to be reported as frequencies and percentages for organizations that use a Level: (a) in any of their programs, (b) over half their programs, and (c) most of their programs. Therefore, using percentage of programs as a measure of the amount an evaluation Level was used provided data which could be reported in different forms that could be better compared with most previous studies.

In order, the means for the percentage of programs reported as using each evaluation Level 1 through 4 are 72.74%, 47.05%, 33.73%, and 20.82% respectively. The standard error for these means ranged from 2.93 to 3.44 and the standard deviation ranged from 30.90 to 36.24. Table 3 presents the descriptive statistics for the percent of programs using each of the four Levels of evaluation. The number of organizations reported as using each Level of evaluation is presented as frequencies for zero and for five intervals covering the range from 1 to 100 % in Table 4. These intervals were chosen to match the intervals used on the survey for the percentage of programs using each evaluation method.

Table 3
Percent of Programs using Each Level of Evaluation

Statistic	Levels			
	1	2	3	4
Mean	72.74%	47.05%	30.54%	20.82%
Standard Error	3.36%	3.44%	3.20%	2.93%
Standard Deviation	35.40%	36.24%	33.73%	30.90%
Minimum	0%	0%	0%	0%
Maximum	100%	100%	100%	100%

Note. $n = 111$

Table 4
Frequencies and Percentages of Programs Using Each of Kirkpatrick's Four Levels of Evaluation for Selected Percentage Intervals

Intervals	1		2		3		4	
	f	% of n	f	% of n	f	% of n	f	% of n
0%	9	8.1%	18	16.2%	39	35.1%	52	46.8%
1-19%	8	7.2%	17	15.3%	16	14.4%	24	21.6%
20-39%	3	2.7%	13	11.7%	15	13.5%	11	9.9%
40-59%	11	9.9%	16	14.4%	18	16.2%	5	4.5%
60-79%	9	8.1%	15	13.5%	6	5.4%	7	6.3%
80-100%	71	64.0%	32	28.8%	17	15.3%	12	10.8%

Note: $n=111$ and $c\% = 100$ for all four levels

One common method for reporting the use of evaluation is the percentage of companies or organizations using each Level. When the data from percentage of programs using an evaluation Level is reported as the percentage of organizations using an evaluation Level, the definition of use greatly affects the outcome. The data collected in this study were separated into three use ranges. This was done by choosing three different minimum percentages of programs required to qualify a responding organization as using that level of evaluation. The organizations were separated into ranges using the following criteria:

1. organizations using the Level in any percent of their programs other than zero (some use)
2. organizations using the Level in 60 % or more of their programs (more than half)
3. organizations using the Level in 80 % or more of their programs (most)

Each range includes the respondents in the following range or ranges. This was done to match the following possible survey questions from other studies: (a) Do you use this Level of evaluation?, (b) Do you use this Level of evaluation in more than half your programs?, and (c) Do you use this Level of evaluation in most of your programs?

Sixty percent was chosen as a breakpoint because it matched the breakpoint used in the survey's questions on methods and helps maintain consistency in reporting. Following the same reasoning, 80 % was the natural breakpoint for "used in most programs" since it was the range for the highest amount of use in the survey's questions on methods.

Table 5 shows the proportion programs reported as using evaluation Levels 1 - 4. The decreasing proportion of programs using evaluation from Level 1 to Level 4 was independent of the breakpoint used to qualify an organization as using each Level of evaluation.

Table 5

Percentages and Frequencies for Organizations Using Each Level of Evaluation in Some, More than Half, and Most of Their Programs

Proportion of Programs	Level 1		Level 2		Level 3		Level 4	
	f	%	f	%	f	%	f	%
Some	102	92	93	84	72	65	59	53
More than half	80	72	47	42	23	21	19	17
Most	71	64	32	29	17	15	12	11

Note. $n = 111$ for all four levels

evaluation data at Level 1 are well documented. To reduce the length of the survey form and increase response rates, no questions were asked regarding Level 1 evaluation methods. For Levels 2, 3, and 4 the respondents were given a list of various accepted methods and asked what percentage of their programs used each method. The methods listed were based on a review of the literature and varied for each Level. The respondents were also provided with the opportunity to write in any additional methods they used which were not included on the provided list.

To make reporting these data as easy as possible, the respondents were asked to estimate the percentage of programs using a method and were provided with the following ranges from which to select; (a) 0%, The method is never used, (b) 1-19%, the method is used even if only in a few of the organization's programs, (c) 20-39%, the method is used in less than half the organization's programs, (d) 40-59%, the method is used in approximately half the organization's programs. (e) 60-79%, the method is used in more than half the organization's programs, and (f) 80-100%, the

method is used in most of the organization's programs. The preceding ranges were chosen to give the respondents more choices than some, about half, and most without providing so many categories that the respondents would have to calculate actual use.

For each Level, there were many missing values in the three areas of the survey that asked for information on the use of various evaluation methods. There were 806 missing values, approximately 21% of the total possible responses. The reporting for the rest of the survey contained few missing values. To find out why this occurred, the researcher attempted to telephone each of the 30 respondents with multiple missing values. A total of 10 respondents was contacted.

Each respondent was asked if the missing responses should be treated as missing values or if a missing response meant that the method was not used. Eight respondents said that missing responses should be recorded as 0% use of the method. Two of the respondents in this phone survey said that they did not have time to discuss the survey and could not remember how they answered. The response rate for the accessible population was 80 %. Based on the response from the phone survey, the researcher used missing values in the three sections on evaluation methods as 0% use.

Frequencies for the reported use of each method are presented in tables. Use for each method covered the entire range from one to six representing categories from "never used" to "used in most of the organization's programs." The frequencies are used to show the reported use for all responding organizations. This provides an indication of which methods were most widely used by the respondents. Additionally,

since the data were ordinal in nature, medians were calculated. Upon inspection it was found that the median values did not accurately reflect use. Therefore, they are not presented.

The amount the evaluation methods were used was calculated for those organizations using methods in 1 % or more of their programs. The frequency data for respondents who reported that their organizations used a method is displayed in chart form in the following three sections. This data reflects how much each method is used for those organizations that employ the method as an evaluation tool. Respondents were removed only from those methods in which they reported zero use. Therefore, the n for each method varied.

Evaluation methods used at Level 2

To quantify the use of each evaluation method at Level 2, the survey instrument asked the respondents to estimate the percentage of programs in which their organization used each of the various methods provided in the item. The methods for Level 2 evaluation are listed in Table 6. The respondents were provided with six categories related to percentage ranges. The percentage ranges were (a) 0%, (b) 1-19%, (c) 20-39%, (d) 40-59%, (e) 60-79%, and (f) 80-100%. This was done to simplify the respondents' estimation of the percentage of programs using each method. The respondents were also provided with an opportunity to describe and quantify the use of any methods that were not listed on the survey.

The frequencies for the reported use of Level 2 evaluation methods are presented in Table 6 and percentages are presented in Table 7. Table 6 lists each Level 2 evaluation method in descending order based on the number of respondents reporting that their organization never uses a given method. This was done by placing the smallest frequency for "never" at the top of the table and the method with the greatest frequency for "never" at the bottom.

Table 6

Frequencies for the "Percentage of Programs Using Each Level 2 Evaluation Method" for All Respondents

Method	Scale value and percent of programs					
	1	2	3	4	5	6
	0%	1-19%	20-39%	40-59%	60-79%	80-100%
Skill Demonstrations	34	12	18	19	14	14
Post only	44	21	17	7	11	11
Pre-Post	50	32	6	8	5	10
Simulations	53	19	9	10	11	9
OJT Demonstrations	55	10	15	11	11	9
Work samples	62	15	9	10	11	4

Note. n = 111 for all methods

More respondents reported that each method was never used, Category 1, than reported using a method in any of the other categories, two through six, as shown by the frequencies in column two of Table 6. (The term, "category or categories," is used as an alternative to "percentage range or 0 % and the percentage ranges two through six.") Skill demonstrations were reported as the most used method.

Table 7

Percentages for the "Percentage of Programs Using Each Level 2 Evaluation Method" for All Respondents

Method	Scale value and percent of programs					
	1	2	3	4	5	6
	0%	1-19%	20-39%	40-59%	60-79%	80-100%
Skill Demonstrations	31	11	16	17	13	13
Post only	40	19	15	6	10	10
Pre/Post	45	29	5	7	5	9
Simulations	48	17	8	9	10	8
OJT Demonstrations	50	9	14	10	10	8
Work samples	56	14	8	9	10	4

Note. $\Sigma \% = 100$ for all methods

Table 7 lists each Level 2 evaluation method in descending order based on the percentage of respondents reporting that their organization never uses a given method. Eleven to 17 % of the respondents reported using skill demonstrations in each category, two through six. This was the method that had the least respondents in category one (the method was never used). The only method that had a reported use by more than 20 % of the respondents was pretest-posttest in category two (the method was used in from 1-19% of their programs).

Of the 14 responses to the question asking what other methods of Level 2 evaluation are used (see Appendix D, Table D1), only the four on the list below could not be placed in the categories provided:

- Actual outcomes-success rate with new approach or method (Level 3)
- Self-evaluation: i.e. skill level rating

- ▶ Competency based training with self-evaluations to follow up
- ▶ Peer review

Six of the responses are forms of posttests and three are forms of skills tests or demonstrations. One response was a Level 1 evaluation technique.

Figure 3 reports the frequencies for only the respondents that make use of each method. The methods are listed in descending order of use based on the number of respondents using the methods in category six. The methods reported as the most used by those organizations using them are at the top of figure 3 and the least used at the bottom.

Respondents were removed from only those methods in which they reported zero use. This made it possible for the number of respondents in each category to differ. The range was 77 to 49. Skill demonstration was the only method used by more than ten respondents in every category. Skill demonstrations, posttest only and pretest-posttest were reported by 10 or more respondents as used in 80 to 100 % of their programs. Pretest-posttest, posttest only, simulations, and work samples had the largest number of organizations reporting that they were used in less than 20 % of their programs. Because the data in Table 6 and Figure 3 were remarkably similar, the correlation between number of organizations reporting using a method in 80 % or more of their programs (Category 6) and the number of organizations reporting using each method (n for each method) was taken. The correlation was .95. This very high association, based on Davis' (1971) descriptors, may suggest a relationship between

the two measures. Methods reported as most widely used by all organizations (those with greatest frequencies and percentages in Tables 5 and 6 as well as the largest ns in Figure 3) have the greatest frequencies for methods used in 80 to 100 % of programs.

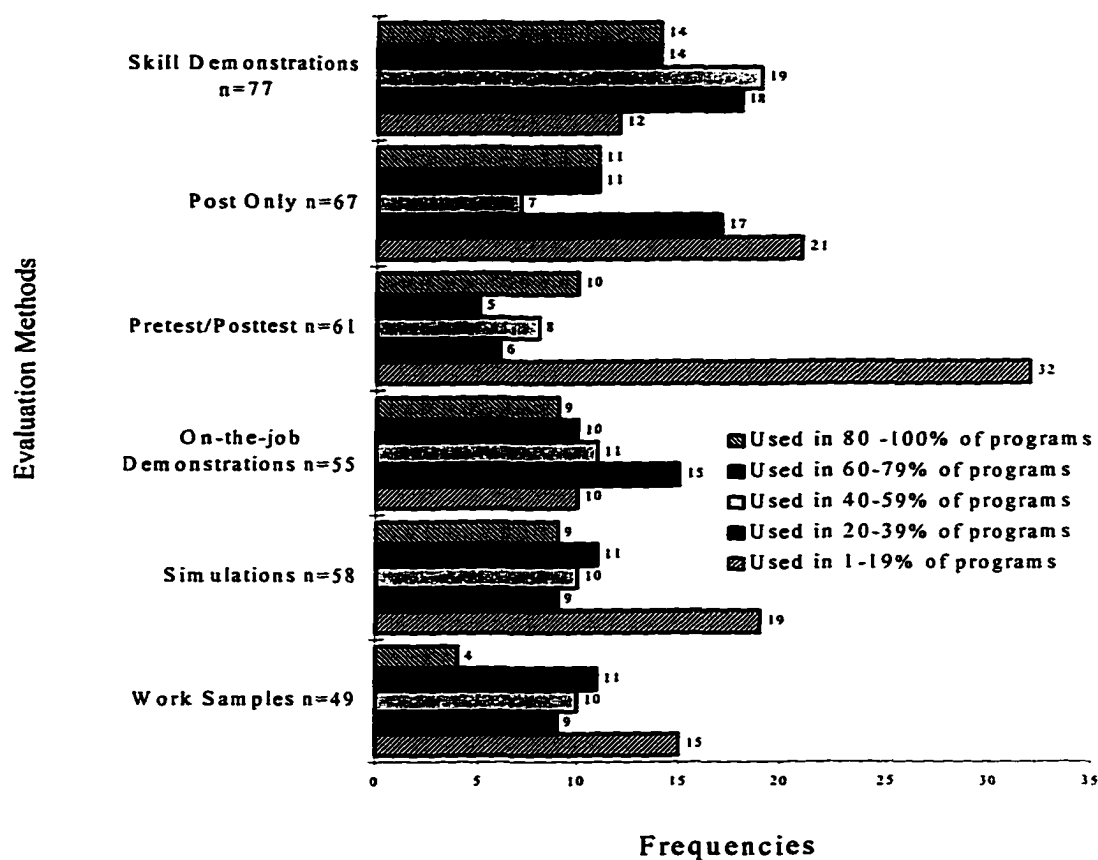


Figure 3. Frequencies for Level 2 evaluation methods in all use categories other than zero

Evaluation methods used at Level 3

The question used to elicit responses on the Level 3 evaluation methods used by each organization was; "Please estimate the percentage of programs in which your organization uses each of the various methods listed on the right to evaluate **the use of learning on the job.**" The methods listed are presented in Table 8. The respondents were also provided with an opportunity to list any methods they used which were not on the survey.

The frequencies for all Level 3 evaluation methods in each use category are shown in Table 8. Table 8 lists each Level 3 method in descending order based on the number of respondents reporting that their organization never uses a given method. The method with the lowest number of responses for "never" is at the top of the table and the method with the highest number of responses for "never" at the bottom. The method reported most used by all organizations is at the top of the table, and the method reported least used at the bottom. Approximately half the respondents (60, $n = 108$ and 52, $n = 107$) reported using observation and performance appraisals respectively.

