

TRAINING PROFESSIONALS' USAGE AND UNDERSTANDING OF
KIRKPATRICK'S LEVEL 3 AND LEVEL 4 EVALUATIONS

by

Perri Estes Kennedy

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in Instructional & Performance Technology

Boise State University

August 2012

© 2012

Perri Estes Kennedy

ALL RIGHTS RESERVED

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

Perri Estes Kennedy

Thesis Title: Training Professionals' Usage and Understanding of Kirkpatrick's Level 3 and Level 4 Evaluations

Date of Final Oral Examination: 18 April 2012

The following individuals read and discussed the thesis submitted by student Perri Estes Kennedy, and they evaluated her presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Seung Youn Chyung, Ed.D. Chair, Supervisory Committee

Donald J. Winiecki, Ed.D., Ph.D. Member, Supervisory Committee

Robert O. Brinkerhoff, Ed.D. Member, Supervisory Committee

The final reading approval of the thesis was granted by Seung Youn Chyung, Ed.D., Chair of the Supervisory Committee. The thesis was approved for the Graduate College by John R. Pelton, Ph.D., Dean of the Graduate College.

ACKNOWLEDGEMENTS

For their assistance in the completion of this thesis, I am in debt to many people.

Dr. Yonnie Chyung has been my endlessly patient advisor and sounding board throughout the entire process. She has always had a sensible solution to each of my new-researcher qualms and I could not have completed this project without her experience and wisdom to help me. From Dr. Don Winiecki, I learned how to channel my natural cynicism into intellectual inquiry and to never mistake assertions for truth. Dr. Robert Brinkerhoff's methodology guided me through this research and I am grateful for his encouraging words.

Thank you to the LinkedIn group managers who allowed me to solicit their members for this research. I would particularly like to thank Wendy Kirkpatrick (Kirkpatrick Evaluation), Christopher Pappas (Instructional Design and E-Learning Professionals Group), and Jenny Thompson (eLearning Guild) for their assistance.

I have learned a tremendous amount from my peers in the IPT program and continue to learn from them beyond the classroom. My special thanks go to Chester Stevenson and Shannon Rist with whom I'd worked from the first group project to the culminating activities. We've spent a lot of late nights talking about things like metacognition and ice cream, sometimes in the same sentence, and I'm so happy to have gone through this program with such good friends.

Most of all, I want to thank Erik Kennedy for his loving support even though he still isn't entirely sure what I've been studying for the last two years.

ABSTRACT

Behavior-based (Level 3) and results-based (Level 4) evaluations of training interventions can provide organizations with substantive proof of the value of those interventions. Training professionals have long acknowledged the necessity of conducting such evaluations, yet Level 3 evaluations are conducted for only about half of all training interventions and Level 4 for about one-third. This research examined the frequency with which training professionals currently conduct Level 3 and Level 4 evaluations, their perceptions on the utility of Level 3 and Level 4 evaluations, and the factors that facilitate or obstruct their attempts to perform such evaluations.

The research was conducted using Brinkerhoff's Success Case Method as its framework. Sixty-eight training professionals completed an online survey to assess their usage and understanding of Level 3 and Level 4 evaluations, indicate their success or non-success to conduct these evaluations, and rate the factors which may have contributed to their success or non-success. Twenty-two of the survey participants were interviewed to collect more in-depth information about their perceptions of these factors and how they impacted attempts to evaluate training interventions at their organizations.

The survey found that 43.47% of the training professionals surveyed conducted Level 3 evaluations at least some of the time, with only 26.08% conducting them on more than 60% of their training interventions. At Level 4, 18.41% of the training professionals conducted evaluations at least some of the time, with 13.15% conducting them on more

than 60% of their training interventions. The three key factors identified by survey respondents as impacting their ability to conduct Level 3 and Level 4 evaluations were the availability of resources such as time and personnel, managerial support, and expertise in evaluative methodology. The interview data supported the survey findings but also showed that expertise also extended to an understanding of what a results-based evaluation can measure and how it is relevant to an organization; if the training professional cannot clarify the relevance of evaluating training interventions in terms of organizational goals, the organization may not see the value in expending the resources needed to conduct evaluations.

The research findings indicated a need to further explore how training professionals interpret Level 3 and Level 4 and how they can better develop their evaluative expertise, which in turn may increase the effectiveness in gaining organizational support for evaluation efforts.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
ABSTRACT	vi
LIST OF TABLES	xii
LIST OF FIGURES	xiv
CHAPTER ONE: INTRODUCTION.....	1
Introduction.....	1
Significance of the Problem.....	7
Research Questions	8
Definition of Terms.....	10
CHAPTER TWO: LITERATURE REVIEW	13
Part 1: Human Performance Technology.....	13
Human Performance Technology and the Role of Evaluation	13
Kirkpatrick’s Four Levels of Training Evaluation.....	13
Part 2: Evaluation of Training Programs	17
The Importance of Instructional Evaluation	17
Gauging the Effectiveness of Training	18
Making the Business Case for Training	19
Frequency of the Levels.....	20
Barriers to Evaluation at Level 3 and Level 4	23

Part 3: The Importance of Evaluation to Human Performance Technology	26
CHAPTER THREE: METHODOLOGY	28
Population and Sampling	29
Instruments.....	30
Overall Methodology Framework.....	32
Implementing the Success Case Method	34
CHAPTER FOUR: DATA ANALYSIS AND FINDINGS	42
Part 1: Survey Data	42
Demographics	42
Frequency of Evaluations	44
Perceptions About Evaluation.....	46
Level 3 Stakeholders.....	49
Level 3 Reasons for Evaluation	50
Level 3 Factors Potentially Impacting Evaluation.....	51
Level 4 Stakeholders.....	52
Level 4 Reasons for Evaluation	53
Level 4 Factors Potentially Impacting Evaluation.....	54
Level 3 and Level 4 Facilitating Factors.....	55
Part 2: Interview Data	57
Facilitating Factors for Level 3 and Level 4 Evaluations	59
Obstructing Factors for Level 3 and Level 4 Evaluations	64
The Organizational Factor	67
CHAPTER FIVE: DISCUSSIONS AND CONCLUSIONS	71

Discussion	71
Frequency of Level 3 and Level 4 Evaluations	71
Stakeholders	72
Reasons for Evaluation	72
Barriers and Facilitators	73
Conclusions.....	76
Limitations of the Study.....	77
Recommendations for Future Studies.....	78
REFERENCES	80
APPENDIX A.....	83
Gilbert’s Behavior Engineering Model.....	83
APPENDIX B	86
Boise State University Institutional Review Board Notification of Approval	86
APPENDIX C	89
Solicitation for Survey Participants	89
APPENDIX D.....	91
Survey Cover Letter	91
APPENDIX E	94
Full Survey.....	94
APPENDIX F.....	116
Written Informed Consent Letter and Verbal Informed Consent Script.....	116
APPENDIX G.....	121
Interview Scripts	121

APPENDIX H.....	127
Reprint Permission from John Wiley & Sons for Use of Copyrighted Figures.....	127

LIST OF TABLES

Table 1. Gilbert's Behavior Engineering Model (BEM).....	3
Table 2. Kirkpatrick's Four Levels of Evaluation.....	4
Table 3. Pulichino's Survey Results on Frequency of Organizational Kirkpatrick Usage, Listed by Percentage.....	5
Table 4. Pulichino 2006 Survey Results on Usage of Kirkpatrick Levels.....	21
Table 5. ASTD 2009 Survey Results on Usage of Kirkpatrick Levels.....	21
Table 6. Pulichino 2007 Survey Results on Value of Evaluation Data.....	22
Table 7. ASTD 2009 Survey Results of Value of Evaluation Data.....	22
Table 8. Pulichino 2007 Survey Results on Perceived Barriers to Level 3 Evaluation....	24
Table 9. Pulichino 2007 Survey Results on Perceived Barriers to Level 4 Evaluation....	24
Table 10. ASTD 2009 Survey Results on Perceived Barriers to Level 3 and Level 4 Evaluations.....	25
Table 11. Structure for Impact Model Based on Gilbert's BEM Components.....	35
Table 12. Demographic Data for Survey Respondents.....	42
Table 13. Estimates of % of Training Evaluated, By Level.....	45
Table 14. Perceived Sufficiency of Evaluation Results for Judging Training Effectiveness.....	46
Table 15. Perceived Importance of Conducting Each Level of Evaluation.....	47
Table 16. Comparison of Perceived Sufficiency and Importance of Level, $n=59$	48
Table 17. Perception of Positive Correlation Between Kirkpatrick Levels, $n=68$	48
Table 18. Number of Respondents by Evaluation Level and Category.....	49