More than 86 % of the respondents reported never using audit action plans, peer surveys, focus groups, performance contracts, peer interviews, or subordinate interviews. The other five methods, (a) supervisor interviews, (b)existing records, (c) follow-up assignments, (d) records produced for evaluation, and (e) subordinate surveys, were reported as never used by 64 to 82 % of the respondents. In other

Table 8

Frequencies for the "Percentage of Programs Using Each Level 3 Evaluation Method" for All Respondents

Method	Scale value and percent of programs						n
	1	2	3	4	5	6	
	0%	1-19%	20-39%	40-59%	60-79%	80-100%	
	f	f	f	f	f	f	
Observation	48	19	10	9	11	11	108
Performance appraisals	55	6	4	13	9	9	107
Supervisor surveys	69	14	10	5	8	2	108
Anecdotal	74	16	9	5	2	3	109
Supervisor interviews	77	15	5	7	2	2	108
Action plans	78	11	6	10	2	1	108
Self report surveys	78	16	3	8	2	1	108
Trainee Interviews	79	15	6	5	3	0	108
Existing records	80	7	6	7	7	1	108
Follow-up assignments	83	9	5	6	4	1	108
Records produced for evaluation	87	5	4	6	3	3	108
Subordinate surveys	89	10	6	0	3	0	108
Audit Action plans	92	8	3	3	1	0	107
Peer Surveys	92	10	2	2	1	1	108
Focus groups	94	8	2	2	2	0	108
Performance contracts	94	5	2	6	1	0	108
Peer interviews	97	8	1	0	1	1	108
Subordinate interviews	98	8	0	0	2	0	108

words, the preceding five methods were only used by 18 to 36 % of the respondents even when the all the respondents in use categories 2 through 6 are used as a measure. (see Table 9)

Table 9

Percentages for the "Percentage of Programs Using Each Level 3 Evaluation Method" for All Respondents

Method	Scale value and percent of programs					
	1	2	3	4	5	6
	0%	1-19%	20-39%	40-59%	60-79%	80-100%
	%	%	%	%	%	%
Observation	44	18	9	8	10	10
Performance appraisals	51	6	4	12	8	19
Supervisor surveys	64	13	9	5	7	2
Anecdotal	68	15	8	5	2	3
Supervisor interviews	71	14	5	6	2	2
Action plans	72	10	6	9	2	1
Self report surveys	72	15	3	7	2	1
Trainee Interviews	73	14	6	5	3	0
Existing records	74	6	6	6	6	1
Follow-up assignments	77	8	5	6	4	1
Records produced for evaluation	81	5	4	6	3	3
Subordinate surveys	82	9	6	0	3	0
Audit Action plans	86	7	3	3	1	0
Peer Surveys	85	9	2	2	1	1
Focus groups	87	7	2	2	2	0
Performance contracts	87	5	2	6	1	0
Peer interviews	90	7	1	0	1	1
Subordinate interviews	91	7	0	0	2	0

Note. $\sum \% = 100$ for all methods

Of the 14 responses to the question asking what other methods of Level 3 evaluation are used, all can be categorized as: peer review (1), existing records (6) or records produced for evaluation (7). Of the 14 responses, 6 responses involved measures of customer satisfaction. (see Appendix D, Table D2)

At Level 3, the frequencies for the reported "percentage of programs using each evaluation method" followed the same pattern as the frequencies for the respondents reporting that their organizations used a method in 80 to 100 percent of their programs. However, the patterns were not as closely matched as they were at Level 2. The first six items at the top of figure 4 include the first five items reported as the most used for all organizations tables 8 and 9. The correlation between reported percentage of programs using each method for all organizations and the methods used in 80 to 100 percent of the programs for those reporting making use of each method was .74 as compared to .95 for Level 2. This high association, based on Davis' (1971) descriptors, may indicate a relationship between the two measures.

Some methods were reported as used only by only a few organizations but were commonly used within those organizations. For example "existing records" of were only used by 26 percent of the respondents but over half of those that made use existing records, 15 out of 28, used them in 40 percent or more of their programs. However, the correlations show that for most evaluation methods those used by the most organizations are also most frequently used within the individual organizations that make use of them.

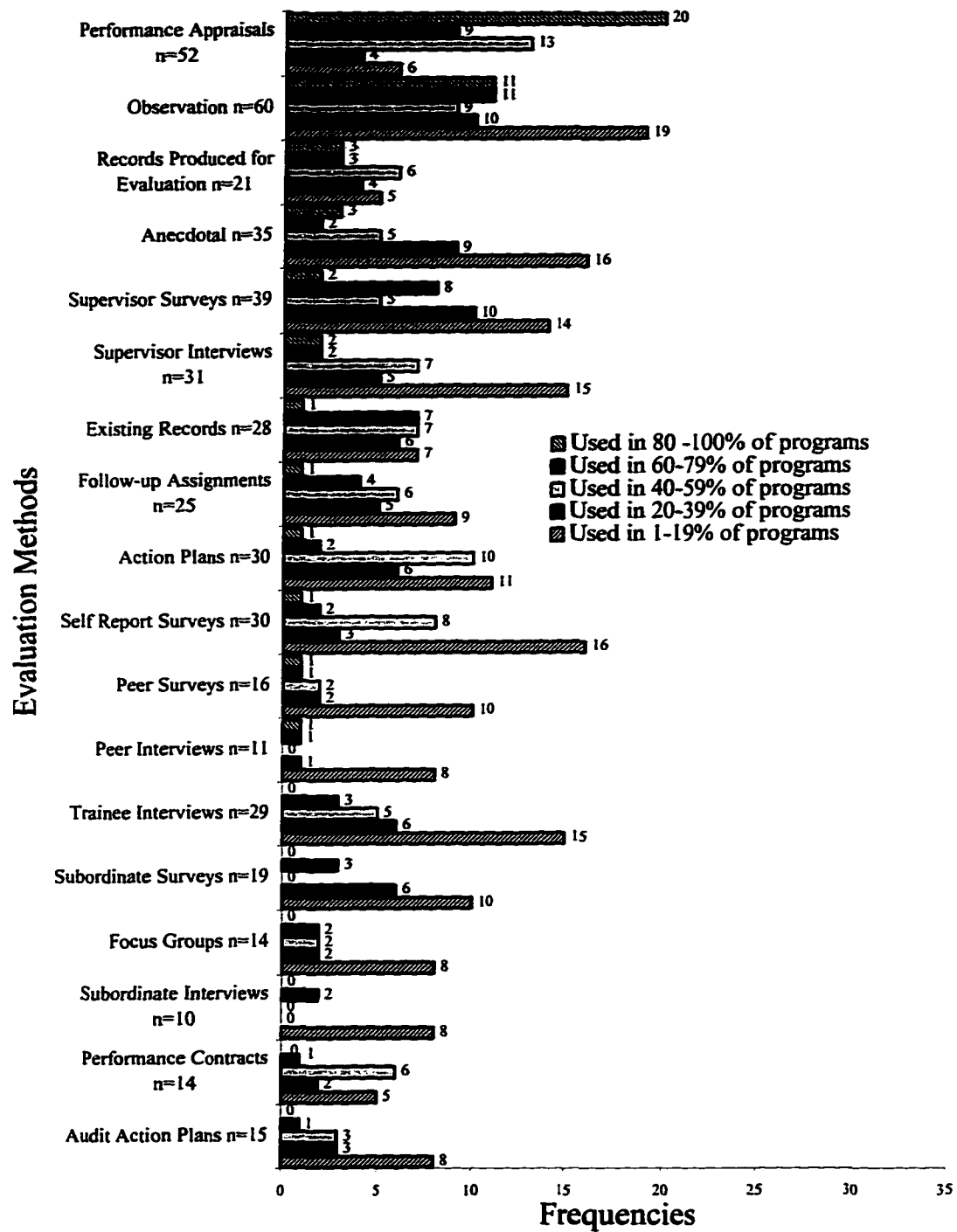


Figure 4. Frequencies for Level 3 evaluation methods in all use categories other than zero

Evaluation methods used at Level 4

The following question was used to gather information on the evaluation methods used at Level 4: "Please estimate the percentage of programs in which your organization uses each of the various methods listed on the right to evaluate **organizational results**." The methods listed are in the first column of Table 10.

The frequencies for all Level 4 evaluation methods in all use categories are shown in Table 10 and show the reported range of use for each method for all the respondents' organizations. Table 10 lists each Level 4 evaluation method in descending order based on the number of respondents reporting that their

Table 10

Frequencies for the "Percentage of Programs Using Each Level 4 Evaluation Method" for All Respondents

Method	Scale value and percent of programs					
	1	2	3	4	5	6
	0%	1-19%	20-39%	40-59%	60-79%	80-100%
Productivity Estimates	64	15	10	9	5	5
Productivity Measures	69	14	5	7	8	5
Regulation Compliance	71	15	4	4	7	7
Before and after Measures Related to Training Goals	73	12	7	4	7	5
Cost Savings	79	13	5	1	7	3
Anecdotal Information	81	9	6	6	2	4
ROI	87	7	4	6	1	3

Note. n = 108 for all methods

organization never uses a given method. The method with the lowest frequency for "never" is at the top and the method with the highest frequency for "never" at the bottom. The method reported as used by the most organizations is at the top of the table and the method reported as used by the least number of organizations at the bottom. Many respondents reported never using any of the methods listed.

Productivity estimates and productivity measures were reported as used by more of the respondents than any of the other methods. ROI (return on investment) and cost saving were among the three methods with the lowest reported use. These two methods are highly recommended in both training journals and evaluation textbooks as Level 4 evaluation methods that are the most effective in showing the value of training.

Table 11 shows the percentage of the respondents reporting the use of each method. Seventy-three percent never use cost savings and 81 % never use ROI. The highest percentages of use were for productivity estimates, productivity measures, and regulation compliance. These methods were reported as used by 13 or 14 % of the respondents in 1 to 19 % of their programs.

The respondents were provided with an opportunity to list other evaluation methods used in Level 4 evaluation. The following is a list of those responses: (a) observed changes in behavior, (b) we survey our dealers and they report 25% to 40% increases after training, (c) safety records, and (d) sales results. Item (a) in the above list is the only item that does not fit in the predefined categories and is a Level 3

method. The other three items are either changes in productivity (b, c, and d) or regulation compliance (c).(see Appendix D, Table D3 for use categories.)

Table 11

Percentages for the "Percentage of Programs Using Each Level 4 Evaluation Method" for All Respondents

Method	Scale value and percent of programs					
	1	2	3	4	5	6
	0%	1-19%	20-39%	40-59%	60-79%	80-100%
Productivity Estimates	59	14	9	8	5	5
Productivity Measures	64	13	5	6	7	5
Regulation Compliance	66	14	4	4	6	6
Before and after Measures Related to Training Goals	68	11	6	4	6	5
Cost Savings	73	12	5	1	6	3
Anecdotal Information	75	8	6	6	2	4
ROI	81	6	4	6	1	3

Note. $\Sigma \% = 100$ for all methods

Figure 5 shows the frequencies for use for categories two through six in descending order of use from top to bottom. These frequencies are for the amount of use in those organizations that use each Level 4 method. Category 1 is not shown. The correlation between reported use for all organizations, r for each category on Figure 5, and the frequency for Category 6 was .70. This high association, based on Davis' (1971) descriptors, may indicate a relationship between the two measures.

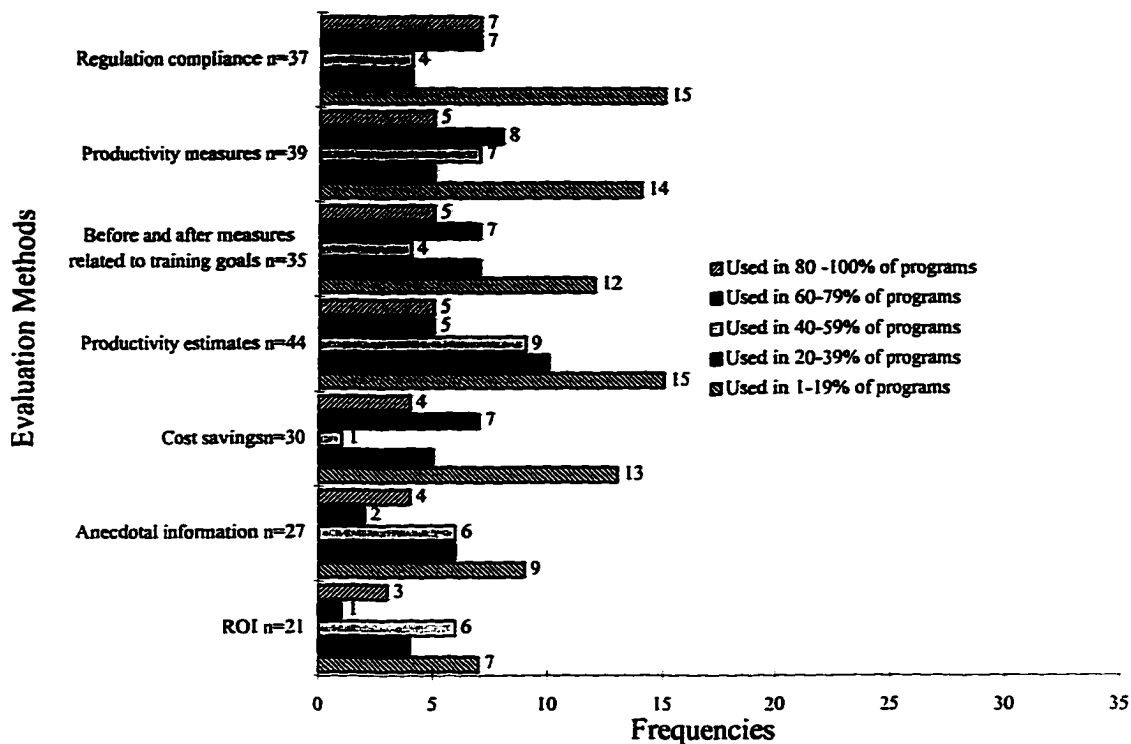


Figure 5. Frequencies for level 4 evaluation methods in all use categories other than zero

Various Reasons for Not Evaluating

Objective three was to determine the reasons for not evaluating at each Level.

For those times when evaluation was not done at any of the four Levels, the respondents were asked to indicate the reasons. Four questions, one for each Level, were used to gather data on the reasons for not evaluating:

1. For those times when you **do not** gather information on **participant reactions**, what are the reasons?
2. For those times when you **do not** evaluate **learning** that took place during a program, what are the reasons?

3. For those times when you **do not** evaluate the **use of learning on-the-job**, what are the reasons?
4. For those times when you **do not** evaluate **organizational results**, what are the reasons?

The following options were provided:

- ▶ of little value to the organization
- ▶ the cost in person hours and capital
- ▶ evaluation takes too much time from the course
- ▶ lack of training or experience in using this form of evaluation
- ▶ union opposition
- ▶ not required by the organization
- ▶ policy prohibits the evaluation of employees by the training department
- ▶ training is done only to meet legal requirements

The most commonly reported reason for not evaluating across all Levels of evaluation was that evaluation was not required by the organization. After not required, the three most frequently reported reasons for not evaluating were the cost in person hours and capital, lack of training or experience in using this form of evaluation, and not required by the organization. The two reasons least cited were policy prohibits the evaluation of employees by the training department and union opposition. Level 2 had the highest percentage of respondents reporting that each reason applied to their organization. The percentage of the respondents reporting that each reason applied to

their organization is shown in Table 12. Each of the four Levels are reported separately.