Table 19. Stakeholders for Level 3 Evaluations.	49
Table 20. Reasons for Evaluation at Level 3.	50
Table 21. Reasons for Non-Evaluation of Level 3.	51
Table 22. Availability of Level 3 Evaluation Factors, By Respondent Category.....	51
Table 23. Stakeholders for Level 4 Evaluations.	52
Table 24. Reasons for Evaluation at Level 4.	53
Table 25. Reasons for Non-Evaluation at Level 4.....	54
Table 26. Availability of Level 4 Evaluation Factors, By Respondent Category.....	55
Table 27. Facilitating and Obstructing Factors For Both Levels.....	56
Table 28. Interviewees by Success Categories.	58
Table A.1. Gilbert's Behavior Engineering Model Including Definitions.....	85

LIST OF FIGURES

Figure 1. Human Performance Technology (HPT) Model.	2
Figure 2. Deductive coding structure sample.	59

CHAPTER ONE: INTRODUCTION

Introduction

The International Society for Performance Improvement (ISPI) (2012) explains Human Performance Technology (HPT) as a systematic approach to improving productivity and performance in the workplace through influencing human behaviors and accomplishments. The HPT approach examines the current and desired levels of performance (performance analysis), the probable reasons for the difference between the two (cause analysis), and solutions to resolve the difference (intervention selection, design, and implementation). Figure 1 illustrates the systematic approach of HPT, with analysis leading to interventions.

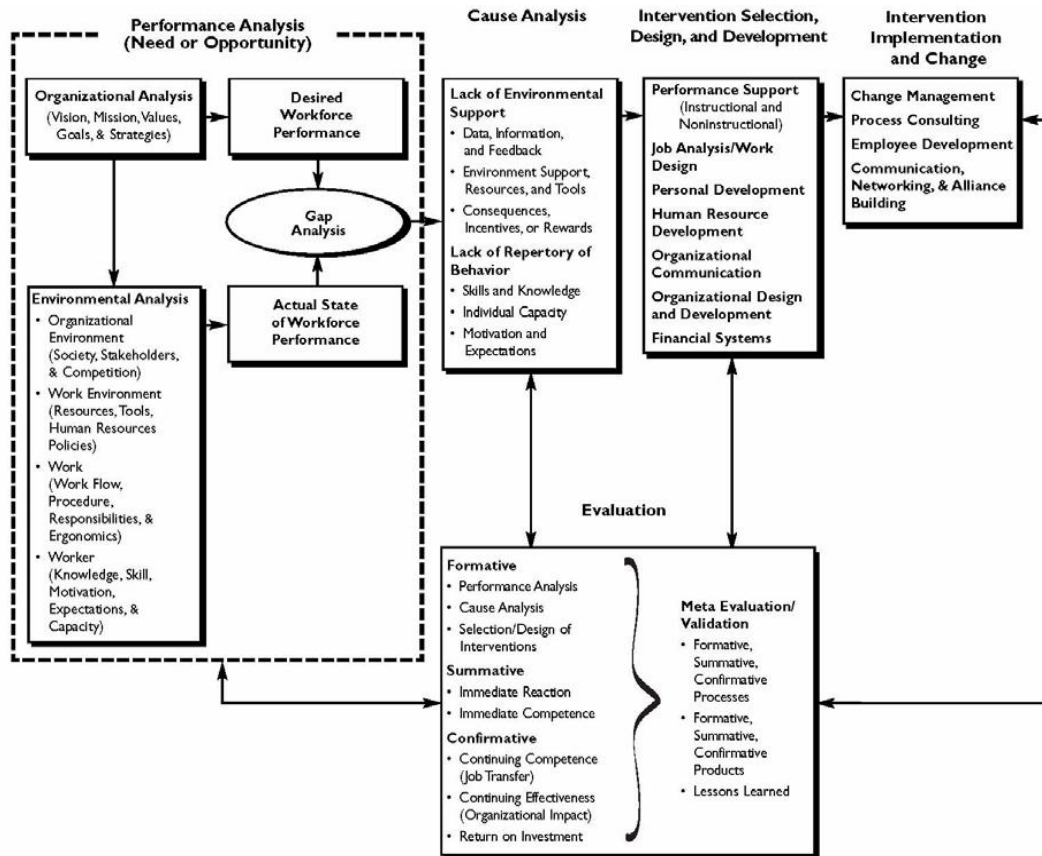


Figure 1. Human Performance Technology (HPT) Model.

Note. Reprinted with permission. Copyright 2004. The International Society for Performance Improvement. HPT model is from page 3 of *Fundamentals of Performance Technology*, Second Edition by D.M. Van Tiem, J.L. Moseley, and J.C. Dessinger. All rights reserved.

Tying the HPT process together is the concept of evaluation, which Scriven (1991) describes as the systematic determination of the intrinsic merit or extrinsic worth of something. Performance analysis and cause analysis are formative evaluations designed to give feedback for performance improvement and to guide decision-making when weighing intervention alternatives. After implementation, performance interventions are subject to summative evaluation to determine whether the interventions accomplished their intended results (Dessinger, Moseley, & Van Tiem, 2011).

One tool commonly used for cause analysis is Thomas Gilbert's (2007) Behavior Engineering Model (BEM), shown in Table 1 (for more information about the BEM, see Appendix A). The BEM identifies information, instrumentation, and motivation as the three types of workplace behavior, with each existing at both the organizational and individual levels.

Table 1. Gilbert's Behavior Engineering Model (BEM).

	Information	Instrumentation	Motivation
Environmental Factors (Organizational)	Data	Resources	Incentives
Personal Repertoire (Individual)	Knowledge	Capacity	Motives

According to Gilbert (2007), workplace performance can be improved through using the BEM to take a systematic look at what he determined to be the six possible causes of performance deficiencies and then directly or indirectly changing the specific behaviors identified as those that prevent optimal performance.

Organizations often turn to instructional interventions when faced with real or perceived knowledge deficiencies that affect workplace performance. Training is a generic term for instructional interventions and refers to any effort to change behavior through learning activities, whether it be teaching repetitive psychomotor tasks or developing analytical skills or anything in between. In order to remedy knowledge deficiencies, an organization's training department is tasked with creating training programs with the goal of changing workplace behavior that, in turn, will produce the desired outcome for the organization.

U.S. organizations spent over \$78 billion on internal training programs in 2007 (American Society for Training & Development, 2010); thus, it would seem prudent that organizations attempt to determine, through valuations of on-the-job relevancy and performance results, if this was money well spent.

One means for doing this is the Kirkpatrick scale. Kirkpatrick (2010), in his doctoral dissertation research, proposed what later came to be known as his Four Levels for evaluation of training programs. The four levels are described in Table 2.

Table 2. Kirkpatrick's Four Levels of Evaluation.

Level	Focus of Evaluation
1: Reaction	How favorably the learners react to the instruction
2: Learning	How well the learners learn the knowledge or skills imparted during the instruction
3: Behavior	To what level the learners apply the new knowledge or skills in their on-the-job behaviors
4: Results	To what level the instruction achieves the intended impact on workplace outcomes

Kirkpatrick's Four Levels are broadly accepted as the standard model in use by the field, and training professionals frequently use Kirkpatrick's terminology when referring to evaluations of training interventions (Holton III, 1996).

Harless (1976) defined "true" training evaluation as that which examined on-the-job relevancy of the instructional content and the results produced by the training; this correlates to evaluations at Level 3 and Level 4 in Kirkpatrick's terminology.

Professional organizations and academic programs in training-related fields emphasize the importance of Level 3 and Level 4 evaluations in order to accurately gauge the relevance and effectiveness of training programs.

What happens in actual practice? Pulichino (2007) surveyed over 400 training professionals in 2006 and discovered that although a large percentage of them conducted Level 1 (reaction) and Level 2 (learning) assessments, far fewer conducted evaluations at Level 3 (behavior) and Level 4 (results). Table 3 represents his findings.

Table 3. Pulichino's Survey Results on Frequency of Organizational Kirkpatrick Usage, Listed by Percentage.

Kirkpatrick Level	Frequency of Usage				
	Never	Rarely	Sometimes	Frequently	Always
1: Reaction	3.1	5.2	7.2	23.3	61.2
2: Learning	4.0	11.0	28.0	37.0	19.1
3: Behavior	13.0	34.6	32.5	14.1	5.8
4: Results	33.0	40.5	12.8	9.4	4.3

Despite the training field acknowledging the value of assessing training interventions at all levels as demonstration of their organizational impact (Rummler & Brache, 1995, p. 198), only 19.9% of the surveyed professionals reported that their organizations always or frequently assessed post-training behaviors (Level 3) and 13.7% always or frequently assessed post-training business outcomes (Level 4) (p. 81).

If training organizations and academic programs advise professionals that the goal of a training intervention is to elicit preferred behaviors and desired outcomes, why do so few professionals evaluate their effectiveness in those goals? What are the barriers to performing such evaluations? Pulichino's survey results suggest that the barriers may be both organizational and individual. For example, when asked to select one or more reasons why Level 4 evaluations were not performed, 90.5% of respondents cited difficulty in accessing the data needed, and 64.7% blamed their own lack of expertise in conducting such evaluations (pp. 87-91).

One factor that Pulichino did not explore was the respondents' interpretation of the Kirkpatrick levels. Did they believe that Kirkpatrick's Four Levels was a model describing a progression of linked evaluations? Critics of Kirkpatrick's work point to this variable interpretation as a liability, as the concept of "levels" implies a series of linked evaluations that build upon prior results. Alliger and Janak (1989) discussed the common assumptions of causality, correlation of results, and relative value and concluded that there was insufficient evidence to validate any such relationships between the levels. Holton (1996) characterized Kirkpatrick's Four Levels as taxonomy rather than as a functional model. So how do training professionals in the field see the levels? If they believe the common assumptions, does this impact how, or if, they conduct evaluations of behavior or results? If they believe in the causality between levels, would they assume that positive results automatically follow positive behavior?

Another question of interpretation is how "results" are defined by training professionals. If there is no universally accepted definition of what a Level 4 evaluation examines, does that affect the likelihood that Level 4 evaluations are conducted in different organizations? In its survey on evaluation, the American Society for Training and Development (ASTD) (2009b) found that organizations that reported conducting Level 4 evaluations most often measured post-training customer satisfaction (39%) and employee satisfaction (37%) levels, followed by measurements of post-training competency levels and the perception of training impact by both learners and their supervisors. However, while improvements in these areas may have a positive business impact, these measurements do not look at how well the training program achieved its ultimate purpose. What is the desired outcome of a training program on customer service

skills? Is it designed to improve customer satisfaction survey results, or to improve customer retention? Of the eight potential measurements listed on ASTD's survey, the one least often selected as one performed by organizations as part of a Level 4 evaluation was business outcomes.

Significance of the Problem

U.S. organizations spend billions of dollars each year on training interventions in an effort to improve the performance of their employees and in turn improve the business outcomes of the organization (ASTD, 2010). However, an ASTD survey in 2009 revealed that 54.5% of training professionals reported that recent economic conditions have required at least a moderate adjustment to their departments' operations, and that 34.7% reported that their yearly training budget had been reduced (ASTD, 2009a). If a training department hopes to retain its budget and establish its importance to the organization, it needs a way to substantiate its real worth to the organization (McLinden, 1995; Rummier & Brache, 1995).

Stufflebeam and Shinkfield (2007) define the evaluation process as a way to show the reliability, effectiveness, and efficiency of a given program. Evaluation of training interventions is a concrete way a training department can measure its results and thus show both its value to the organization and its commitment to producing a positive return on the organization's investment in training.

Measurements of learners' reactions to a training intervention (Kirkpatrick Level 1) and knowledge level right after the intervention (Kirkpatrick Level 2) are common but only show the immediate impact of the intervention. If the goal of the intervention was to change on-the-job behaviors in an effort to achieve a desired business outcome, a training

department intending to prove its business value should evaluate how the intervention actually changed those behaviors (Kirkpatrick Level 3) and impacted business outcomes (Kirkpatrick Level 4).

There appears to be scant published research on the extent to which training professionals evaluate their product at the behavior and outcome levels and the factors that prevent or promote such evaluations. This thesis project was proposed as a way to add to that research, in an attempt to better understand when and how training professionals conduct behavior-based and results-based evaluations in their organizations. If training professionals can recognize the potential barriers to evaluation within themselves and their organizations, they can create strategies to reduce or remove such barriers. Conversely, if they can identify the factors that facilitate evaluation, they may be more willing or able to make use of these factors.

Research Questions

This research was conducted to understand training professionals' conduct of behavior-based (Kirkpatrick Level 3) and results-based (Kirkpatrick Level 4) evaluations of training interventions and to examine factors that influence their conduct of such evaluations. This study was focused on individuals whose primary job duties include internal training functions for the organizations with which they are employed.

In this research, training professionals were classified into three categories:

Success Cases: These training professionals have successfully conducted Level 3 and/or Level 4 evaluations of at least one training program in their organizations, with results submitted to the stakeholders.

Non-Success Cases: These training professionals have attempted to conduct Level 3 or Level 4 evaluations of at least one training program in their organizations but were prevented by one or more factors from obtaining results that could be submitted to the stakeholders.

Non-Evaluators: These training professionals have not attempted to conduct Level 3 or Level 4 evaluations in their organizations.

The purpose of this research was to answer two research questions:

Research Question 1: With what frequency do training professionals conduct Level 3 and/or Level 4 evaluations for their organizations?

Sub-question 1a: Who are the stakeholders for these evaluations?

Sub-question 1b: For what reasons do training professionals conduct or attempt to conduct the evaluations?

Research Question 2: What factors act as facilitators or barriers to conducting Level 3 and Level 4 training evaluations?

Sub-question 2a: For Success Cases, what are the facilitating factors and the barriers, and how did they impact the performance of evaluations?

Sub-question 2b: For Non-Success Cases, what are the facilitating factors and the barriers, and how did they impact the attempts to perform evaluations?

Sub-question 2c: For Non-Evaluators, what are the facilitating factors and the barriers, and why were evaluations not attempted?

Prior research studies had identified factors that enhance or restrict the ability of training professionals to conduct behavior-based and results-based training evaluations. The promoting factors tended to be environmental in nature, driven by the needs and requirements of the organization. When Moller and Mallin (1996) surveyed training professionals, they found that evaluation duties were included in the official job description of 60% of respondents, and 77% of the organizations with formal training programs require the conducting of evaluations. Pulichino (2007) found that organizations that based their training expenditures on the need to combat competitive pressure and to maintain a knowledgeable workforce were the most likely to conduct Level 3 and Level 4 training evaluations. Barriers to conducting training evaluations could be environmental, such as a culture resistant to sharing confidential data, or individual, such as a lack of expertise in evaluation methods. The data collection instruments for this research were designed to look at both environmental and individual factors.

Definition of Terms

Training professionals are individuals whose primary job functions involve the improvement of the workplace-related knowledge and skills of an organization's employees through educational methods. The training professional may be involved with assessing knowledge and skill deficiencies, designing and developing educational programs, delivering educational programs to employees, and evaluating the effectiveness of training interventions in terms of improved employee performance.

A training intervention or training program is an educational program developed by training professionals in order to resolve a knowledge or skills deficiency. Training programs may vary in duration and complexity depending on an organization's needs.