Table 12
Reasons for Not Evaluating at Each Level 1 through 4

Level 1		
Reason for not evaluating	Organizations Reporting	n
Not required	28.83%	106
Little Value	18.92%	106
Cost	10.81%	106
Not legally required	9.91%	106
Lack of training	9.01%	106
Time	8.11%	105
Union	1.80%	106
Prohibited	.90%	106
Level 2		
Reason for not evaluating	Organizations Reporting	n
Not required	36.94%	106
Lack of training	23.42%	106
Time	21.62%	106
Little Value	19.82%	104
Cost	18.02%	106
Not legally required	14.41%	106
Union	4.50%	106
Prohibited	3.60%	107

table continued

Level 3		
Reason for not evaluating	Organizations Reporting	n
Not required	44.14%	110
Cost	36.94%	110
Lack of training	34.23%	110
Little Value	13.51%	110
Not legally required	7.21%	110
Time	3.60%	110
Union	2.70%	110
Prohibited	1.80%	110
Level 4		
Reason for not evaluating	Organizations Reporting	n
Not required	42.34%	108
Lack of training	39.64%	108
Cost	36.94%	108
Little Value	15.32%	107
Not legally required	8.11%	108
Time	5.41%	108
Prohibited	1.80%	108
Union	1.80%	108

Figure 6 shows the frequencies for respondents reporting a reason for not evaluating applied to their organization. The reasons are grouped by Level and a fifth

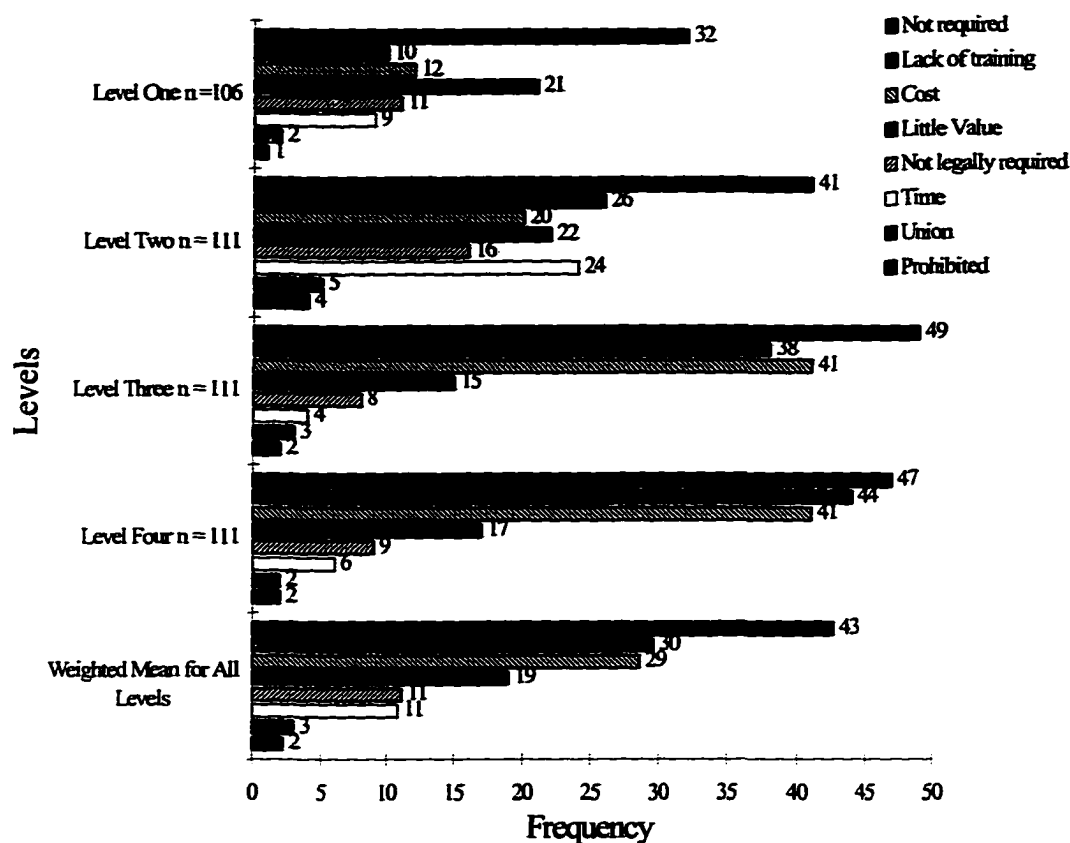


Figure 6. Reasons for Not Evaluating at Each Level

group is used to show the weighted means for all four levels. There were more responses for not required by the organization than for any other reason. The four reasons with the highest mean frequencies were (a) not required by the organization, (b) lack of training in evaluation techniques, (c) cost, and (d) lack of value from evaluation results. The respondents were provided with an opportunity to list reasons for not evaluating in addition to those provided on the survey. For Level 1, there were a total of nine responses providing other reasons for not evaluating. Five of the responses fell into the not required, lack of training, lack of time, or lack of value

categories. There were two reasons for not evaluating at Level 1 that do not fit the categories which were provided on the survey form (a) the instructor may fail to implement and (b) participants' refusal to participate in evaluation. (see Appendix E)

Of the 18 responses to the other methods of evaluation used for Level 2 evaluation: (a) two suggest that evaluation does occur, (b) two describe instructors' memory lapse, (c) one replicates the time category, (d) six deal with management opposition and organizational culture matching the prohibited category either directly or indirectly, (e) six are based on measurement problems (lack of evaluation training), and (f) the last "not designed into the course" may fall into any of several categories such as "lack of training" or "not required." (see Appendix E)

The respondents were provided with an opportunity to list reasons for not evaluating at Level 3 that were not provided on the survey. There was a total of 20 responses providing other reasons for not evaluating. Seven of the responses relate to a lack of control of what occurs after training. Six of the responses suggest a lack of resources (time or budget) placing them in the cost or time categories. A lack of ability to measure outcomes at this Level was reported as the reason for not evaluating at this Level by three of the respondents. The information provided by the respondents makes it impossible to decide how many responses could also be placed in the "lack of training" category. One response shows some difficulty in differentiating Level 2 and Level 3 evaluation for on-the-job training. Another response, "We try to eliminate

fear," shows internal resistance to evaluation and how the information from evaluation may be used. Finally, one response described a form of evaluation. (see Appendix E)

There were 15 responses at Level 4 describing barriers to evaluation. More than one third of the responses described measurement problems and when combined with those responses discussing a lack of evaluation methods in place, nearly half the responses (7) are related to a lack of training in evaluation methods. Four of the remaining responses suggest that a lack of resources is another barrier to evaluation as was organizational culture (two responses). The other two responses question the value of evaluation in specific situations and fall into any of several categories such as "little value," "lack of training," or "not required." (see Appendix E)

Exploratory Correlations Between Business and Industry Variables and Evaluation Variables

Objective four was to determine if a relationship existed between the function of a business or industry and the Level of evaluation used or method of evaluation used. Additionally, objective four was to determine if a relationship existed between the size of a business or industry (as determined by the number of employees) or the number of employees trained each year and the Level of evaluation used.

Exploratory correlations between organizational function and the Level of evaluation used

Correlations were drawn between the reported organizational functions and the amount each Level is used. Use was determined by the percentage of programs using a Level of evaluation. Eighty-eight correlations were drawn. The correlation coefficient

was less than .30 for all cases. The only correlation that was significant at the .05 level was between aviation and Level 1 and that correlation was .20.

Exploratory correlations between organizational function and the method of evaluation used

To determine if a relationship existed between the function of a business or industry and how much an evaluation method is used, correlations were drawn between the functions and the use of each method of evaluation. Use was determined by the percentage of programs in which the method was used. The correlation coefficient was .30 or greater for two cases; banking and audit action plans with an r of .31, $p < .01$ and textiles and supervisor interviews with an r of .32, $p < .01$. There were seven organizations in the banking function and one in textiles. Based on Davis' (1971) scale for interpreting correlations this is a moderate association. However, the small number of organizations in both cases and the large number of correlations drawn, 638, should be noted.

Exploratory correlations between an organization's size and the Level of evaluation used

To determine if a relationship exists between the size of a business or industry (as determined by the number of employees) or the number of employees trained each year and the Levels of evaluation used, correlations were drawn between "number of employees" and "number of employees trained" and the percentage of use of each Level of evaluation. The correlation coefficients were less than .30 for all cases and for Levels 2 through 4 they were very small, ranging from .004 to .09. However, there

was a .19, $p < .05$ correlation between the number of employees and the percentage of programs using Level 1 evaluation.

Exploratory Correlations Between Training Managers' Perceptions of the Importance of a Level and the Level of Evaluation Used

Objective five was to determine if a relationship existed between the training manager's perception of the importance of a Level of evaluation in improving training, gaining upper management's support for training, or reaching organizational goals, and how much each Level was used. Correlations between the variables perception and percentage of programs using a Level were calculated. The correlation coefficients were less than .30 for all cases. However, all the correlations were positive. The correlations between a manager's perception of the value of Levels 1 through 4 in improving training and the percentage of programs using a Level were in order .29, .27, .25, .19 (with $p < .01$, $p < .01$, $p < .01$, and $p < .05$ respectively). Based on Davis' (1971) descriptors this is a low positive association between the manager's perception of the value of Levels 1 through 4 in improving training and the amount of evaluation at Levels 1 through 4. Three other correlations were significant at the .05 level: (a) .19 between demonstrating the value of on-the-job training (OJT) and Level 1, (b) .19 between demonstrating value in reaching organizational goals and Level 2, and (c) .20 between demonstrating value in reaching organizational goals and Level 4. There was a total of 16 correlations drawn for this objective.

Exploratory Correlations Between Training Managers' Experience and the Level of Evaluation Used

Objective six was to determine if a relationship exists between a training manager's experience and the percentage of programs using each Level of evaluation. The measure of a training manager's experience was the number of years the respondent reported as having performed the training function. The correlation between these two interval variables (experience and percentage of programs using a Level of evaluation) was calculated. For Levels 1 through 4, the correlations were -.15, .08, .22 ($p < .05$), .12 respectively.

Exploratory Correlations Between Age of the Training Program and the Level of Evaluation Used

Objective seven was to determine if a relationship existed between the age of a training program in years and the percentage of programs using each Level of evaluation. The correlation between these two interval variables (experience and percentage of use) was calculated. All correlations were less than .30. Correlations between program age and the use of each Level of evaluation, 1 through 4, were in order .01, .11, -.04, and .15. None of the correlations reached significance at the $p < .05$.

Exploratory Correlations Between Various Organizational Training Practices and the use of each Level or Method of Evaluation

Objective eight was to determine if a relationship exists between organizational training practices and how much each Level or method of evaluation is used. Correlations were calculated where relationships could logically be expected.

The organizational training practices were (a) types of training programs, (b) when training evaluation is planned, (c) to whom evaluation results are reported, (d) why training is done, (e) the percent of training personal involved in evaluation, (f) the percent of training staff trained in evaluation, (g) evaluation techniques used that matched research techniques, (h) program funding, and (i) the overall funding of the training function. Correlations were calculated between organizational training practices and (a) the percent of programs using Levels 1 through 4 or (b) the percent of programs using various methods of evaluation. Correlations were drawn only where relationships could logically be expected.

Training methods and the percentage of programs using each evaluation Level 1 through 4

Data were gathered on the training methods used by the respondents with the following item: "Technical training can be accomplished using many methods and/or combinations of methods. For your organization, please indicate the percent of programs that fall into each category." The categories were (a) informal OJT (following another employee's lead), (b) formal OJT (planned and monitored), (c) apprenticeships, (d) self-study, (e) individual training events addressing specific needs, (f) a curriculum based on organizational goals, and (g) work team initiated training.

Correlations were drawn between the percent of programs using each training method and the percent of programs using Levels 1 through 4. The strength of the associations is discussed using Davis' (1971) scale for interpreting correlation

coefficients. Only informal OJT showed low negative or negligible correlations with all Levels of evaluation. Apprenticeship showed the only other negative correlation with a low association at Level 1. The highest positive correlations were the following three moderate associations between: (a) formal OJT (planned and monitored) and Level 2, $r = .35$, $p < .01$; curriculum based on organizational goals and Level 2, $r = .37$, $p < .01$; individual training events addressing specific needs and Level 1, $r = .33$, $p < .01$. The only other correlations approaching .30 were between individual training events addressing specific needs and Level 2, $r = .29$, $p < .01$ and between team initiated training and Level 4, $r = .29$, $p < .01$. The data from these correlations is reported in Table 13.

Table 13

Correlations Between Training Methods and the Percentage of Programs Using Each Level of Evaluation

Training Methods	Levels			
	1	2	3	4
Informal OJT	-.23*	-.05	-.15	.02
Formal OJT	.06	.35**	.19*	.09
Apprenticeships	-.12	.24*	.03	.05
Self-study	.06	.22*	.07	.20*
Individual training events	.33**	.29**	.17	.11
Goal based curriculum	.26	.37**	.06	.17
Team initiated training	.15	.20*	.10	.29**

Note. * $p < .05$. ** $p < .01$.

Integration of evaluation in programs and the percentage of programs using each evaluation Level 1 through 4

An item on the survey asked respondents to "please indicate the percentage of programs in which your organization starts planning the evaluation process at each of the stages listed at the right. The list contained the following items: (a) prior to program development, (b) as the first step in program development, (c) during program development, (d) after program completion, (e) when training results must be documented, and (f) evaluations are not implemented. The item was developed to gather data on whether evaluation was part of the program development process or was a separate function external to program development. The responses to this item were collapsed into two measures. The measure of programs in which evaluation was an integral part of program development was based on the sum of the frequencies for the following items: (a) the evaluation process is started prior to program development, (b) the evaluation process is the first step in program development, or (c) the evaluation process is started during program development. The measure of programs in which evaluation was not an integral part of program development was based on the sum of the frequencies for the following items: (a) the evaluation process is started after program completion, (b) the evaluation process is started when training results must be documented, and (c) evaluations are not implemented.

Correlations were drawn between (a) programs in which evaluation was or was not an integral part and (b) the percentage of programs using each of the four Levels of evaluation. The strength of the association is discussed using Davis' (1971) scale

for interpreting correlation coefficients. All the correlations were positive. The correlation between Level 4 evaluation and programs in which evaluation was not an integral part was negligible. Correlations between programs in which evaluation was an integral part were all significant at the .01 or .05 level. Correlations between programs in which evaluation was an integral part and Levels 1 and 2 was moderate, based on Davis' (1971) descriptors. All other correlations had low associations. The results of the correlations are reported in Table 14.

Table 14

Correlations Between Integration of Evaluation in Programs and the percentage of programs using each Level of evaluation

Amount of use	Integral Part of Programs	Independent of Programs
Level 1	.38**	.18
Level 2	.45**	.17
Level 3	.22*	.16
Level 4	.22*	.07

Note. * $p < .05$. ** $p < .01$.

Perceived importance of evaluation in proving value to upper management and the percentage of programs reporting evaluation results to management

Four survey items asked respondents how much value they thought each Level of evaluation had in demonstrating the value of training to upper management. A second item asked the respondents to indicate the percentage of programs in which the results for each were reported to the following:

- stockholders
- CEO and upper level managers
- direct managers of those receiving training

- ▶ program participants
- ▶ training department members not directly involved in the evaluation process
- ▶ all employees

The frequencies for the responses to the percentage of programs in which the results for each were reported to (a) stockholders, (b) CEO and upper level managers, and (c) direct managers of those receiving training were summed as a measure of the frequency of reporting to management overall. Correlations were drawn between the respondents' perception of the value of each Level in demonstrating value to upper management and the frequency of reporting to management for all four Levels. The results of these correlations in order for Levels 1 through 4 are .34, .16, .18, and .17.

None of the correlations reached significance at the .05 level.

Perceived importance of evaluation in improving programs and the percentage of programs reporting outcomes to participants

A survey item asked the respondents to indicate the percentage of programs in which the results were reported to program participants. Correlations were taken between the respondent's perceived importance for Levels 1 through 4 in improving programs and the responses to the percentage of programs in which the evaluation results for each were reported to program participants. The correlation for the reported relationship of the manager's perception of the value of the four Levels and how often the results of evaluations are reported to participants were .19, .12, .04, and -.02. None of the correlation reached significance at the .05 level.

Programs delivered to all the members of a target audience and the Level of evaluation used

Data was gathered on the reasons for training using the following item.

"Employee development programs are delivered for a variety of reasons and have different levels of participation. Please indicate the percent of your programs that match the descriptions on the right." The response categories were (a) employees are sent to this program as a reward, (b) all employees at this level have always had this program, (c) all employees involved in an activity or specific group attend this program, (d) participants will acquire new attitudes by attending this program, (e) participants in this program will be able to perform at a set level, and (f) a change in organizational outcomes will result from this course.

The frequencies for "all employees at this level have always had this program" and "all employees involved in an activity or specific group attend this program" were summed as a measure of "all the members of a target audience" Correlations were drawn between "all the members of a target audience" and the percent of programs using each of the four Levels of evaluation. The correlations in order for Levels 1- 4 were .22, .23, .02, and -.02. None of the correlations reached the .05 significance level.

Programs that are delivered to change performance or organizational outcomes and the Level of evaluation used

Data was gathered on the reasons for training using the following item.

"Employee development programs are delivered for a variety of reasons and have

different participation rates. Please indicate the percent of your programs that match the descriptions on the right." Two of the response categories were (a) participants in this program will be able to perform at a set level and (b) a change in organizational outcomes will result from this course. The items (a) participants in this program will be able to perform at a set level and (b) a change in organizational outcomes will result from this course were combined by summing the frequencies for each respondent. This summation was used as a measure of programs presented for the reason of producing a change in performance in the participants or the organization. Correlations were drawn between this measure and the use of the four Levels of evaluation. Using Davis' (1971) descriptors, there was a low positive association with all four Levels and programs with specific expected outcomes. From Level 1 to Level 4 the correlations were in order .22, $p < .05$; .25, $p < .01$; .20, $p < .05$; and .15, $p > .05$.

Percent of staff involved in evaluation and the Level of evaluation used

Respondents were asked what percentage of the employee development staff was involved in evaluation. Correlations were taken between the percentage of the employee development staff involved in evaluation and the percentage of programs using each of the four Levels. The only correlation more than .30 showing a moderate association based on Davis' (1971) descriptors is for Level 1 and its value is .31, $p < .01$. The correlations for Levels 2 through 4 are .13, .06, and .08 respectively, showing

a low association with Level 2 and a negligible association for Levels 3 and 4; none reached significance at the .05 level.

Percent of staff with formal training in evaluation and the Level of evaluation used

Respondents were asked what percentage of the employee development staff had formal training in evaluation. Correlations were taken between percentage of the employee development staff reported as having formal training in evaluation and the percentage of programs using Levels 1 through 4. There was a low positive association, based on Davis' (1971) descriptors, for the use of all four Levels and formal training in evaluation. The correlations between percentage of the employee development staff reported as having formal training in evaluation and the percentage of programs using Levels 1 through 3 were .16, .12, .11 respectively; none reached significance at the .05 level. The correlation for Level 4 was .29, $p < .01$.

Scientifically accepted research techniques and reasons for delivering training

Data was gathered on the use of accepted research techniques using the following item: "What percentage of your training programs use the evaluation techniques listed on the right." The list included (a) a matching (randomly chosen) control group, (b) a control group selected from a similar work unit, and (c) multiple measures taken before and after a program.

Data was gathered on the reasons for training using the following item: "Employee development programs are delivered for a variety of reasons and have different levels of participation. Please indicate the percent of your programs that

match the descriptions on the right." The categories were (a) employees are sent to this program as a reward, (b) all employees at this level have always had this program, (c) all employees involved in an activity or specific group attend this program, (d) participants will acquire new attitudes by attending this program, (e) participants in this program will be able to perform at a set level, and (f) a change in organizational outcomes will result from this course.

Correlations were taken between the reported percentage of programs using each research method and the reported percentage of programs implemented for the various reasons for training. There was a moderate association between "repeated measures" and "participants will acquire new attitudes by attending this program." All the other correlations were negligible to low using Davis' (1971) descriptors for associations. Random selection had the highest association across all the reasons for training. Training with specific expected outcomes (participants will acquire new attitudes by attending this program, participants in this program will be able to perform at a set level, and change in organizational outcomes will result from this course) had the highest correlation with repeated measures. The data for all the correlations is reported in Table 15.

Percentage of programs dependent on evaluation for funding and the Levels of evaluation used

Data was gathered to determine the percentage of training programs in each organization that were dependent on evaluation for funding. The respondents were

Table 15

Correlations Between Evaluation Techniques and the Reasons for Training

Reason for Program Delivery	Evaluation Technique		
	Random selection	Matching	Repeated measures
Reward	.18	.13	-.05
Everyone takes this program	.27**	.05	-.14
Specific groups receive this training	.10	.02	.09
Attitude change	.11	.12	.31**
Attain performance required	.12	.22*	.19*
Change organizational outcomes	.15	.18*	.27**

Note. * $p < .05$. ** $p < .01$.

asked approximately what percentage of their programs were dependent on evaluation for continued funding. The mean for this item was 13.47 % with a range of 0 to 100 % and a standard error and deviation of 2.90 and 30.32 respectively. Correlations were taken between the percent of programs dependent on evaluation for funding and the percentage of programs using each of the four Levels. Using Davis' (1971) descriptors for correlations, Level 3 had a low correlation .21, $p < .05$ with the percent of programs dependent on evaluation for funding and the other three Levels had negligible correlation coefficients.

Percentage of budget dependent on evaluation for funding and the Levels of evaluation used

Data was gathered to determine the percentage of training budget that was dependent on evaluation for funding. The respondents were asked approximately what percentage of their budget was dependent on evaluation for continued funding. The mean for this item was 8.68 % with a range of 0 to 100 % and a standard error and

deviation of 2.01 and 21.00 respectively. Correlations were taken between the percentage of training budget that was dependent on evaluation for funding and the percentage of programs using each of the four Levels. Using Davis' (1971) descriptors for correlations, Level 1 had a low correlation, .24, $p < 5$ with the percentage of training budget that was dependent on evaluation for funding, and the other three Levels had negligible correlation coefficients.

Funding methods for training and the Level of evaluation used

The respondents were asked how training was funded within their organization and provided with three options: separate training budget, training as a separate profit center, or some of both. Correlations were taken between funding methods and the use of each Level of evaluation. There was a low association with Level 1 and separate profit centers. All the other associations were negligible. The descriptors used above were based on Davis (1971). The correlations are reported in Table 16.

Table 16

Correlation Between Funding and the Use of Each Evaluation Level

	Separate training budget	Separate profit center	Some of both
Level 1	-.04	.12	-.03
Level 2	-.07	-.01	.08
Level 3	-.08	-.01	.09
Level 4	-.01	-.06	.05

Table 16
Correlation Between Funding and the Use of Each Evaluation Level

	Separate training budget	Separate profit center	Some of both
Level 1	-.04	.12	-.03
Level 2	-.07	-.01	.08
Level 3	-.08	-.01	.09
Level 4	-.01	-.06	.05

CHAPTER 5

SUMMARY

Introduction

American industry must continuously upgrade its workforce's skills through training and education to stay globally competitive. U. S. business and industry spend a tremendous amount of money providing technical training to its employees.

America, its industries, and its workers all have a tremendous stake in training. The costs are both in dollars and in time. Production workers and administrative employees received an estimated 750 million hours of training in 1996 at a cost of more than 20 billion dollars (Industry Report, 1996). Is this training effective? Does it produce a change in the employees who receive it? Does it produce changes in organizations that increase their ability to function at a reasonable cost?

Evaluation is one way to assure that the benefits expected from training are delivered and the cost of providing training is justified. Evaluation can assure that the benefits possible from training are realized. Evaluation can also provide information allowing business and industry to decide if the dollars spent on training are producing the maximum benefit. Evaluation, both formative and summative, is a method to assure that both training and education are efficient (completed in the shortest time and for the least cost possible) and effective (produce the desired employee and organizational change).

The most common model for evaluation in business and industry classifies evaluation into types:

- Participant reaction
- Attainment of learning objectives
- Actual changes in on-the-job performance
- The effect of training on the organization (Medsker & Roberts, 1992. p. 1)

This model was introduced by Donald Kirkpatrick. "In the November 1959 article, I used the term 'four steps.' But someone, I don't know who, referred to the steps as 'levels.' The next thing I knew, articles and books were referring to the four Levels as the Kirkpatrick model" (Kirkpatrick, 1996).

Most of the available literature reports that training is not adequately evaluated. Most of the articles in the literature report that Level 1 evaluation is common across business and industry. They also report that the use of each Level becomes less common from 1 to 4 (Dixon, 1990, p. 1, Gordon, 1991, p.21, Carnevale & Schulz, 1990, p. s-24, Robinson and Robinson, 1989, p. 170-171). Not all the literature agrees with the surveys discussed above. The Corporate HRD Executive Survey of the American Society of Training and Development in their Survey #11 Report (1989), reports only 57 % of the companies surveyed used participant reaction forms in technical training.

Objectives

The primary purpose of this study was to determine the percentage of programs using each of the four Levels of evaluation in technical training.

Additionally, the study attempted to determine what evaluation methods were used at each Level, what were the barriers to evaluating, and if there were any relationships between a company's training environment and evaluation practices.

Eight objectives were developed.

- 1. Determine the extent to which business and industry are using evaluation Levels 1, 2, 3, or 4 to evaluate technical training.**
- 2. In instances where they are used, determine the methods of Level 2, 3, or 4 evaluation used.**
- 3. Where a Level of evaluation was not used, determine some of the reasons for not using that Level of evaluation.**
- 4. Determine if a relationship existed between selected industry demographics and the amount a Level or method of evaluation was used. The demographics to be studied were a business' or industry's function, the number of people employed, and the number of individuals trained per year.**
- 5. Determine if a relationship existed between the training manager's perceptions of the importance of a Level of evaluation to selected organizational functions and the amount a Level of evaluation was used. The functions to be studied were improving**

training, gaining upper management's support for training, and reaching organizational goals.

6. Determine if a relationship existed between a training manager's experience in training and the percentage of programs using of each Level and method of evaluation.

7. Determine if a relationship existed between the length of time a technical training program had been in existence and the percentage of use of each Level and method of evaluation.

8. To determine if a relationship existed between organizational training practices and how much each Level or method of evaluation was used. The organizational training practices were (a) types of training programs, (b) when training evaluation was planned, (c) to whom evaluation results are reported, (d) why training was done, (e) the percent of training personnel involved in evaluation, (f) the percent of training staff trained in evaluation, (g) evaluation techniques used that match research techniques, (h) program funding, and (i) the overall funding of the training function. Correlations were calculated between organizational training practices and (a) the percent of programs using Levels 1 through 4 or (b) the percent of programs using various methods of evaluation.

Methods

A total of 348 survey forms were sent to organizations with members in the ASTD professional practice area of Technical and Skills Trainers. Additional mailings

followed Dillman's TDM (Total Design Method) dealing with survey mailings (1978). Population errors identified from returned surveys were addressed by mailings to additional members of the original sample which randomized the entire accessible population to provide for replacements. A total of 146 surveys were returned for an overall response rate of 42%. One-hundred-twelve usable surveys were returned making the response rate for usable surveys 35%.

The survey consisted of seven sections. Sections 1 through 4 included questions on the percent of programs evaluated at each Level, types of evaluations used, and reasons for not evaluating at each Level. Section 5 consisted of questions used to determine the training practices of the organization. The questions asked what methods of program delivery were used, when evaluations were planned, to whom evaluation results were reported, the purpose of training, number of personnel involved in evaluation and their training in evaluation practices, research practices used, and the relation of evaluation to budgeting. Section 6 gathered information on the respondents' thoughts on the value of evaluation to improve training, demonstrating the value of training in improving job performance, demonstrating the value of training to upper management, and demonstrating the value of training in attaining organizational goals. This section was based on earlier research that showed a relationship between training manager's thoughts on training evaluation and how much evaluation was done. Section 7 gathered demographic data.

A sample of twenty-two questions was drawn from the original questionnaire and used as a telephone interview guide for the non-respondents. Since it would be expected that at least some differences would be found as an artifact of a large number of comparisons, the respondents and non-respondents were considered different only if they differed on more than 10% of the questions or, in this case, if more than two significant differences were found. The groups differed significantly on only two questions. Additionally, the results of this survey on the use of each Level of evaluation roughly parallels that of previous studies (Dixon, 1990, p. 1, Gordon, 1991, p.21, Carnevale & Schulz, 1990, p. s-24, Robinson and Robinson, 1989, p. 170-171). The researcher concluded that the respondents and non-respondents were not significantly different.

Results

The extent to which business and industry are using evaluation Levels 1, 2, 3, or 4 to evaluate technical training

This research found that technical training used each of Kirkpatrick's four Levels of evaluation in the following percentage of their programs: Level 1 - 72.74%, Level 2 - 47.05%, Level 3 - 33.73%, and Level 4 - 20.82%.

Evaluation methods used at Levels 2, 3, and 4

Use of the various evaluation types or methods was addressed in two different ways. The first measure was the percentage of programs using the method. The second was the amount of use by those organizations that reported using the method. It was therefore possible for a method to be used by a small percentage of the reporting

organizations but to have a high percentage of use within those organizations using the method.

The Level 2 evaluation method reported as most used was skill demonstrations. This method was also used in the highest percent of programs by the organizations using the method. The second most common evaluation method at Level 2 was the posttest with no pretest. On-the-job demonstrations, simulations, and work samples were the methods used in the fewest number of programs and organizations. Common use of a method across organizations did not always indicate that a method was used regularly to evaluate programs. Posttests were reported as used by well over half the organizations but were reported most commonly used to evaluate only one fifth or less of the programs taught by those organizations.

Observation was reported as the most commonly used Level 3 evaluation method. It was used by slightly over half the reporting organizations. Performance appraisals were the second most common evaluation method and were reported as used by slightly less than half the respondents. However, for the organizations using performance appraisals, they were reported as used in more programs than any other Level 3 method. Other methods commonly reported as used were existing records, records produced for evaluation, action plans, and performance contracts. Surveys and interviews were not commonly used.

Productivity estimates, productivity measures, and regulation compliance were reported as used by one third or more of the organizations for Level 4 evaluation.

However, these three methods were only used in a few of the programs taught. ROI measures were used by less than one fifth of the respondents, but those using ROI used it in approximately half the programs taught. All other Level 4 evaluation methods were used in less than one fifth of the programs taught.

Reasons for not evaluating

For those instances when evaluation was not done at any of the four Levels, the respondents were asked to indicate the reasons. The most commonly reported reason for not evaluating was that evaluation was not required. The three reasons reported by the highest percent of organizations for all Levels of evaluation were not required, lack of training, and cost. The two reasons least cited were prohibited by the organization or prevented by the union.

Exploratory correlations between various organizational variables and the 4 Levels of evaluation

Exploratory correlations were run between various organizational variables and the four Levels of evaluation. Additionally, exploratory correlations were run between various organizational variables and methods of evaluation. It is recognized that with many correlations, relationships may be a construct of the number of correlations and not replicable in future studies. Because of this, only those correlations greater than .30 are discussed. The existence of strong relationships between the variables would have suggested areas for further study. However, few correlations were found and, of those, only one exceeded .40.

Exploratory correlations between business and industry function and Level or method of evaluation used. No significant correlations were found between business and industry function and Level of evaluation used. When comparing the function of a business and methods of evaluation, two cases were found with small correlations. Banking and audit action plans plus textiles and supervisor interviews were correlated. However, these correlations were based on a small number of respondents and have little practical significance. Across technical training there is almost no statistical relationship between an organization's function and the Level or method of evaluation used.

Exploratory correlations between business and industry size and Level of evaluation used. No correlations more than .30 were found between business and industry size and Level of evaluation used. Nor were any correlations more than .30 found between the number of employees trained and the Level of evaluation used. However, there was a .19, $p < .05$ correlation between the number of employees and the percentage of programs using Level 1 evaluation.

Exploratory correlations between training managers' perceptions of the importance of a Level and its use. An earlier study showed that within the banking industry there was a relationship between a training manager's perceptions of the importance of a Level and its use (Gutek, 1988). This research did not find any correlations greater than .30 between perceptions and use. However, smaller significant correlations between a manager's perception of the value of Levels 1

through 4 in improving training and the percentage of programs using a Level were found. The correlation between the perceived value of evaluation in improving programs and the four Levels were in order .29, .27, .25, .19 (with $p < .01$, .01, .01, and .05 respectively). Based on Davis' (1971) descriptors, this is a low positive association between perceived value of evaluation in improving programs and the use of each Level of evaluation. Two correlations of .20, $p < .05$ were found between demonstrating value in reaching organizational goals and Levels 2 and 4. There was a total of 16 correlations drawn for this objective.