A training evaluation is the systematic analysis of a training program to determine whether it has achieved the goals established for the program. To develop an evaluation process, the training professional must determine whether the purpose of the evaluation is to determine a program's quality (merit) or value to the organization (worth), to look for aspects of the program that need improvement, or both (Davidson, 2005).

Behavior-based evaluation is a systematic analysis of a training program to determine whether it has changed actual on-the-job behaviors of the employees who completed the program. In other words, a behavior-based evaluation examines what was learned and how much of it was put to use by the employees. If the intended behavior resulting from a training program is a more consistent use of Customer Relationship Management (CRM) software for a sales team, the behavior-based evaluation might examine the extent to which the sales employees apply the knowledge learned to use the software and for what purposes. Level 3 of Kirkpatrick's (1996) Four Levels of evaluation is a behavior-based evaluation.

Results-based evaluation is a systematic analysis of a training program to determine whether it has contributed towards the organizational goal that necessitated the training program. In other words, a results-based evaluation examines why a training program was needed and if it met that need after changing employee behaviors. If the intended organizational goal for the aforementioned CRM training program was to increase sales to small client companies previously overlooked by the sales team, a

results-based evaluation might examine the change in sales income for this specific client category and the change in client satisfaction levels. Level 4 of Kirkpatrick's (1996) Four Levels of evaluation is a results-based evaluation.

CHAPTER TWO: LITERATURE REVIEW

Part 1: Human Performance Technology

Human Performance Technology and the Role of Evaluation

Pershing (2006) defined Human Performance Technology (HPT) as the design and development of results-oriented systemic interventions meant to improve organizational productivity. The International Society for Performance Improvement (ISPI), a professional organization that promotes the use of HPT, presents it as a systematic process that starts with performance analysis, continues with cause analysis, intervention selection and implementation, and ends with evaluation, all phases of which are also interconnected.

The HPT model developed by Van Tiem, Moseley, and Dessinger (2004), seen in Figure 1 (p. 13), includes formative evaluation as an essential conclusion to each phase that validates the findings of that phase and guides the organization in its approach to the next phase. The model suggests summative evaluations to gauge immediate reactions to interventions and a confirmative (outcome) evaluation to examine how the interventions impacted job performance and met organizational goals.

Kirkpatrick's Four Levels of Training Evaluation

In 1959, the Journal for the American Society of Training Directors published four articles, collectively entitled "Techniques for Evaluating Training Programs,"

written by Donald Kirkpatrick, now an emeritus professor of Management at the University of Wisconsin (Kirkpatrick, 1996). Each of the articles, based on his doctoral dissertation about evaluating training for supervisors, focused on one “step” of evaluation.

The steps were widely adopted within the training community (Newstrom, 1978), as they formed a simple yet flexible way of identifying different criteria for evaluation. Although Kirkpatrick originally referred to the criteria as steps, at some point the community re-labeled them as levels (Kirkpatrick, 1996).

Level 1 is Reaction, defined by Kirkpatrick (1996) as how well trainees like a particular training program. Level 1 measures the learners’ immediate response to the training and may encompass anything from how relevant the course content was to their job duties to how comfortable were the classroom chairs. Kirkpatrick recommended that Level 1 evaluations be performed through written, anonymous comment sheets designed to collect quantifiable data relevant to the learners’ satisfaction with the training program. It is easy for training professionals to conduct Level 1 evaluations, as they have access to the learners at the time of training and have control over the collection of Level 1 data.

Level 2 is Learning, which tests the knowledge or skills acquired during the training session. Kirkpatrick and L’Allier (2004) noted that Level 2 evaluations may require a pre-assessment test as well as a post-training test in order to accurately measure how much knowledge the learners have gained through the training session. Like Level 1, Level 2 data collection can be controlled by the training professionals.

Level 3 is Behavior, which measures the transfer of training, or how effectively the learners are using their new knowledge on the job. It is one thing to recall the

concepts learned during training immediately after the session ends, but as Tyler (1942) emphasized, learning is intended to change behaviors. If the workers do not use their newly-acquired knowledge on the job, the training has a poor return on the resources invested because it has not accomplished its primary purpose. Evaluating worker behaviors is considerably more complex than gathering reactions and assessing immediate learning. Training professionals seeking to conduct such evaluations must collect data from multiple sources to generate a thorough view of the post-training behaviors. Some data may be quantitative and objective, like production numbers pre- and post-training. Other data, like performance reviews from supervisors, are subjective in nature; Kirkpatrick and L'Allier (2004) recommended that training professionals develop a methodology for observing and quantifying the behavior in order to make it measureable and subject to analysis in consistent and efficient ways.

Level 4 is Results, which is a measure of to what extent the training program contributed to meeting the organization's objectives for the program. A simple goal that is not greatly affected by external variables can be evaluated on its results with relative ease; Kirkpatrick (1996) mentioned a typing class as such an example, wherein results would be measured by comparing current typing speed and accuracy to the speed and accuracy measured prior to the class. For more complex training content and more complex organizational goals, Level 4 evaluation methods should be developed along with the training program to ensure that the training has a clear set of measurable objectives.

One drawback to the reliance on Kirkpatrick's Four Levels as a guide to training evaluation is that the levels simply define four different criteria for evaluation.

Kirkpatrick did not create or recommend a specific methodology for conducting evaluations for any of the four criteria, and stated that his Four Levels are best regarded as a categorization scheme (Kirkpatrick & L'Allier, 2004).

Another issue with relying on Kirkpatrick's Four Levels as a developed model is the lack of a systemic approach to evaluating both behavior and results. If one accepts the assumption of a direct progression through the levels, one will assume training to be the direct cause of behavior and results. However, training itself only directly affects one possible contributor to performance gaps (for example, the Knowledge component as included in Gilbert's (2007) Behavior Engineering Model (BEM)). A change or stagnation in behavior and results may be the result of a variety of environmental and individual factors. Human Performance Technology (HPT) requires a systemic element to evaluation that takes a holistic view of performance. The HPT model (Van Tiem, Moseley, & Dessinger, 2004) shown in Figure 1 (p. 13) illustrates how instructional support is just one component to performance improvement. If training professionals only focus on a change to knowledge as the solution to performance issues, they will not have an accurate picture of either the effectiveness or relevance of their training product. Also, evaluation of a training program should be designed with this holistic view in order to understand various factors in the system that affect the effectiveness of the training program on performance outcomes.

Part 2: Evaluation of Training Programs

The Importance of Instructional Evaluation

When Tyler (1949) defined his views on curriculum and instruction in American schools, he identified four key questions that were essential when developing any instructional plan:

1. What educational purposes should the school seek to attain?
2. What educational experiences can be provided that are likely to attain these purposes?
3. How can these educational experiences be effectively organized?
4. How can we determine whether these purposes are being attained?

The fourth question focused on educational evaluation. Tyler's (1942) concept of effective learning was a process that achieved not merely the delivery of information, but a permanent change in the learner's behavior patterns. Simple knowledge tests were inadequate as evaluation as they could not demonstrate that the instruction actually affected learner behaviors and thus could not demonstrate the actual effectiveness of the instruction in achieving its goals:

Since the program seeks to bring about certain changes in the behavior of students, and since these are the fundamental educational objectives, then it follows that an evaluation of the educational program is a process for finding out to what degree these changes in the students are actually taking place. (p. 496)

Why is it important for training departments to conduct evaluations of their training programs beyond simple knowledge tests? What Tyler believed was essential for academic evaluation could be equally well applied to workforce training evaluation. If the purpose of workforce training is to teach workers new behaviors to use on the job in

order to achieve a business outcome, the purpose of a training evaluation should be to examine if workers are using the new behaviors on the job and if by doing so they are achieving the desired business outcome.

Gauging the Effectiveness of Training

The primary purpose of a training evaluation is, quite logically, to evaluate the training program. Scriven (1991) defined evaluation as “the process of determining the merit, worth, or value of something” (p. 139). Stufflebeam and Shinkfield (2007) explained that evaluations confirm worth, value, opportunities for improvement, accreditation, accountability, and ultimately whether a program should be kept or discontinued.