Exploratory correlations between training managers' experience and the use of each Level of evaluation. No correlations of .30 or more were found between a training manager's experience and the use of each Level of evaluation. However, one significant correlation was found between experience and percentage of programs using a Level of evaluation. That correlation was between experience and percentage of programs using Level 3, .22, $p < .05$.

Exploratory correlations between age of the training program and the use of each Level of evaluation. No significant correlation was found between the length of time an organization had been training and the use of each Level of evaluation. There was a small positive correlation between Level 1 evaluation usage and percentage of training accomplished by individual training events that were not part of an overall curriculum. There was also a small positive correlation between the percentage of training that was part of a curriculum and Level 2 evaluation usage.

Exploratory correlations between various organizational training practices and the use of each Level or method of evaluation

Correlations were calculated between organizational training practices and how much each Level or method of evaluation is used. The organizational training practices were (a) types of training programs, (b) when training evaluation is planned, (c) to whom evaluation results are reported, (d) why training is done, (e) the percent of training personnel involved in evaluation, (f) the percent of training staff trained in evaluation, (g) evaluation techniques used that match research techniques, (h) program funding, and (i) the overall funding of the training function. Correlations were calculated between organizational training practices and (a) the percent of programs using Levels 1 through 4 or (b) the percent of programs using various methods of evaluation.

Training methods and the percentage of programs using each evaluation Level

1 through 4. There was a low negative correlation significant at $p < .05$ and a moderate positive correlation significant at $p < .01$ between Level 1 and (a) informal OJT and (b) individual training events respectively. Level 2 had low or moderate correlations significant at $p < .01$ or $p < .05$ with all training methods except informal OJT. The only significant correlation at Level 3 was a low positive correlation with formal OJT. The only significant correlations at Level 4 were low positive correlations with self-study and team initiated training.

Correlation between integration of evaluation in programs and the use of evaluation at each Level. The strongest correlations in the study were in the

relationship between Levels 1 and 2 and the integration of evaluation into training program design. Those programs in which evaluation was an integral part of the design process showed moderate correlations (Davis, 1971) with the use of the first two Levels of evaluation and positive correlations at all four Levels. The correlations for Levels 1 through 4 were .38, $p < .01$; .45, $p < .01$; .22, $p < .05$; and .22, $p < .05$ respectively.

Correlation between perceived importance in demonstrating value to management and frequency of reporting to management. None of the correlations between a training manager's perception of the importance of a Level of evaluation in demonstrating value to management and the frequency of reporting evaluation outcomes to managers reached significance at $p < .05$. However, the correlation between a training manager's perception of the importance of a Level of evaluation in demonstrating value to management and the frequency of reporting evaluation outcomes to managers was .34 for level 1.

Correlation between perceived importance in improving programs and frequency of reporting to participants. No significant correlation ($p < .05$) was found between the perceived importance in improving programs and frequency of reporting to participants.

Programs that are delivered to change performance or organizational outcomes and the Level of evaluation used. Correlations were drawn between programs delivered to change performance or organizational outcomes and the use of the four

Levels of evaluation. None of these correlations were .30 or higher. However, using Davis' (1971) descriptors, there was a low positive association with all four Levels and programs with specific expected outcomes. From Level 1 to Level 4 the correlations were in order .22, $p < .05$; .25, $p < .01$; .20, $p < .05$; and .15, $p > .05$.

Correlation between percent of staff involved in evaluation and the Level of evaluation used. The only correlation more than .30 showing a moderate association based on Davis' (1971) descriptors was for Level 1 and its value was .31, $p < .01$. The larger the percentage of training staff involved in evaluation, the more likely Level 1 evaluation will take place.

Percent of staff with formal training in evaluation and the Level of evaluation used. Correlations were taken between percentage of the employee development staff reported as having formal training in evaluation and the percentage of programs using of Levels 1 through 4. None of these correlations were .30 or higher. However, the correlation for Level 4 was .29, $p < .01$.

Correlation between scientifically accepted evaluation techniques and reason for delivering training. There were low to moderate associations significant at the $p < .05$ between "programs with expected outcomes" and both (a) matching and (b) repeated measures. A low positive correlation significant at $p < .01$ was also found between random selection and everyone takes this program.

Percentage of programs dependent on evaluation for funding. Few of the programs operated by organizations in this survey were dependent on evaluation for

funding. There were low associations significant at the $p < .05$ between (a) separate profit centers and (b) Level 1 and (a) separate training budget and (b) Level 2.

Training managers' perceived values for each evaluation Level and the Level of evaluation used. Four survey items asked respondents how much value they thought each Level of evaluation had in four areas: improving training, demonstrating the value of training in improving on the job performance, demonstrating the value of training to upper management, and demonstrating the value of training in attaining organizational goals. There were low positive correlations significant at $p < .01$ between improving training and Levels 1, 2, and 3. There were also low positive correlations significant at $p < .05$ between demonstrating value in organizational outcomes and Levels 2 and 4.

Conclusions

The extent to which business and industry are using evaluation Levels 1, 2, 3, or 4 to evaluate technical training

Previous research reports the use of evaluation as a percent of the respondents' programs that use each Level or the percent of respondents' organizations that use each Level. Since the various studies used different methods to determine use, a direct comparison is not possible. However, the reported use of evaluation in technical training closely matches the values in previous reports. Additionally, the values in this study tend to more closely match the upper ends of the ranges in other studies. Finally, as reported in the past, the amount of evaluation done at each Level shows consistent declines from Level 1 to Level 4.

If the measure of the amount evaluation is used is "the percent of programs using each Level of evaluation", this study replicates previous findings. Level 1 is used to evaluate more than half of technical training programs. Level 2 evaluation is used in less than half of technical training programs. Evaluation Levels 3 and 4 are only used to evaluate a small percentage of technical training programs.

The percentage of organizations using each Level is a measure of the use reported in the literature (Dixon, 1990; Industry Report, 1996; Phillips, 1991). Selecting different minimum percentages of programs using a Level of evaluation as criteria for concluding that respondents use a Level of evaluation affects the reported results. This could lead to misinterpretation and makes comparisons across studies very difficult.

This research yields percentages slightly above those reported in Training's 1996 Industry Report when the same criterion is used. Training's 1996 Industry Report included all organizations that used a Level of evaluation in any percentage of programs greater than zero. That report was for all types of training and listed the following percentage of organizations as using each of the four Levels; (a) Level 1 - 86 percent, (b) Level 2 - 71 percent, (c) Level 3 - 65 percent, and (d) Level 4 - 49 percent. This research found the slightly higher values that follow: (a) Level 1 -92 percent, (b) Level 2 -84 percent, (c) Level 3 - 65 percent, and (d) Level 4 -53 percent.

Using 60 % and 80 % as minimum percentages of programs using a Level of evaluation as criteria for concluding that respondents use that Level, this research

produced numbers that closely match results reported in the rest of the literature (Carnevale & Schulz, 1990; Dixon, 1987; Phillips, 1991; Robinson and Robinson, 1989; Survey #11, 1989). Robinson and Robinson's (1989) survey indicated that 77 % of training directors use Level 1 evaluation in 81 to 100 % of their courses. Survey #11, the Corporate HRD Executive Survey of the American Society of Training and Development, reported that in technical training only 57 % used participant reaction forms (1989). The reported use for Level 1 in this study was 71 % when the criterion "used in more than half the organization's programs" was used and 63 % when "used in most of the organization's programs" was used as the criterion.

Robinson and Robinson (1989) report only 22 % of the training directors surveyed use Level 2 evaluation in 80 % or more of their courses. The American Society for Training and Development's Survey Report #11 (1989) shows this type of evaluation at 45 %. The reported use in this research falls between 42% (used in more than half the organization's training programs) and 29% (used in most of the organization's training programs). Again this study's findings closely match those of other studies.

Two of the four surveys in the literature reported the use of Level 3 evaluation between 10 and 15 % (Carnevale & Schulz, 1990). Six to 13 % of training managers use Level 3 evaluations for 80 % of the programs taught. The reported use for Level 3 in this study was between 15 and 21 %.

Carnevale and Schulz (1990) report that "Employee training was only evaluated at the organizational result level about 25 % of the time, despite new pressures on training practitioners to assess the economic worth of HRD activities." Robinson and Robinson (1989, p. 171) report that for 80 % of the courses taught, less than 6 % of training managers use Level 4 evaluation and 59 % of managers never used Level 4 evaluation. Again the reported use in this study (11% to 17%) was close to that of other studies (6% to 25%).

When following the literature back to the development of the four levels, there has been little change in the amount of evaluation within business and industry for at least the last 25 years. Gutek (1988) cited a 1968 study that asked if organizations used each of the four levels. That study reported that 78 % used Level 1, 50 % used Level 2, 54 % used Level 3, and 45 % used Level 4. The usage in the older study is similar to that of this study for Levels 1 and 2 as discussed above but not for Levels 3 and 4. However, the older study stated that the evaluation at Levels 3 and 4 were superficial and subjective. Using a less strict interpretation of use to match the older study ("used in some of the organization's programs" or "used in more than half of the organization's programs") this study had a reported use of 21% to 66% for Level 3 and 17% to 55% for Level 4. The range of findings in this study include the values from this 25-year-old study. As early as 1953, Wallace and Twichell were discussing the need for and lack of training evaluation. Today's literature contains parallel comments and the lack of training evaluation still exists.

Evaluation methods used

The literature discusses a wide range of evaluation methods. Thirty-one were placed in the survey and the respondents were provided with an opportunity to write in additional methods. Only six methods were used by more than half the reporting organizations. Skill demonstrations, posttests, pretests and posttests, and simulations were used by more than half the respondents as Level 2 evaluation methods. The most commonly reported Level 3 evaluation methods were observation and performance appraisals. Observation and performance appraisals were used by approximately half the organizations in this research. No Level 4 evaluation method was used by more than half the responding organizations. However, productivity estimates, productivity measures, and regulation compliance were used by more than a third of the reporting organizations as Level 4 measures. The literature does not discuss the amount each method is used nor does it infer that some methods are generally more effective than other. However this study found that of the many evaluation methods available, only a few are commonly used by technical trainers.

Reasons for not evaluating

The three most common obstacles to evaluation are cost, lack of training, and not required. Organizations seldom require training departments to evaluate (a mean of 13 % of the time). Even when funding is dependent on evaluation, there were no moderate or substantial correlations between funding and evaluation. Was the requirement more stringently imposed, training departments would not have the

knowledge and skills required to perform the evaluations. Additionally the cost of evaluations is seen to outweigh the benefits. It does not appear that either organizational culture or labor organizations play a major role in preventing the evaluation of training outcomes.

Exploratory correlations between business and industry variables and evaluation variables

This study found no business and industry variables that had any practically significant relationship ($r > .30$) to the levels or methods of evaluation used. The only statistically significant relationship which might be of interest was a low positive association between the number of employees and the use of Level 1 evaluation.

Exploratory correlations between training managers' perceptions of the importance of a Level and the Level of evaluation used

There was no practically significant relationship between training managers' perceptions of the value of each level of evaluation and their use in technical training. However, nearly half the correlations, seven out of 16, were positive and statistically significant at $p < .05$ or $p < .01$ and all the correlations between a manager's perception of the value of a Level and improving training were positive and statistically significant. Gutek (1988) found that banking training managers' perceptions of the value of a Level of evaluation were the controlling factor in the use of that Level. The correlations in this research were not large enough to extend Gutek's (1988) findings to the area of technical training. However, they lend support to those findings since all the correlations were positive.

Exploratory correlations between training managers' experience and the Level of evaluation used

The number of years a technical training manager has performed the training function showed no practical relationship to the Levels of evaluation used. However there was a low, positive, statistically significant correlation between Level 3 and a technical training manager's experience. More experienced technical training managers are more likely to use Level 3 evaluations.

Exploratory correlations between various organizational training practices and the use of each Level or method of evaluation

These correlations were taken to determine if there were any relationships between organizational training practices, a measure of organizational attitude toward evaluation (Phillips, 1991), and the amount evaluation is used. This was done because it has been suggested the attitude toward evaluation affects how much evaluation is done.

Training methods and the percentage of programs using each evaluation Level 1 through 4. Programs with preplanned outcomes, both formal OJT and curriculum based on organizational goals were associated with direct program outcomes (Level 2), not changes in behavior or organizational goals. Individual training events were associated most strongly with Level 1 evaluation. Only team initiated training was associated with measuring changes directly related to training and changes in organizational goals. No substantial associations were found. The term substantial is based on Davis' descriptors (1971). Davis' scale (1971) is less restrictive than the

Hinkle, Wiersma, and Jurs scale (1988). The Hinkle, Wiersma, and Jurs scale (1988) described all the correlations in this research except one as having little if any correlation and the one exception would have been described as a low positive correlation.

Integration of evaluation in programs and the percentage of programs using each evaluation Level 1 through 4. The strongest correlations between organizational variables and evaluation were moderate positive associations between Levels 1 and 2 and the integration of evaluation into training program design. This correlation may occur because measurements of how much evaluation is designed into programs and how much evaluation is done are measuring the same thing and would have to show a strong positive correlation. Although if this were the case, a much stronger correlation should have been found. There was only a low positive association between integration and Levels 3 and 4. The data from this research supports both curriculum design and evaluation literature that states evaluation is more likely to take place if it is designed into the program than when evaluation is a separate process. However, the association between integration and the use of any Level of evaluation is not substantial based on Davis' descriptors (1971).

Perceived importance of evaluation and the percentage of programs reporting outcomes. No statistically or practically significant relationships were found between a training manager's perceptions of the importance of a Level in (a) improving training, (b) demonstrating the value of training in improving on the job performance,

(c) demonstrating the value of training to upper level management, or (d) demonstrating the value of training in attaining organizational goals. A technical training manager's perceived importance of a Level shows only a negligible relationship with reporting the results of evaluation to management or program participants.

Training staff evaluation variables and the Level of evaluation used. There was a moderate positive association between the percent of staff involved in evaluation and the use of Level 1 evaluation. Additionally, there was a low association between percent of staff with formal training in evaluation and the use of Level 4 evaluation. Both associations were significant at $p. < .01$. The number of staff involved in technical training evaluation and the amount of formal training the staff has in evaluating training have little association with the percent of programs using each Level of evaluation.

Scientifically accepted research techniques and reasons for delivering training. There was a moderate association between "repeated measures" and "participants will acquire new attitudes by attending this program." All the other correlations were negligible to low using Davis' (1971) descriptors for associations. No substantial associations were found.

Funding variables and the percentage of programs using each Level of evaluation. A small percentage of programs are dependent on evaluation, only a few training departments' budgets are dependent on evaluation, and more than half the

respondents reported that they were not functioning as a separate profit center. Only low to moderate correlations were found between any funding variables and percentage of programs dependent on evaluation for funding, the percentage of training budget that was dependent on evaluation for funding, and the Levels of evaluation used. No substantial associations were found between funding and the percentage of programs using any of the four Levels of evaluation.

Recommendations for Practice

If an organization wants to increase the amount of formal evaluation applied to its training programs, the following recommendations show promise:

1. require evaluation reports at the specific evaluation Levels of interest
2. provide training in evaluation to those involved in evaluation
3. ensure that the evaluation process will cost less than the value of the information gained
4. provide specific time in training programs for evaluation
5. make evaluation an early step in program design

Suggestions for Further Research

The consistent lack of evaluation across business and industry suggests two possibilities: either (a) billions of training dollars are spent with no measure of their effect on the participant's knowledge and skills, the organization's functioning, and/or the organization's profitability or (b) business and industry are using other methods to measure the effect of training. Therefore, it is suggested that further research focus on

two different areas: the value of evaluating using Kirkpatrick's Model and identifying what other methods are being used to evaluate training.

Research questions based on Kirkpatrick's Model

1. What is the present cost in time and capital to evaluate at each Level?
2. What are the most cost effective methods of evaluating at each of Kirkpatrick's four Levels?
3. What is the value of the information gathered by each of Kirkpatrick's four Levels?
4. What methods are best for training personnel to evaluate training outcomes?

Research questions based on identifying what other methods are being used to evaluate training

1. What criteria are used by supervisors to determine whether training should be initiated?
2. What criteria are used by supervisors to determine whether training was effective?
3. What criteria are used by supervisors to determine whether training changed performance?
4. What criteria are used by supervisors to determine whether training changed the way the group under their supervision functions and if the changes are positive or negative?

5. What criteria are used by supervisors to determine whether training changed the profitability for the area under their supervision and if the changes are positive or negative?
6. What criteria are used by supervisors to determine whether to continue using a form of training?
7. What criteria are used by middle and upper management to determine whether training should be initiated?
8. What criteria are used by middle and upper management to determine whether training was effective?
9. What criteria are used by middle and upper management to determine whether training changed performance?
10. What criteria are used by middle and upper management to determine whether training changed the way the group under their supervision functions and if the changes are positive or negative?
11. What criteria are used by middle and upper management to determine whether training changed the profitability for the area under their supervision and if the changes are positive or negative?
12. What criteria are used by middle and upper management to determine whether to continue using a form of training?

Implications

The lack of evaluation

These thoughts are based on the researcher's experience in completing this research and its relationship to the researcher's experiences while training in industry. Television news, news magazines, economic reports, management journals, training literature, politicians, and business figures all discuss the need to increase the rate of growth in productivity, downsizing, cost-cutting, and ever increasing competitiveness. The training literature presents evaluation as a necessary component in providing training that can help organizations increase productivity, reduce the required number of employees, cut costs, and increase competitiveness. All the literature on how much evaluation is used by business and industry suggests that less than half of the training programs in place are evaluated for objective outcomes. Additionally, less than one third of training programs are evaluated in any way that measure changes in organizational goals or profitability. Either business and industry are not concerned with the costs or outcomes of training or they are using some method of justifying costs and checking outcomes that does not include Donald Kirkpatrick's four Levels (Kirkpatrick, 1996).

After at least 40 years of bemoaning the lack of evaluation, supporting the value of evaluation, developing methods of evaluation, and pushing the evaluation cause, there appears to have been little change in the amount, types, or quality of evaluation in business and industry. As in many other endeavors, history keeps

repeating itself. Program developers and trainers see the obvious value of evaluation because it would affirm the value of what they do and provide an objective basis for improving both programs and teaching. However in the face of business and industry's resistance to **formal** evaluation based on Donald Kirkpatrick's four Levels (1996), those in the training field must ask, is it truly possible that business and industry spend billions of dollars without verifying the value of what they purchase? This is highly unlikely. It is time to ask, "how are business and industry placing value on training?" instead of bemoaning the fact that they do not evaluate training by the accepted methods reported in the literature. Additionally, why aren't evaluations a recognized management tool?

Evaluating for an audience

Public education programs were evaluated extensively in the 60s and 70s to discover the value of large federally funded programs. Evaluation was a requisite part of each program. When the results of those evaluations were reported, Congress became impatient with evaluations that could not state simply whether the funded programs were producing the desired results. In response, the evaluation community developed new methods and practices. Finally, members of the evaluation community created the Joint Committee on Standards for Educational Evaluation. The Joint Committee developed Standards for Evaluations of Educational Programs, Projects, and Materials (Standards, 1981).

The Standards for Evaluations of Educational Programs, Projects, and Materials has very little in common with Donald Kirkpatrick's four Levels (1996). This is not a negative comment on Donald Kirkpatrick's work. Donald Kirkpatrick never meant his four articles on evaluation as the format for all forms of evaluation in business and industry. Indeed, he was surprised by their acceptance as a system of evaluation. They are and always have been a description of the various outcomes or possible effects of training that should be of interest to those purchasing and consuming training (Kirkpatrick 1996).

The Standards for Evaluations of Educational Programs, Projects, and Materials are a set of guidelines for developing an evaluation that will produce the required information for the audience (customer, manager, the consumer of research, etc.) at a cost that is reasonable compared with the value of the results. No where in the Standards (1981) is the information required for a given type of education identified. The Standards for Evaluations of Educational Programs, Projects, and Materials (1981) focuses on the customer (audience), value, quality, reliability, and timeliness. This focus is well known in business and industry. Customer oriented organizations are touted as the basis for success and profitability in the management literature.

"People in all fields make choices, and it is inconceivable that they should do so without assessing the worth or merit of options"(Standards, 1981). "It is inconceivable" (Standards, 1981) that consumers of technical training should have to

make choices about that training without "assessing the worth or merit of "(Standards, 1981) its programs. Either the consumers are assessing the value of training programs using methods outside the scope of research based on Donald Kirkpatrick's four Levels (1996) or they are doing the inconceivable. If training organizations want evaluation to be carried out regularly and to be appreciated, it might be of value to ask the customers, internal or external, what they want or need to know and what form of information would be useful before starting any evaluation process.

If the training department will be using evaluation results, then using Kirkpatrick's Model will probably provide the information required by the customer. However if the training department is the supplier, how can it hope to provide the customer with the required information without first asking the customer what they want and need from the evaluation process? The customer, not the service provider, decides what will be purchased and sets the criteria for judging the product. It is ironic that those who are continually trying to train people to focus on the customer appear to be ignoring their own training.

BIBLIOGRAPHY

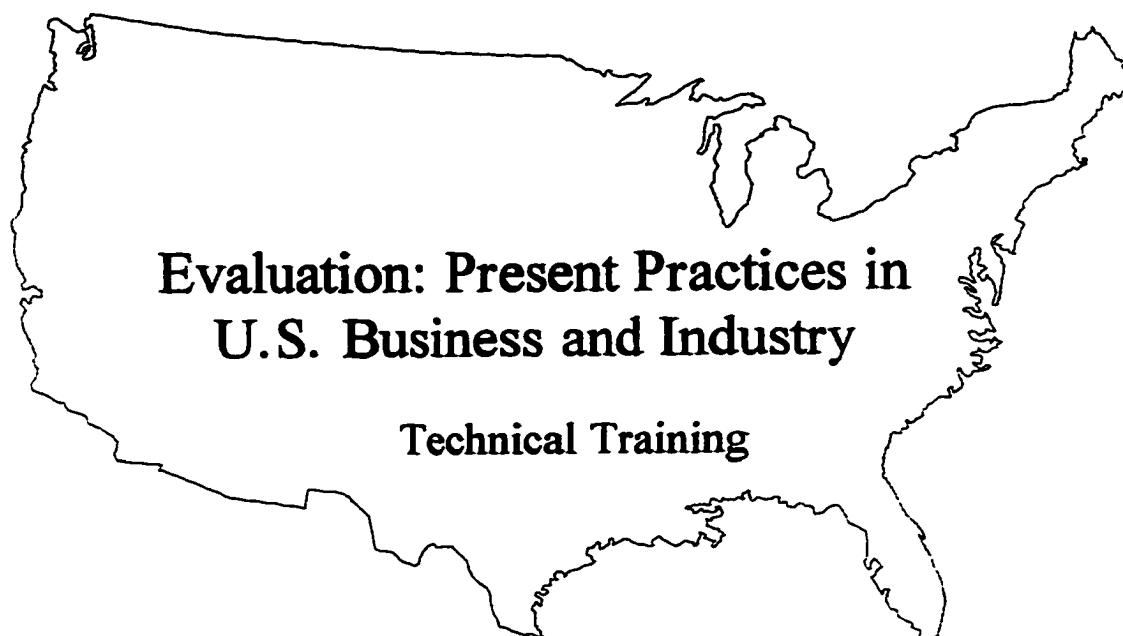
- American Society for Training and Development. (1991). The evaluation of training. Alexandria, VA: Author.
- Bernardin, H. J. & Betty, R. W. (1984). Performance appraisal: Assessing human behavior at work. Boston: Kent.
- Brinkerhoff, R. O. (1991). Achieving results from training. San Francisco: Jossey-Bass.
- Broad, M. L. & Newstrom, J. W. (1992). Transfer of training: action-packed strategies to ensure high payoff from training investments. Reading, MA.: Addison-Wesly.
- Bushnell, D. S. (1990). Input, process, output: A model for evaluating training. Training and Development Journal, 44(3) 41-43.
- Callahan, M. R. (Ed.). (1986). Essentials for Evaluation. Alexandria, VA: American Society for Training and Development.
- Carnevale, A. P. (1991) America and the new economy: How competitive standards are radically changing American workplaces. San Francisco: Jossey-Bass.
- Carnevale, A. P., Gainer, L. J., & Schulz, E. (1990). Training the technical work force. San Francisco: Jossey-Bass.
- Carnevale, A. P., & Schulz E. (1990). Economic accountability for training: Demands and responses. Training and Development Journal, 44(7), s-1 - s-32.
- Cascio, W. F. (1982). Costing human resources: The financial impact of behavior in organizations. Boston: Kent.
- Cocheu, T. (1990). Training from the start. Training and Development Journal, 44(10) 22-27.
- Connolly, S. M. (1991) Integrating evaluation, design, and implementation. In American Society for Training and Development (Ed.). The Evaluation of Training. Alexandria, VA: editor.

- Corporate HRD Executive Survey. (1989). Survey #11 Report. Alexandria, VA: American Society for Training and Development.
- Darraugh, B. (Ed.). (1991). Measuring attitudinal and behavioral change. Alexandria, VA: American Society for Training and Development.
- Davis, J. A. (1971). Elementary survey analysis. Englewood Cliffs, New Jersey, Prentice Hall
- Dillman, D. D. (1978). Mail and telephone surveys: The total design method. New York: John Wiley & Sons. Inc.
- Dixon, N. M. (1990). Evaluation: a tool for improving quality. San Diego: University Associates, Inc.
- Dixon, N. M. (1987). Meeting training's goals without reaction forms. Personnel Journal, 66(8) 108-115.
- Finch, C. R. & Crunkilton J. R. (1989). Curriculum development in vocational and technical education: Planning, content, and implementation. Boston: Allyn and Bacon, Inc.
- Giusti, J. P. (1990). Training keeps AMP connected to its customers. Training and Development Journal, 44(10) 73-80.
- Gordon, J. (1991, August). Measuring the goodness of training Training, 30(8), 19-25.
- Gutek, S. P. (1988). Training program evaluation: An investigation of perceptions and practices in nonmanufacturing business organizations. Dissertation Abstracts International, 49/05A, 1040
- Hinkle, D. E.; Wiersma, W.; & Jurs, S. J. (1988) Applied Statistics for the Behavioral Sciences. Boston: Houghton Mifflin.
- Industry report 1996. (1996, October). Training, 33(10), 36-49.
- Industry report 1988. (1988, October). Training, 25(10), 31-34.
- Jones, J. E. (1990). Don't smile about smile sheets. Training and Development Journal, 44(12) 19-21.

- Kirkpatrick, D. L. (1975a). Techniques for evaluating training programs: part 1: Reaction. In D. L. Kirkpatrick (Ed.), Evaluating training programs (pp. 1-5). Madison, Wisconsin: American Society for Training and Development.
- Kirkpatrick, D. L. (1975b). Techniques for evaluating training programs: part 2: Learning. In D. L. Kirkpatrick (Ed.), Evaluating training programs (pp. 6-9). Madison, Wisconsin: American Society for Training and Development.
- Kirkpatrick, D. L. (1975c). Techniques for evaluating training programs: part 3: Behavior. In D. L. Kirkpatrick (Ed.), Evaluating training programs (pp. 10-13). Madison, Wisconsin: American Society for Training and Development.
- Kirkpatrick, D. L. (1975d). Techniques for evaluating training programs: part 4: Results. In D. L. Kirkpatrick (Ed.), Evaluating Training programs (pp. 14-17). Madison, Wisconsin: American Society for Training and Development.
- Kirkpatrick, D. L. (1994). Evaluating training programs: the four levels. San Francisco: Berrett-Koehler.
- Kirkpatrick, D. L. (1996, January). Revisiting Kirkpatrick's Four-level model. Training & Development, 48(1), 54-59.
- Mager, R. F. (1984a). Measuring instructional results: or gotta match (2nd ed.). Belmont, CA: Lake.
- Mager, R. F. (1984b). Preparing instructional objectives. Belmont, CA: Lake.
- Mager, R. F., & Pipe, P. (1970). Analyzing performance problems: or 'you really oughta wanna'. Belmont, CA: Fearon Pitman.
- McEvoy, G. M. & Buller, B. F. (1990). Five uneasy pieces in the training evaluation puzzle. Training and Development Journal, 44(8) 39-42.
- Medsker, K. L. & Roberts, D. G. (Eds.). (1992). Trainer's toolkit: Evaluating the results of training. Alexandria, VA: American Society for Training and Development.
- Ostroff, C. (1991). Training effectiveness measures and scoring schemes: A comparison. Personnel Psychology, 44(2) 353-374).
- Phillips, J. J. (1991). Handbook of training evaluation and measurement methods. Houston: Gulf.

- Robinson, D. G. & Robinson, J. C. (1989) Training for impact: How to link training to business needs and measure the results. San Francisco: Jossey-Bass.
- Russ-Eft, D. (1992). Zenger-Miller. In K. L. Medsker & D. G. Roberts (Eds.), Trainer's toolkit: Evaluating the results of training (pp. 193-206). Alexandria, VA: American Society for Training and Development.
- Schilling, C. W. (1991). The role of company training programs to improve employee performance and company productivity. (Doctoral Dissertation, Ball State University, 1991). Dissertation Abstracts International. 52-6, A3058
- Skills gap? What skills gap? (1990). Training & Development Journal. 44(9), 9-12.
- The Joint Committee on Standards for Educational (1981). Evaluation in Standards for Evaluations of Educational Programs, Projects, and Materials. New York: McGraw-Hill.
- Toffler, A. (1970). Future shock. New York: Random House.
- Vanpelt, D. D. (1992, August/September). Education drives continuous improvement. Technical & Skills Training. 3(6), pp. 9-12.
- Wallace, S. R. Jr. & Twichell, C. M. (1953). Evaluation of a Training Course for Life Insurance Agents. Personnel Psychology. 6(1) 25-43.
- Wimbiscus, J. J. (1995). A classification and description of human resource development scholars. Human Resource Development Quarterly. 6(1) 5-34.
- Hudson Institute. (1987, May). Workforce 2000. Indianapolis, IN: Author.
- Zemke, R. (1991, May). Do performance appraisals change performance? Training. 28(5), 34-39.

APPENDIX A SURVEY



The purpose of this research, directed by Skip Twitchell, is to gather data on the use of evaluation in the technical training community. Further, this study is coordinated with a larger study (under the auspices of the ASTD research committee) directed by Dr. Elwood F. Holton III and Dr. Jack J. Phillips which is examining evaluation across all training functions. In kind support has been provided by the ASTD in the form of a mailing list.

Research associates in this study

Skip Twitchell
Training Consultant and
Graduate Fellow
School of Vocational Education
Louisiana State University
Baton Rouge, LA 70803

Dr. James Trot
Associate Dean and
Graduate Committee Chair
College of Agriculture
Louisiana State University
Baton Rouge, LA 70803

Dr. Elwood F. Holton III
Assistant Professor
School of Vocational Education
Louisiana State University
Baton Rouge, LA 70803

Dr. Jack J. Phillips
Performance Resources
Organization
P.O. 380637
Birmingham, AL 35238

Copyright 1994

The information gathered by this survey will be the basis for nationwide research into the use of various types of evaluation to measure the effectiveness of technical training. (For the purposes of this study, technical training includes any training that addresses the skills and knowledge needed to construct, assemble, operate, or repair machines, utilities, or structures.) The first four sections of this survey address Kirkpatrick's four levels of evaluation: reaction, learning, on-the-job behavior, and organizational results. Section five seeks information about your organization and organizational climate in relation to evaluation. If your area of responsibility includes training addressing other subjects or training external to the United States, please respond to each question based only on technical training which occurs in the United States.