It would seem logical that an organization that takes the time to evaluate its training programs would try to determine if the workers actually learned and used the desired knowledge, and then use this information to make any necessary improvements or other changes to the training program. When ASTD (2009b) surveyed its membership on their use of training evaluations, 87.7% of the respondents believed their organizations should use evaluation results to determine if the workers learned what they were supposed to, and 91.3% believed that their organizations should use evaluation results to improve training programs. However, when asked what their organizations actually did, only 39.1% of the organizations examined what workers learned, and only 52.9% of the organizations used the results to improve their training programs.

This discrepancy between what should be done with evaluation results and what actually is done with them could be explained by what kind of data organizations collect for their evaluations. Are they measuring the use of learned behaviors on the job, or are

their “evaluations” no more than quick learner surveys on how enjoyable the class was? If the latter is true, the organizations would lack any useful data which would help them determine the merit, worth, or value of the training programs in relation to the objectives of the training. ASTD (2009b) found that 75.0% of its survey respondents believed that behavior-based training evaluations had a high or very high value to their organizations, yet only 52.5% reported that their organizations actually conducted any behavior-based training evaluations. Pulichino’s (2007) survey of training professionals produced an almost identical result, finding that 52.4% of respondents’ organizations had conducted any behavior-based training evaluations; in terms of the value of such evaluations, his respondents reported that 97.3% of their organizations valued behavior-based training evaluations as a way to measure training effectiveness and 94.6% as a way to measure a change in job performance.

Making the Business Case for Training

Why should an organization spend money on training its workers? ASTD’s 2010 State of the Industry Report stated that U.S organizations spent nearly \$126 billion on workforce training in 2009, with over \$78 billion spent on internal training. They estimated that an average of 62.5% of those internal training expenditures went for training personnel, administration, and development costs. ASTD noted that organizations had increased their expenditures on outsourced training. This does not mean that internal training groups were performing less training, however. Expertus Inc. and Training Industry Inc. (2008) surveyed training managers about organizational pressures on their departmental budgets; 54% reported significant or intense pressure

from upper management to cut training costs, and 60% reported that they were expected to expand training programs without any additional department funding.

Vance (2011) advised that a training department should function like a business unit within the organization, with a clear mission, a compelling business plan, and a strategy for measuring accomplishments. By doing so, the training department would both be perceived as a strategic business partner and maximize the return on learning investment. Berk (2005) suggested that training leaders focus on collecting data that would show the tangible business impact of training, including measurements of improvements and analysis of the actual effect of training on outcomes. While such metrics would promote the importance of the training department to the organization's leaders, the Expertus Inc. and Training Industry Inc. (2008) survey found that only 20% of the surveyed training managers conducted evaluations of training which reflected its impact on productivity, expenditures, and other aspects of importance to upper management.

Kirkpatrick and L'Allier (2004) summed up the value of demonstrating the alignment between training results and business needs as follows:

The creation of well-constructed evaluations can be useful in generating support for budget expenditures, continuing and modifying existing programs, and measuring the impact specific training programs have on the organization's business goals. (p. 30)

Frequency of the Levels

Two recent studies surveyed training professionals to determine how often they conduct Level 3 (behavior-based) and Level 4 (results-based) training evaluations, how

their organizations valued the evaluation data, and what their reasons were for conducting such evaluations.

In 2006, Pulichino (2007) surveyed 446 members of the eLearning Guild, an organization of training professionals focused on the development and implementation of the electronic delivery of instruction. He asked about their organizations' usage of training evaluation types as defined by Kirkpatrick's levels. See Table 4 for his findings.

Table 4. Pulichino 2007 Survey Results on Usage of Kirkpatrick Levels.

Kirkpatrick Level	Frequency of Usage				
	Never	Rarely	Sometimes	Frequently	Always
1: Reaction	3.1	5.2	7.2	23.3	61.2
2: Learning	4.0	11.0	28.0	37.0	19.1
3: Behavior	13.0	34.6	32.5	14.1	5.8
4: Results	33.0	40.5	12.8	9.4	4.3

If the data for "Sometimes," "Frequently," and "Always" are combined to represent an acceptable degree of evaluation usage, then 91.7% of the respondents' organizations conducted Level 1 evaluations, and 84.1% conducted Level 2 evaluations. This figure drops to 52.4% for Level 3 evaluations and 26.5% for Level 4 evaluations.

Three years later, ASTD (2009b) surveyed 704 of its members, all professionals in training and related fields. Respondents were asked which of the Kirkpatrick Levels of evaluation were used in their organization to any extent; the findings are in Table 5.

Table 5. ASTD 2009 Survey Results on Usage of Kirkpatrick Levels.

Kirkpatrick Level	Frequency of Usage
Reactions of participants (Level 1)	91.6%
Evaluation of learning (Level 2)	80.8%
Evaluation of behavior (Level 3)	54.6%
Evaluation of results (Level 4)	36.9%

Pulichino (2007) asked respondents how much value their organizations perceived they received from Level 3 and Level 4 evaluations. Their responses are in Table 6.

Table 6. Pulichino 2007 Survey Results on Value of Evaluation Data.

Kirkpatrick Level	Application of Evaluation Data	Not At All Valuable or Not Very Valuable	Fairly Valuable, Very Valuable, and Highly Valuable
Level 3	Effectiveness of training programs	2.7%	97.3%
	Desired change in job performance	5.4%	94.6%
Level 4	Effectiveness of training programs	3.0%	97.0%
	Desired business or organizational results	3.0%	97.0%

ASTD's (2009b) survey asked respondents how much they thought their organizations valued the results of training evaluations. Their responses are in Table 7.

Table 7. ASTD 2009 Survey Results of Value of Evaluation Data.

Kirkpatrick Level	No Value	A Little Value	Some Value	High Value	Very High Value
Reactions of participants (Level 1)	1.2%	14.2%	48.6%	22.8%	13.1%
Evaluation of learning (Level 2)	0.5%	4.2%	40.4%	38.1%	16.8%
Evaluation of behavior (Level 3)	1.3%	2.8%	21.0%	46.4%	28.6%
Evaluation of results (Level 4)	8.6%	6.0%	25.9%	26.7%	32.7%

If the data for "High Value" and "Very High Value" are combined to represent a strong degree of value to the organization for each Kirkpatrick level, then 35.9% strongly valued Level 1 evaluations, 54.9% strongly valued Level 2 evaluations, 75% strongly

valued the results of Level 3 evaluations, and 59.4% strongly valued the results of Level 4 evaluations.

In any event, all of these studies demonstrated a large gap between the level of value placed on Level 3 and Level 4 training evaluations by organizations and the actual performance of these evaluations.

Barriers to Evaluation at Level 3 and Level 4

If the results of Level 3 and Level 4 evaluations are so highly valued by organizations, why are these evaluations conducted so infrequently?

Moller and Mallin (1996) gathered qualitative data from their survey of instructional designers, asking them to discuss the presence of organizational barriers that hindered their efforts to conduct evaluations at any level. 88% of the respondents stated that such barriers existed in their organizations, with the dominant barrier being related to a lack of time and other resources to conduct evaluations. Moller and Mallin (1996) also documented responses related to organizational cultures resistant to evaluation, lack of access to the data required to conduct evaluations. They reported that only a small number of respondents identified a lack of ability or methodology as a barrier. However, their survey included questions about the methodology in use to conduct evaluations at each level, and they found that respondents frequently identified methods which were inadequate or inappropriate for the evaluation level. For example, post-training interviews with instructors was considered by a respondent to be accurate data collection for a Level 3 evaluation. Moller and Mallin (1996) noted this seemed to stem from a lack of knowledge about evaluations, speculating that respondents do not truly understand what measurements indicate changes in behavior or how to take those measurements.

Pulichino (2007) asked survey respondents to rate the importance of possible reasons why their organizations do not conduct Level 3 and level 4 training evaluations. Table 8 represents the choices presented to the survey respondents as barriers to Level 3 evaluations and the percentage who rated each choice as fairly important, very important, or highly important. Table 9 represents the same information for Level 4 evaluations.

Table 8. Pulichino 2007 Survey Results on Perceived Barriers to Level 3 Evaluation.

Reason	Rated as important
Difficulty accessing the data required for Level 3 evaluations	86.2
No management support for Level 3 evaluations	85.1
Level 3 evaluations too time-consuming	83.3
Level 3 evaluations not considered an important priority for the training department	69.6
Level 3 evaluations too costly	67.2
We do not have the required expertise for Level 3 evaluations	52.2
Level 1 and/or Level 2 evaluations are sufficient for determining the effectiveness of training programs	32.7

Table 9. Pulichino 2007 Survey Results on Perceived Barriers to Level 4 Evaluation.

Reason	Rated as important
Difficulty accessing the data required for Level 4 evaluations	90.5
Level 4 evaluations too time-consuming	85.1
No management support for Level 4 evaluations	79.5
Level 4 evaluations too costly	75.6
Level 4 evaluations not considered an important priority for the training department	75.4
We do not have the required expertise for Level 4 evaluations	64.7
Level 1 and/or Level 2 evaluations are sufficient for determining the effectiveness of training programs	41.5

For both levels of evaluation, survey respondents identified lack of access to needed data to be the most important barrier to conducting evaluations, with the lack of time and management support as the next most important. Over half of the respondents

felt they lacked the necessary expertise to evaluate behavior, and nearly two-thirds lacked the expertise to evaluate results.

Pulichino (2007) made particular note of the respondents' answers about the priority placed on training evaluations by their training departments:

Although not one of the most important reasons cited by participants, the fact that Level 3 and 4 evaluations are not considered important or urgent by training professionals suggests that a "priority divide" may exist between the training department and the business units they serve. Perhaps too many training professionals are satisfied with measuring only what they can directly control, that is, student reaction and learning, and are content to leave the measurement of behavior and results to some other group in the organization. (p. 120-121)

ASTD (2009b) asked its survey respondents to what extent various factors negatively impacted their organizations' ability to conduct Level 3 and Level 4 evaluations. The factors and the percentage of respondents who rated the impact of each as high or very high are represented in Table 10.

Table 10. ASTD 2009 Survey Results on Perceived Barriers to Level 3 and Level 4 Evaluations.

Factor	High or Very High impact
Too difficult to isolate training's impact on results versus other factors' influence	51.7
Our LMS does not have a useful evaluation function	40.8
Evaluation data is not standardized enough to compare well across functions	38.0
It costs too much to conduct higher-level evaluations	32.2
Leaders generally don't care about evaluation data	24.1
Evaluation data is too difficult to interpret for most	18.9
Evaluations are not seen as credible	14.5

The ASTD (2009b) report speculated that the factor that respondents identified as the greatest barrier, the perception of difficulty in proving the actual impact of training on

results, may be a self-fulfilling prophecy. ASTD's analysis showed a negative correlation between the perception that proving the impact of training is too difficult to accomplish and the overall effectiveness of an organization's evaluation efforts, and concluded that organizations that avoid evaluation due to perceived difficulty may be sabotaging their training efforts through inaction.

Part 3: The Importance of Evaluation to Human Performance Technology

ISPI (2012) developed the Ten Standards of Performance Technology as a guideline for the systematic practice of HPT. Those seeking the professional credential of Certified Performance Technologist (CPT) are required to document how they have demonstrated each of the ten standards in their practice.

The behavior-based or results-based evaluation of a training-based intervention, when completed and presented to the stakeholders, would fulfill five of the ten ISPI standards as follows:

Focus on results: A behavior-based or results-based training evaluation should be designed to demonstrate to what degree the training intervention achieved the goals for which it was designed.

Systemic approach: A behavior-based or results-based evaluation will use evidence from multiple sources to better understand if the training intervention achieved its goals and what individual or environmental factors may have had an impact on the training outcome.

Add value to your work: By conducting behavior-based or results-based training evaluations, the HPT practitioner helps an organization to determine if it is using its resources effectively when developing and implementing training interventions.

Collaborating with clients: The HPT practitioner cannot conduct behavior-based or results-based evaluations in isolation from the client. He or she must earn support from the client organization for the evaluation process by educating it on the value of evaluation towards achieving business goals and then work in partnership with the organization at all levels during the evaluation process.

Systematic evaluation: The HPT practitioner must approach the evaluation process with a clear system that includes the design, development, and implementation of an effective evaluation method, followed by communication of the results to the organization.

Evaluation is an integral part of the HPT process; without it, the HPT practitioner has little concrete evidence to prove the value of an intervention. Training programs are widely used in organizations, which spend a great deal of money and other resources on these programs. Through systematic evaluation of the behavior and results outcomes of training programs, the HPT professional brings value to the organization by assessing how effectively those resources have been used.

CHAPTER THREE: METHODOLOGY

The purpose of this research was to examine how often training professionals evaluate the success of their training programs in producing the on-the-job behaviors and business outcomes that the programs were intended to create. The research also looked at the individual and organizational factors that affected the ability of training professionals to conduct such evaluations and how those factors helped or limited the successful performance of the evaluations. This research was approved by Boise State University's Institutional Review Board (see Appendix B).

This research was designed to answer two primary questions and associated sub-questions:

Research Question 1: With what frequency do training professionals conduct Level 3 and/or Level 4 evaluations for their organizations?

Sub-question 1a: Who are the stakeholders for these evaluations?

Sub-question 1b: For what reasons do training professionals conduct or attempt to conduct the evaluations?

Research Question 2: What factors act as facilitators or barriers to conducting Level 3 and Level 4 training evaluations?

Sub-question 2a: For Success Cases, what are the facilitating factors and the barriers, and how did they impact the performance of evaluations?

Sub-question 2b: For Non-Success Cases, what are the facilitating factors and the barriers, and how did they impact the attempts to perform evaluations?

Sub-question 2c: For Non-Evaluators, what are the facilitating factors and the barriers, and why were evaluations not attempted?

Population and Sampling

This study was focused on individuals whose primary job duties include internal training functions for the organizations with which they are employed. Additional qualitative data was collected from individuals who are involved in training functions but not as a primary job duty or not for internal clients.

The U.S. Department of Labor's Bureau of Labor Statistics (2012) estimated that in 2011, almost 233,000 individuals worked as training and development managers and specialists in the United States. For the purposes of this study, the population was initially limited to those training professionals who are members of online professional forums on LinkedIn.com. With the permission of forum leaders, an invitation to participate in the research project (included in Appendix C) was posted to each of the following LinkedIn groups:

- ASTD National (approximately 45,000 members)
- Kirkpatrick Partners (approximately 2,900 members)
- Instructional Design & E-Learning Professionals' Group (approximately 22,000 members)
- E-Learning Guild (approximately 23,000 members)
- Instructional & Performance Technology Network – Boise State University (243 members)

Four study participants contacted the researcher directly to ask permission to re-post the study solicitation message to their own professional online networks; this permission was granted. The survey instrument did not track respondents by source, so it is unknown which outlets were the most successful for soliciting participants.

Instruments

Questions for the survey instrument were based primarily on the factors identified by Pulichino (2007) and ASTD (2009b) in their research on training evaluation usage and barriers. The questions were refined through consultation with the members of the thesis committee, who contributed their experience in research and the HPT field to increase the relevancy and clarity of the questions and available answers.

An examination of prior research on the frequency of training evaluations and the factors that impact evaluations indicated some consistency in the factors considered as important. Some of the common factors were environmental, in that they were the result of organizational structure, culture, and actions. Other common factors were individual and were the result of knowledge, experience, and perception.

The first page of the survey was a Survey Cover Letter (Appendix D) which served as an informed consent form, as required by Boise State University's Institutional Review Board (IRB) to protect human research subjects. Respondents who declined to continue with the survey exited before providing any data.

The full survey is included as Appendix E. The first question in the survey instrument eliminated respondents who did not meet the primary target population of

individuals whose primary duties involved internal training and development functions. The eliminated respondents were invited to answer two free response questions; this qualitative data was included in the final analysis.