Section 1 Measures of Participant Reaction

Section 1 relates to the use of participant reaction forms. The questions refer to any formal measure of a participant's perceptions of training taken after the completion of a program.

1. What percent of your organization's currently active training programs use participant reaction forms or other methods of gaining information concerning the participant's thoughts or feelings about various aspects of a program such as content, instruction, facilities, materials, or usefulness. These measures are taken directly after completion of the program. _____ %

If you entered 0% for question 1 above, please skip to question 3.

2. For those times when you do not gather information on participant reactions, what are the reasons?

Check all that apply. There is space for additional reasons below.

- of little value to the organization _____ ☐
 the cost in person-hours and capital _____ ☐
 evaluation takes too much time from the course _____ ☐
 lack of training or experience in using this form of evaluation _____ ☐
 union opposition _____ ☐
 not required by the organization _____ ☐
 policy prohibits the evaluation of employees by the training department _____ ☐
 training is done only to meet legal requirements _____ ☐

Other reasons:

Section 2 Measures of Learning

Section 2 relates to evaluation methods that measure the amount of learning resulting from a training program. These measures are normally in the form of post-tests (written or skills demonstrations) at course end.

3. What percent of your organization's currently active training programs use evaluation methods measuring the amount of learning resulting from training. _____ %

For the following questions, use the scale below when asked to estimate percentages. If you entered 0% for question 3 above, please skip to question 5.

1	2	3	4	5	6
0%	1-19%	20-39%	40-59%	60-79%	80-100%

4. Please estimate the percentage of programs in which your organization uses each of the various methods listed on the right to evaluate learning. Please circle the number corresponding to the percent of use.

written pre-test/post-test _____ 1 2 3 4 5 6
 written post-test only _____ 1 2 3 4 5 6
 simulations _____ 1 2 3 4 5 6
 work samples _____ 1 2 3 4 5 6
 skill demonstrations _____ 1 2 3 4 5 6
 on-the-job demonstrations _____ 1 2 3 4 5 6

In the space below, please write in any additional evaluation methods used and circle the number corresponding to the percent of use.

1 2 3 4 5 6
 1 2 3 4 5 6

5. For those times when you do not evaluate learning that took place during a training program, what are the reasons?

Check the box that applies. There is space for additional reasons below.

no time taken for this evaluation	<input type="checkbox"/>
the person doing the evaluation is not available	<input type="checkbox"/>
evaluation takes too much time from the course	<input type="checkbox"/>
lack of training or experience in using this form of evaluation	<input type="checkbox"/>
unknown position of the person doing the evaluation	<input type="checkbox"/>
not required by the organization	<input type="checkbox"/>
policy prohibits the evaluation of employees by the training department	<input type="checkbox"/>
training is done only to meet legal requirements	<input type="checkbox"/>

Other reasons:

Section 3 Measures of On-the-Job Behavior

Section 3 relates to evaluation methods that measure the amount of learning that is transferred to the job. These measures typically take place several weeks or months after a training program and measure actual use of the knowledge or skills gained during training.

6. What percent of your organization's currently active training programs use evaluation methods that measure the amount of learning that is transferred to the job? _____ %

For the following questions, use the scale below when asked to estimate percentages. If you entered 0% above, please skip to question 8.

1 0%	2 1-19%	3 20-39%	4 40-59%	5 60-79%	6 80-100%
---------	------------	-------------	-------------	-------------	--------------

7. Please estimate the percentage of programs in which your organization uses each of the various methods listed on the right to evaluate the use of learning on the job. Please circle the number corresponding to the percent of use.

anecdotal information	1	2	3	4	5	6
observation	1	2	3	4	5	6
performance appraisals	1	2	3	4	5	6
existing records other than performance appraisals	1	2	3	4	5	6
records produced specifically for evaluation purposes	1	2	3	4	5	6
surveys of the trainee's subordinates	1	2	3	4	5	6
self-report surveys	1	2	3	4	5	6
surveys of the trainee's peers	1	2	3	4	5	6
surveys of the trainee's supervisor	1	2	3	4	5	6
interviews with the trainee's subordinates	1	2	3	4	5	6
interviews with the trainee	1	2	3	4	5	6
interviews with the trainee's peers	1	2	3	4	5	6
interviews with the trainee's supervisor	1	2	3	4	5	6
focus groups	1	2	3	4	5	6
follow-up assignments	1	2	3	4	5	6
action plans	1	2	3	4	5	6
audit action plans	1	2	3	4	5	6
performance contracts with supervisor	1	2	3	4	5	6

In the space below, please write in any additional evaluation methods used and circle the number corresponding to the percent of use.

_____	1	2	3	4	5	6
_____	1	2	3	4	5	6

8. For those times when you do not evaluate the use of learning on-the-job, what are the reasons?

Check all that apply. There is space for additional reasons below.

of little value to the organization	<input type="checkbox"/>
the cost in person-hours and capital	<input type="checkbox"/>
evaluation takes too much time from the course	<input type="checkbox"/>
lack of training or experience in using this form of evaluation	<input type="checkbox"/>
union opposition	<input type="checkbox"/>
not required by the organization	<input type="checkbox"/>
policy prohibits the evaluation of employees by the training department	<input type="checkbox"/>
training is done only to meet legal requirements	<input type="checkbox"/>

Other reasons:

Section 4

Measures of Results

Section 4 relates to evaluation methods that measure **organizational change (results)**, due to a change in performance as a result of learning that occurred in a training program. These measures usually compare conditions prior to training to conditions after training has been completed and link the change to the training program.

9. What percent of your organization's currently active training programs use evaluation methods that measure **organizational results** that occur after a training program. _____ %

For the following questions, use the scale below when asked to estimate percentages. If you entered 0% above, please skip to question 11.

1 0%	2 1-19%	3 20-39%	4 40-59%	5 60-79%	6 80-100%
---------	------------	-------------	-------------	-------------	--------------

10a. Please estimate the percentage of programs in which your organization uses each of the various methods listed on the right to evaluate **organizational results**. Please circle the number corresponding to the percent of use.

In the space below, please write in any additional evaluation methods used and circle the number corresponding to the percent of use.

anecdotal information	_____	1	2	3	4	5	6
estimates of improved productivity	_____	1	2	3	4	5	6
objective measures of productivity	_____	1	2	3	4	5	6
before and after measures related to the training goal	_____	1	2	3	4	5	6
benefit/cost comparison (ROI)	_____	1	2	3	4	5	6
cost savings	_____	1	2	3	4	5	6
compliance with federal, state, and local regulations	_____	1	2	3	4	5	6
	_____	1	2	3	4	5	6
	_____	1	2	3	4	5	6

10b. If you use any type cost/benefit analysis to determine training's return-on-investment, what is the average return on investment (ROI) for those programs in which you use this type of evaluation? _____ %

11. For those times when you do not evaluate **organizational results**, what are the reasons?

Check all that apply. There is space for additional reasons below.

of little value to the organization	<input type="checkbox"/>
the cost in person-hours and capital	<input type="checkbox"/>
evaluation takes too much time from the course	<input type="checkbox"/>
lack of training or experience in using this form of evaluation	<input type="checkbox"/>
union opposition	<input type="checkbox"/>
not required by the organization	<input type="checkbox"/>
policy prohibits the evaluation of employees by the training department	<input type="checkbox"/>
training is done only to meet legal requirements	<input type="checkbox"/>

Other reasons:

Section 5a Organizational Training Practices

For questions 12 through 21, please circle the number to the right of each item that best answers each question using the percentage ranges on the scale below.

1 0%	2 1-19%	3 20-39%	4 40-59%	5 60-79%	6 80-100%
---------	------------	-------------	-------------	-------------	--------------

12. Technical training can be accomplished using many methods and/or combinations of methods. For your organization, please indicate the percentage of programs that fall into each category.

informal OJT (following some other employee's lead) 1 2 3 4 5 6
 formal OJT (planned and monitored) 1 2 3 4 5 6
 apprenticeships 1 2 3 4 5 6
 self-study 1 2 3 4 5 6
 individual training events addressing specific needs 1 2 3 4 5 6
 a curriculum based on organizational goals 1 2 3 4 5 6
 work team initiated training 1 2 3 4 5 6

In the space below, please write in any additional methods used and estimate the percent of use. The total should approximate 100%.

1 2 3 4 5 6

1 2 3 4 5

13. Please indicate the percentage of programs in which your organization starts planning the evaluation process at each of the stages listed at the right.

prior to program development 1 2 3 4 5 6
 as the first step in program development 1 2 3 4 5 6
 during program development 1 2 3 4 5 6
 after program completion 1 2 3 4 5 6
 when training program results must be documented 1 2 3 4 5 6
 evaluations are not implemented 1 2 3 4 5 6

14. Please indicate the percentage of time that program evaluation results are reported to each of the audiences on the right.

stockholders 1 2 3 4 5 6
 CEO and other upper level managers 1 2 3 4 5 6
 direct managers of those receiving training 1 2 3 4 5 6
 program participants 1 2 3 4 5 6
 training department members not directly involved in the evaluation process 1 2 3 4 5 6
 all employees 1 2 3 4 5 6

15. Employee development programs are delivered for a variety of reasons and have different levels of participation. Please indicate the percent of your programs that match the descriptions listed on the right.

employees are sent to this program as a reward 1 2 3 4 5 6
 all employees at this level have always had this program 1 2 3 4 5 6
 all employees involved in an activity or specific group attend this program 1 2 3 4 5 6
 participants will acquire new attitudes by attending this program 1 2 3 4 5 6
 participants in this program will be able to perform at a set level 1 2 3 4 5 6
 a change in organizational outcomes will result from this course 1 2 3 4 5 6

16. Approximately what percentage of the employee development staff is involved in evaluation? _____ %

17. Approximately what percentage of the employee development staff has formal training in evaluation? _____ %

18. What percentage of your training programs use the evaluation techniques listed on the right?

a matching (randomly chosen) control group _____ 1 2 3 4 5 6
 a control group selected from a similar work unit _____ 1 2 3 4 5 6
 multiple measures taken before and after a program _____ 1 2 3 4 5 6

19. Approximately what percentage of your training programs are dependent on evaluation for continued funding? _____ %

20. Approximately what percentage of your budget is dependent on evaluation results? _____ %

21. How is employee development funded in your organization?
 separate training budget _____ ☐
 training as separate profit center _____ ☐
 an administrative budget and some form of charge back for program attendance _____ ☐

Section 5b Training Managers' Perceptions of Evaluation

For questions 22 through 25, please indicate what you *think* about the *value* of the four levels of evaluation when used to improve or justify a training program. Based on the scale below, please circle the number to the right of each question that matches your opinion.

1	2	3	4	5	6
none	little	some	considerable	great	essential

22. How much value do you think reaction forms have for each item on the right.

in improving training? _____ 1 2 3 4 5 6
 in demonstrating the value of training in improving _____ 1 2 3 4 5 6
 on the job performance?
 in demonstrating the value of training to upper level management? _____ 1 2 3 4 5 6
 in demonstrating the value of training in attaining organizational goals? _____ 1 2 3 4 5 6

23. How much value do you think end-of-course measures of learning have for each of the items on the right.

in improving training? _____ 1 2 3 4 5 6
 in demonstrating the value of training in improving _____ 1 2 3 4 5 6
 on the job performance?
 in demonstrating the value of training to upper level management? _____ 1 2 3 4 5 6
 in demonstrating the value of training in attaining organizational goals? _____ 1 2 3 4 5 6

24. How much value do you think on-the-job measures of performance have for each of the items on the right.

in improving training? _____ 1 2 3 4 5 6
 in demonstrating the value of training in improving _____ 1 2 3 4 5 6
 on the job performance?
 in demonstrating the value of training to upper level management? _____ 1 2 3 4 5 6
 in demonstrating the value of training in attaining organizational goals? _____ 1 2 3 4 5 6

25. How much value do you think organizational results measures have for each of the items on the right.

in improving training? _____ 1 2 3 4 5 6
 in demonstrating the value of training in improving _____ 1 2 3 4 5 6
 on the job performance?
 in demonstrating the value of training to upper level management? _____ 1 2 3 4 5 6
 in demonstrating the value of training in attaining organizational goals? _____ 1 2 3 4 5 6

Please provide the following information about your organization and yourself.

Number of employees working in the United States

Number of U. S. employees participating in technical training last year

Number of years your organization has been providing technical training

Number of years you have been performing a training function in this or any other position

Gender

Please check the industry group that best describes your organization:

- | | | |
|--|---|--|
| <input type="checkbox"/> Accounting, Auditing, & Underwriting | <input type="checkbox"/> Computer / Data Processing | <input type="checkbox"/> Military Trainers |
| <input type="checkbox"/> Agriculture | <input type="checkbox"/> Criminal Justice System Trainers | <input type="checkbox"/> Newspaper Trainers |
| <input type="checkbox"/> Armed Forces Civilian Trainers | <input type="checkbox"/> Educational Institutions | <input type="checkbox"/> Petroleum & Natural Gas |
| <input type="checkbox"/> Associations & Non-profit Organizations | <input type="checkbox"/> Federal Trainers | <input type="checkbox"/> Retail |
| <input type="checkbox"/> Automotive | <input type="checkbox"/> Food | <input type="checkbox"/> State & Local Government Trainers |
| <input type="checkbox"/> Aviation & Space | <input type="checkbox"/> Forest Products | <input type="checkbox"/> Telecommunications |
| <input type="checkbox"/> Banking | <input type="checkbox"/> Health Care | <input type="checkbox"/> Textiles |
| <input type="checkbox"/> Chemicals & Pharmaceuticals | <input type="checkbox"/> Hospitality | <input type="checkbox"/> Transportation |
| <input type="checkbox"/> Communications, Publishing & Broadcasting | <input type="checkbox"/> Insurance | <input type="checkbox"/> Utilities |
| | <input type="checkbox"/> Manufacturing / Industrial | <input type="checkbox"/> Other |

Thank you for taking the time to fill out this questionnaire. Please use the enclosed, stamped, self-addressed envelope to return this survey to:

**Skip Twitchell
Twitchell Consultants
9879 Kinglet Dr.
Baton Rouge, LA 70809**

This code number is for following up un-returned questionnaires. To maintain your confidentiality, the list that matches your name to this code number will be destroyed after the responses are received and coded into the computer. No information regarding the individual responses will be released to anyone before the list is destroyed.

APPENDIX B

LETTERS AND POST CARDS

Cover letter sent with survey

FIELD(3) FIELD(2)
FIELD(4)
FIELD(5)
FIELD(6)
FIELD(7), FIELD(8)

DATE

Dear FIELD(3) FIELD(2),

Training literature is filled with articles on training evaluation and the need to benchmark organizational functions. Those of us in training need to choose companies that do well in the area of training evaluation to benchmark our own organizations. Present knowledge about evaluation in technical training does not give us a standard by which to choose the best.

Your organization has been chosen as part of a sample from which we can learn what business and industry is doing to evaluate technical training. In any study based on sampling, it is extremely important that every member of the sample respond. Your answers are important to the entire field of technical training. The questionnaire should be completed by the person most knowledgeable on the evaluation of technical training programs within your organization.