After a set of questions about frequency of evaluations and perceptions of their importance, the survey respondents answered a single question about the completion of Level 3 evaluations, which identified them as Success Cases (those who completed at least one Level 3 evaluation and delivered the results to a stakeholder), Non-Success Cases (those who attempted at least one Level 3 evaluation but were unable to complete it), and Non-Evaluators (those who did not attempt any Level 3 evaluations). The Success Cases and Non-Success Cases received similar sets of questions about the factors that affected their evaluations; the only difference was the phrasing of the question to reflect attempts versus completions. The Non-Evaluators received a different set of questions to explore the reasons why no Level 3 evaluations were attempted.

After the questions about Level 3 evaluations, the respondents were asked the same questions about their completions, attempts, or non-attempts of Level 4 evaluations.

At the end of the survey, respondents were invited to volunteer for post-survey interviews. These respondents were categorized according to their success at each evaluation level and then each was contacted by e-mail with a request to schedule a 15-30 minute interview. All interviewees received a letter of Written Informed Consent by e-mail prior to the interview and listened to a script for Verbal Informed Consent at the beginning of the interview; the informed consent documents are included in Appendix F. The interview instrument was a semi-structured script that prompted the informants to speak about their success or lack of success in conducting training evaluations at Level 3

and Level 4. Different scripts were created for Success Case, Non-Success Case, and Non-Evaluator informants. All of the interview scripts are included in Appendix G.

Overall Methodology Framework

Brinkerhoff (2003) developed the Success Case Method (SCM) as a way to analyze the effectiveness of a program, using both quantitative and qualitative data, by looking for successful elements and determining what conditions led to that success. The ASTD (2009b) study on training evaluation singled out the SCM as an evaluation methodology popular with organizations; 47.7% of respondents reported that their organizations conducted interviews with successful learners as a way to evaluate training results.

This research project was structured around Brinkerhoff's (2003) SCM for evaluation. According to Brinkerhoff, the SCM can be used to answer any or all of the following four questions about a given program:

- What is really happening?
- What results, if any, is the program helping to produce?
- What is the value of the results?
- How could the initiative be improved? (pp. 6-13)

The SCM is usually used to evaluate the outcome of specific programs or initiatives, including training programs. Although this research study did not evaluate a specific planned program, the SCM methodology was flexible enough to have been adapted to the needs of the study.

The SCM consists of five steps:

1. Focus and plan the SCM evaluation
2. Create an impact model
3. Survey all program recipients to identify success and non-success cases
4. Interview success and nonsuccess cases and document their stories
5. Communicate findings, conclusions, and recommendations

Steps 3 and 4 allow for the collection of both quantitative data from the survey and qualitative data from the interviews. Previous research studies on the topic of evaluation usage by training professionals had collected only quantitative survey data on the frequency of training evaluations and the factors that affected the evaluations; by following the SCM structure, this study allowed for qualitative data that looked at the stories behind the numbers of those who were successful at conducting evaluations and those who were not.

A unique aspect of the SCM is its use of a survey instrument to identify extreme cases with whom to conduct those narrative interviews. Extreme case selection focuses on the individuals who are farthest from the mean of the sample distribution and thus represent the most unusual cases. In an SCM, the extreme cases represent the most successful and least successful program recipients and are not meant to be representative of the "typical" program recipient. Seawright and Gerring (2008) note that this method of selection is only appropriate in an exploratory study involving open-ended questioning, before a specific hypothesis has been formed.

However, according to Brinkerhoff (2003), a key advantage of the SCM is its use of qualitative data gathered through narrative interviewing, or storytelling. Interview

questions developed for an SCM evaluation allow for semi-structured, open-ended questioning so the program recipients are free to relate their experiences about specific aspects of the program. A similar narrative-based exploration of human performance is Flanagan's (1954) Critical Incident Technique, which asks informants to identify and discuss extreme behaviors that they have witnessed in a specified context.

Implementing the Success Case Method

Step 1: Focus and plan the SCM evaluation

The first step of a SCM evaluation includes defining the purpose of the study, why and to whom it is important, and the overall strategy for collecting the necessary data.

The purpose of this study, as defined by its primary research questions, was to examine the frequency with which training professionals conduct Level 3 and Level 4 evaluations, and to examine the factors that impact this frequency.

As this study does not look at a planned program or a specific organization, there is no formal stakeholder who requested the evaluation or would directly benefit from receiving the results. In this case, the field of workforce training and development was identified as the stakeholder, as the study results were meant to contribute to the base of knowledge about the field.

The study population was not centralized in location or organization, and the timeframe of the study itself was limited by the academic schedule, so the overall strategy relied on rapid communication methods independent of such centralization. The survey participants were solicited and then surveyed online, with the selection of success

and non-success cases based on the analysis of survey responses. Interviews were done by telephone or Voice Over Internet Protocol (VOIP).

Step 2: Create an impact model

Brinkerhoff (2003) defined an impact model as “a carefully articulated and concise forecast of what the most successful performance might be expected to look like” (p. 76). This research study defined a successful performer as a training professional who successfully conducted training evaluations that measured the impact of the instruction on individual employee behaviors and/or achievement of organizational goals and then communicated the evaluation results to the organization for its benefit.

Specific capabilities needed for the successful completion of training evaluations had not been identified prior to the study, so potential capabilities of the population were structured around the components of Gilbert’s (2007) Behavior Engineering Model (BEM), with successful performance defined as the completion of behavior- and results-based evaluations of training interventions (For more information about Gilbert’s BEM, see Appendix A). This study’s impact model is shown in Table 11.

Table 11. Structure for Impact Model Based on Gilbert’s BEM Components.

	Capabilities	Critical Actions	Key Results	Outcomes
Organizational Level	Data	<p>Organization clearly defines its expectations for the evaluation of training, and defines the role of training department in the evaluation process.</p> <p>Organization facilitates the collection of data required for evaluation, and encourages feedback to flow between the training department, stakeholders, and downstream impactees.</p>	<p>Organization receives evidence of the effectiveness of its training interventions for workforce improvement.</p>	<p>Organization improves organizational success by making use of evaluation results and transforms to a learning organization.</p>

	Instruments	Organization allocates sufficient personnel to the training department to allow adequate time for the evaluation process without negatively impacting other departmental priorities. Organization acquires or develops tools for collection and analysis of evaluation data.		
	Incentives	Organization makes visible how it values evaluation as an important component of workforce improvement.		
Individual Level	Knowledge	Training professional uses the terminology, concepts, and models applicable to training evaluation. Training professional communicates effectively with organization at all levels.	Training professional successfully conducts evaluation of training, generates evidence of contribution towards organizational goals involving workforce improvement, and then communicates results to the appropriate groups.	Training professional or training department functions as a strategic business partner and is actively involved in improving organizational success.
	Capacity	Training professional possess the cognitive capability to make data-based evaluative judgments.		
	Motives	Training professional wants to perform the required jobs and has motives aligned with the work and work environment.		

Step 3: Survey all program recipients to identify success and non-success cases

Potential survey participants were solicited online; see Population and Sampling Plan for details. The solicitation message included a link to an online survey hosted by Qualtrics.

Survey respondents were screened for eligibility at the start of the survey. Only those who self-identified as individuals whose primary job function was related to training and development for internal organizational clients were allowed to take the full survey. Those who worked primarily with external clients, or whose primary job functions did not relate to training and development, were given the opportunity to answer two free-response questions.

The full survey instrument created for this research was divided into three parts (see Appendix E). The first part of the survey instrument collected demographic information and asked questions related to respondents' perceptions of training evaluation at Level 3 and Level 4 and the frequency with which their organizations conducted training evaluations at all four levels.

At the end of the first part of the survey, respondents answered a single question about their success in completing of Level 3 evaluations; their responses to this question identified them as Level 3 Success Cases (those who completed at least one Level 3 evaluation and delivered the results to a stakeholder), Level 3 Non-Success Cases (those who attempted at least one Level 3 evaluation but was unable to complete it), and Level 3 Non-Evaluators (those who did not attempt any Level 3 evaluations). For the second part of the survey, each of these groups answered a different set of questions regarding the factors that impacted their ability to conduct Level 3 evaluations.