The names, organizations, and addresses of those participating in this research survey will be kept confidential. When the data collection portion of the study is completed, all records connecting the participants with their responses will be destroyed. However, if your organization would like to be credited with assisting with this effort, please indicate this on the enclosed postcard and you will be given official recognition in the research report.

By actively participating in this study you can be among the first to know how technical trainers evaluate and how much they evaluate. The results of this study will provide information that should help you in benchmarking your organization's technical training evaluation. To receive a summary of the results, please fill out the enclosed postcard and mail it separately. This will ensure that your organization's name is not associated with the information requested.

If you have any questions, please write or call. The telephone number is 504 / 291-3232.

Regards,

Skip Twitchell
Graduate Fellow
School of Vocational Education
Louisiana State University

Dr. James Trott
Associate Dean
College of Agriculture
Louisiana State University

Reminder post card

Dear FIELD(1) FIELD(2),

Several days ago you should have received the survey Evaluation in Technical Training: Present Practices in U.S. Business and Industry. The information gathered by this survey is important to this research and the training community.

This is just a note to thank you for your participation and insure that you did receive a copy of the survey. FIELD(1), if you did not receive a copy the survey please call collect 504 / 291-3232 and we will see that a copy is mailed promptly.

Thanks again for your time and interest in this research.

Regards,

Skip Twitchell & Dr. James Trott

First follow-up letter

FIELD(3) FIELD(2)

DATE

FIELD(4)

FIELD(5)

FIELD(6)

FIELD(7), FIELD(8)

Dear FIELD(3) FIELD(2),

Those of us in technical training need know how evaluation is being used support and improve training. Present knowledge about evaluation in technical training does not give us this information. This information is important to the entire training community. We must know where we are before we can move ahead.

Your organization was specifically chosen as part of the survey sample and your responses are important if this research is to provide information representative of the entire training community. You will be providing important information for all technical trainers. The questionnaire should be completed by your organization's most knowledgeable person concerning the evaluation of your technical training programs. In the event that the original survey was not received for some reason or was misplaced a replacement is enclosed.

The names, organizations, and addresses of those participating in this research survey will be kept confidential. When the data collection portion of the study is completed, all records connecting the participants with their responses will be destroyed. However, if your organization would like to be credited with assisting with this effort, please indicate this on the enclosed postcard and you will be given official recognition in the research report.

By actively participating in this study you can be among the first to know how technical trainers evaluate and how much they evaluate. The results of this study will provide information that should help you in benchmarking your organization's technical training evaluation. To receive a summary of the results, please fill out the enclosed postcard and mail it separately. This will ensure that your organization's name is not associated with the information requested.

Your cooperation is greatly appreciated.

Regards,

Skip Twitchell
Graduate Fellow
School of Vocational Education
Louisiana State University

Dr. James Trott
Associate Dean
College of Agriculture
Louisiana State University

P.S. Several people have written asking when the results will be available. We hope to provide summaries to those who have requested them within the next 90 days.

Second follow-up letter.

FIELD(3) FIELD(2)
 FIELD(4)
 FIELD(5)
 FIELD(6)
 FIELD(7), FIELD(8)

DATE

Dear FIELD(3) FIELD(2),

I am writing to about the study Evaluation in Technical Training: Present Practices in U.S. Business and Industry. We have not yet received your completed survey. We have had a large number of the surveys returned. But, whether we can use the information collected as being representative of all technical trainers in ASTD depends on you and others who have not yet responded.

Your organization may have every different practices than those described in the surveys that have been returned. This is the first nationwide study of this type. Therefore, the results are important to the entire training community. Your responses are important. Only with a large majority of participants responding can we truly reflect what is current practice in technical training.

In case our other correspondence did not reach you, a replacement is enclosed. The survey should be completed by the person that you feel has the most knowledge concerning the evaluation of technical training within your organization.

We will be happy to send you a copy of the results. The survey asks if you would like a free summary of the research and provides a separate card for you to enter the name and address to which this information should be sent. This information will be separated from your responses on our receipt of the completed survey and will in no way be connected to your responses.

Your cooperation in the study will be greatly appreciated.

Regards,

Skip Twitchell
 Graduate Fellow
 School of Vocational Education
 Louisiana State University

Dr. James Trott
 Associate Dean
 College of Agriculture
 Louisiana State University

P.S. Several people have written asking when the results will be available. We hope to provide summaries to those who have requested them within the next 90 days.

APPENDIX C

COMPARISON OF RESPONDENTS AND NON-RESPONDENTS ON SELECTED VARIABLES

Comparison of Respondents and Non-respondents on Selected Variables

No.	Variable	n_1	n_2	T	p
1	Level 1 - Evaluation	20	111	1.242	.216
2	Level 2 - Evaluation	20	111	-0.173	.863
3	Level 3 - Evaluation	20	111	-0.753	.453
4	Level 4 -Evaluation	20	111	-0.198	.843
5	Courses that are informal OJT	20	110	-1.852	.066
6	Courses that are formal OJT	20	110	-0.704	.483
7	Courses that are apprenticeships	20	110	-1.204	.231
8	Courses that are self-study	20	110	0.317	.752
9	Courses that are one-time training events	19	110	-1.294	.198
10	Courses that are part of a curriculum	20	109	-0.262	.794
11	Courses that are based on team initiated training	19	110	-0.065	.948
12	Evaluation planning starts prior to development	20	110	0.194	.847
13	Evaluation planning is the first step in development	20	110	2.013	.046*
14	Evaluation is planned during development	20	110	0.409	.683
15	Evaluation is planned after program completion	20	110	0.952	.343
16	Evaluation is planned only when results must be documented	19	110	1.077	.283
17	Evaluations are not implemented	20	110	-0.077	.939
18	Percentage of programs dependent on evaluation	20	109	2.688	.008*
19	Total number of employees	20	107	-0.115	.909
20	Number of employees in training in 1994	20	103	-1.296	.197
21	How long the company has been providing technical training	20	105	-1.667	.098
22	Number of years experience the respondent has in training	20	109	0.711	.478

Note: $p < .05$ is indicated by an *

Note: n_1 is the number of non- respondents and n_2 is the number of respondents.

APPENDIX D

TABLES FOR THE WRITTEN RESPONSES TO THE METHODS USED TO EVALUATE AT EACH LEVEL

Table D1

Response and "Percentage of Programs Using Each Level 2 Evaluation Method" for Respondents Answering Other

Method	Percentage of programs using the method
Computerized post-tests	60-79%
Post training survey	1-19%
Self-evaluation, i.e.: skill level rating	40-59%
Module tests ¹	80-100%
Comprehensive final ¹	80-100%
Combination Written & skills demo.	60-79%
Oral post-test	20-39%
Performance to labor efficiency standard	80-100%
Competency based training w/self evaluations to follow-up	1-19%
Checklists administered by the trainer	1-19%
Video tape actual performance and review with instructor ²	1-19%
Actual successes/success rate with new approach or method ²	40-59%
Oral review quiz	40-59%
Peer review	60-79%

Note. ^{1, 2} responses from same respondent.

Table D2

Response and "Percentage of Programs Using Each Level 3 Evaluation Method" for Respondents Answering Other

Method	Percentage of programs using the method
Performance review for trainees who are involved in accidents or incidents during 6 month probationary period (number varies)	20-39%
Customer letters (usually positive) ¹	20-39%
Customer satisfaction surveys ¹	60-79%
Measure customer complaints - trained vs non-trained	60-79%
Monthly failure analysis reports. (used to indicate efficiency of repair work)	80-100%
Customer input	40-59%
OJT - performance standards for each job are evaluated during on-the-job training. i.e. claims entered, claims entered correctly etc. for data entry trainees	1-19%
Peer / Practice Reviews	80-100%
Increased sales ²	80-100%
Customer satisfaction surveys of internal & external customers ²	80-100%
Productivity measures ²	80-100%
Survey customer satisfaction levels with outcomes/services ³	40-59%
Gather data on total outcomes ³	60-79%
Commission dollars are a direct measure of training success	80-100%
<u>Note.</u> ^{1, 2, 3} responses from same respondent.	

Table D3

**Response and "Percentage of Programs Using Each Level 4
Evaluation Method" for Respondents Answering Other**

Method	Percentage of programs using the method
Reduced number of field engineer visits/customer satisfaction? 1 call per year X 5 years ...	Missing
Again we survey our dealers and they report 25% to 40% increases after training	40-59%
Sales results	1-19%
Observed changes in behavior	80-100%
Safety Record	80-100%

APPENDIX E

REASONS FOR NOT EVALUATING WRITTEN IN BY THE RESPONDENTS

Level 1

1. Only if instructor "forgets"
2. After a specific class is taught over + over, need to eval. after every class is not useful.
3. We do training for other areas of the company that are not our direct responsibility.
4. We sometimes do not evaluate - usually an oversight.
5. Some classes/students object
6. Instructor neglects to implement
7. These evaluations provide little real information about the success or failure of a program to meet the business need.
8. None of the above - We get information for all courses.
9. Short courses (2-6 hours) are not evaluated. Most of those are required by OSHA or EPA.

Level 2

1. Only if instructor "forgets"
2. Potential for the employee challenge of test validity (i.e., culture bias)
3. Time
4. This type of evaluation not built into course design

5. In prior years, no one has "owned" training and therefore it was not formalized.
6. Measurement of "soft skills" management (leadership types of training) is an ongoing measurement.
7. Difficult to measure a lot of the training provided. Have not located a measurement tool that would work for us.
8. Usually an oversight by the trainer - especially when training area that are not our direct responsibility
9. There are some areas on which we cannot agree on the best method to evaluate.
10. Difficulty in setting measurable ways to evaluate learning in soft-skill situations
11. We have not been able to test new employees because our current employees were flunking the test.
12. This factory has many repetitive manual operations, little thinking is required (or encouraged), Job tasks are demonstrated by lead operators, then its sink or swim baby; you're on your own!
13. Culture - the perceived cost /benefit of evaluating all programs is doubtful
14. Learning is not easily measured in behavior enhancement programs
15. Management opposition
16. Management in some departments does not believe that such evaluations are important, so won't allow time for evaluations (similar to the first 2 choices above)

17. Evaluation method dictated by type of training (skill-based vs. informational or inspirational)
18. Most training is one-on-one so the trainer works with the trainee over and over until the skill/knowledge is mastered.

Level 3

1. Extremely difficult to evaluate on-the-job performance when trainee is from unrelated company
2. Lack of time and interest
3. We ask people to rate their own performance before/after training rather than doing specific testing.
4. Not built into course design; we try to eliminate fear
5. The company doesn't understand how to train successfully so they don't support any activities outside the classroom.
6. Time and budget restraints
7. We train non-company individuals, and do not have access to on-site data
8. Working as a single person training dept. this falls through the cracks.
9. Staffing levels inadequate to conduct
10. No follow up by supervisors
11. No method in place
12. Takes too much time from work
13. Lack of means to measure transfer

14. Lack of resources to create/monitor evaluation techniques. We will be using this more in the future, however.
15. For indirect labor where production standards (labor efficiencies) cannot or have not been established
16. The perceived cost/benefit of evaluating all programs is doubtful
17. Feasibility getting multiple sites to gather data is weighed against project requirements. Paper compliance always a problem!
18. I'm not sure what happens out on the job. The foremen + supervisors don't disclose this information with me. I don't know if they are aware of exactly what is being covered in class.
19. Our training involves individuals all over North America - no structure (formal) exists for continued evaluation (other than success/fail) plus reports from supervisors
20. One-on-one training is on-the-job.

Level 4

1. Constantly changing organizational environment and strategies makes establishing effective of evaluation difficult - and sometimes counter productive
2. Not built into course design
3. We don't know how to measure it at the organizational level.
4. The company doesn't understand training for impact.
5. Time and budget restraint

6. Direct benefit to the Bottom Line is subjective.
7. Staffing levels inadequate
8. Management and institutional inertia
9. Takes time
10. No method in place
11. Training involves people at client organization that do not permit measurement
12. Lack of resources - the training function is a newly - created department.
13. The perceived cost benefit of evaluating all programs is doubtful
14. Difficulty of identifying valid measures
15. Our projects are complex multistage/multi training type - Measuring organizational change due to "a" training would be tedious. Would it provide an ROI? Not sure

VITA

Skip Twitchell was born November 13, 1948, in Norway, Maine. The first 12 years of schooling were in the same three story white clapboard building housing the entire student population of Woodstock, Maine. Those same 12 years included technical training of all types on a small rural farm. The United States Army provided additional training and practice where he earned two Army commendations for meritorious service in actions against hostile forces.

Skip holds an "Airframe and Powerplants" mechanic's license from the Federal Aviation Administration and two associate degrees from Enterprise State Junior College, one in aviation management and the other a pre-engineering science degree, 1975 and 1976. His bachelors of science degree is in Industrial Arts Education from Auburn University, 1979.

From 1968 to 1985 Skip held several teaching and maintenance technician positions. In 1986, he became actively involved in curriculum design, training trainers, and delivering technical courses on industrial automation, basic workplace skills, and teaching older workers. Since 1990 he has been a full time consultant in his own business, Twitchell Consultants.

Skip's technical publications include 17 texts in industrial automation, two articles in refereed journals on methods of instruction and teaching older workers, and many presentations on teaching and automation training.

Currently, he is a candidate for a doctor of philosophy degree in Vocational Education at Louisiana State University. Skip's doctoral studies were started in August of 1990.

DOCTORAL EXAMINATION AND DISSERTATION REPORT

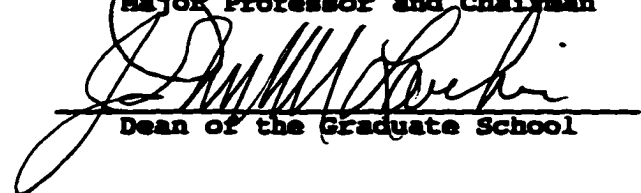
Candidate: Skip Twitchell

Major Field: Vocational Education

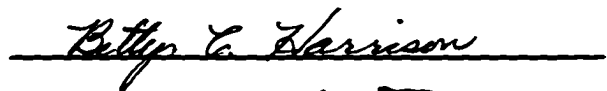
Title of Dissertation: Technical Training Program Evaluation: Present Practices in United States' Business and Industry

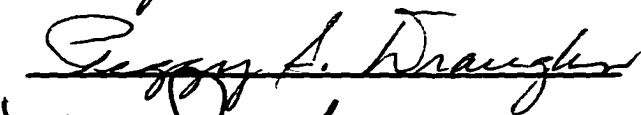
Approved:


Major Professor and Chairman

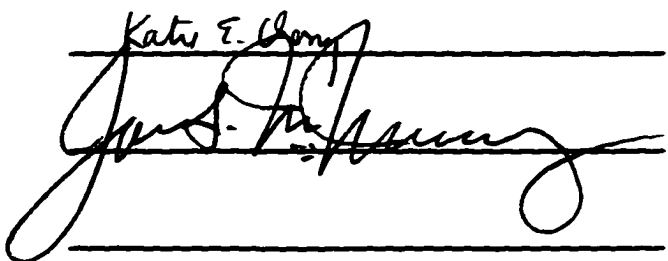

Dean of the Graduate School

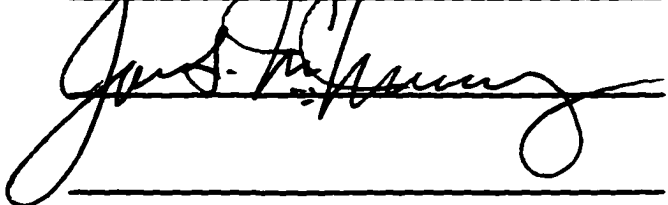
EXAMINING COMMITTEE:


Betty C. Harrison


Peggy A. Draughn


Michael Burnett


Kate E. Long


James H. Harrison

Date of Examination:

March 20, 1997