After the questions about Level 3 evaluations, the respondents were asked a single question about the completion of Level 4 evaluations to identify them as Level 4 Success Cases, Level 4 Non-Success Cases, and Level 4 Non-Evaluators. For the third part of the survey, each of these groups answered a different set of questions regarding the factors that impacted their ability to conduct Level 4 evaluations.

At the end of the survey, respondents were invited to volunteer for post-survey interviews. Only those who volunteered were considered when selecting Success Cases, Non-Success Cases, or Non-Evaluators to be interviewed.

Sixty-eight respondents completed the first part of the survey about demographics and perceptions. An additional 11 respondents did not meet the screening criteria for the

full survey but answered the two free-response questions. Sixty-four respondents completed the second part of the survey about success with Level 3. Sixty-one respondents completed the third part of the survey about their success with Level 4.

In this step, the methodology used for this study diverged from the SCM. Brinkerhoff (2003) stated that interviewees should be selected based on the degree of self-reported success as determined by survey responses. However, the survey instrument for this study did not attempt to measure the degree of success. With no definite common factor among respondents except professional field, it would have been very difficult to establish a rating system that would be universal to such a variety of backgrounds, job functions, and organizational structures; what one respondent considered a completed and successful evaluation might be regarded by another respondent as incomplete and unsuccessful.

A decision was made to maximize the amount of qualitative interview data available, and thus all survey respondents who volunteered for interviews were contacted with requests for interviews.

Step 4: Interview success and nonsuccess cases and document their stories

Thirty-two survey respondents volunteered to be contacted for interviews. The prospective interviewees were contacted by e-mail to schedule a telephone or Voice Over Internet Protocol (VOIP) interview; 22 individuals actually participated in the interview step of this project.

The prospective interviewees were e-mailed a letter of Written Informed Consent, as required by the BSU IRB. Those individuals who consented to an interview listened to

a Verbal Informed Consent script at the beginning of the interview; all interviewees gave their verbal consent and were interviewed. Informants were notified in the Written Informed Consent and Verbal Informed Consent that the interview would be digitally audio recorded unless they denied permission; all interviewees allowed audio recording for the full interviews.

Brinkerhoff (2003) recommended that SCM interviews be 30 to 45 minutes long. However, the interviewees for this study had received no direct benefit from participation in a program or initiative, nor did they have professional or personal obligations to participate. To increase the likelihood of cooperation, potential informants were told that interviews would last only 15 to 30 minutes and would be scheduled and conducted to avoid inconveniencing participants more than necessary. Informants who were willing to talk for a longer period of time were encouraged to do so; actual interview times ranged from 12 to 40 minutes, with the majority lasting 20-25 minutes.

The interviews were designed as semi-structured, which allowed for open-ended responses to scripted and follow-up questions. Separate interview scripts were written for Success Cases, Non-Success Cases, and Non-Evaluators; scripts were chosen for each subject based on his or her self-categorization questions from the survey. In each script, the questions asked the informants to describe the factors that contributed to the completion, non-completion, or non-attempt to conduct evaluations of training at both Level 3 and Level 4.

Step 5: Communicate findings, conclusions, and recommendations

The quantitative data from the survey was compiled in Excel. The quantitative findings are presented in Chapter 4.

The qualitative data for this research was collected through two methods. The online survey included two free-response questions:

1. What do you think is the most critical facilitating factor that organizations should maintain in order to successfully conduct Level 3 and Level 4 evaluations, and why do you think so?
2. What do you think is the most obstructing barrier that organizations should reduce or remove in order to successfully conduct Level 3 and Level 4 evaluations, and why do you think so?

All survey respondents, including those ineligible to take the full survey, were given the opportunity to answer these questions.

The rest of the qualitative data was collected through interviews conducted over the telephone or via a VOIP (Skype) connection.

Responses from the survey and interviews were first sorted using Brinkerhoff's (2003) "bucket" concept, using the following broad domains:

- Barriers to evaluation
- Facilitating factors for evaluation

The contents of each "bucket" were analyzed using deductive coding, which LeCompte and Schensul (1999) referred to as "analysis from the top down." The coding categories followed Gilbert's (2007) Behavior Engineering Model, with each response assigned to one of the six categories below:

- Organizational Data
- Organizational Instruments

- Organizational Incentives
- Individual Knowledge
- Individual Capacity
- Individual Motives

After the responses had been coded into these categories, the analysis focused on the three categories to which almost all responses from both “buckets” had been assigned: Organizational Data, Organizational Instruments, and Individual Knowledge. The qualitative findings are presented in Chapter 4.

CHAPTER FOUR: DATA ANALYSIS AND FINDINGS

The findings for this research are divided into two parts:

Part 1: Quantitative and qualitative data collected through the online survey

Part 2: Qualitative data collected through post-survey interviews with survey participants

Part 1: Survey Data

The survey instrument used in this research project served two purposes. One was to collect quantitative data on the frequency with which training evaluations were performed and the factors that impacted that frequency. The second was to categorize respondents by their success, non-success, or non-attempt to conduct training evaluations at Level 3 and Level 4.

Demographics

Sixty-eight individuals completed the first part of the survey instrument, which collected demographic information and questioned the respondents on their usage and perceptions of Level 3 and Level 4 evaluations. The demographic information is shown in Table 12.

Table 12. Demographic Data for Survey Respondents.

Gender	
Male	33.82%
Female	66.18%

Age Range	
18 - 29	4.41%
30 - 39	25.00%
40 - 49	32.35%
50 -59	30.88%
60 or older	7.35%
Job Function	
Instructional Designer	25.00%
Trainer	19.12%
Human resources professional in training & development	7.35%
Human resources professional (other)	1.47%
Organizational development specialist	1.47%
Manager in training & development, organizational development, or human resources	36.76%
Manager (other)	0.00%
Other	8.82%
Years of Experience	
Less than 2 years	5.88%
2 to 5 years	8.82%
6 to 15 years	50.00%
More than 16 years	35.29%
Number of Employees in Organization	
Fewer than 100	2.94%
100 - 500	27.94%
501 - 1500	20.59%
More than 1500	44.12%
Not sure	4.41%
Highest Degree Earned	
High school diploma	2.94%
Associate's degree	2.94%
Bachelor's degree	35.29%
Master's degree	48.53%
Doctorate	10.29%
Highest Instructional Design Degree Earned	
Bachelor's degree	8.82%
Master's degree	39.71%
Doctorate	8.82%
Certificate	19.12%
none	23.53%
Evaluation-specific Course Completed	
Credit-based university course	27.94%

Non-credit seminar or workshop	19.12%
Both credit and non-credit courses	33.82%
No evaluation-specific course completed	19.12%

Respondents were also asked the type of organization for which they worked; the largest concentration was in the health care industry, which employed 13.24% of the respondents.

The remaining questions in the first part of the survey explored the Research Questions and associated sub-questions for this thesis research:

Research Question 1: With what frequency do training professionals conduct Level 3 and/or Level 4 evaluations for their organizations?

Sub-question 1a: Who are the stakeholders for training evaluations?

Sub-question 1b: For what reasons do training professionals conduct or attempt to conduct training evaluations?

Research Question 2: What factors act as facilitators or barriers to conducting Level 3 and Level 4 training evaluations?

Sub-question 2a: For Success Cases, what are the facilitating factors and the barriers, and how did they impact the performance of evaluations?

Sub-question 2b: For Non-Success Cases, what are the facilitating factors and the barriers, and how did they impact the attempts to perform evaluations?

Sub-question 2c: For Non-Evaluators, what are the facilitating factors and the barriers, and why were evaluations not attempted?

Frequency of Evaluations

Survey respondents were asked to estimate the percentage of their organizations' training programs that were evaluated at each of the four Kirkpatrick levels; respondents

entered their estimates using a slider to indicate approximate percentages. The respondents' estimates were grouped into categories as Rarely (0% to 20% of training evaluated), Infrequently (21% to 40% of training evaluated), Sometimes (41% to 60%), Often (61% to 80%), and Almost Always (81% to 100%). The results are shown in Table 13.

Table 13. Estimates of % of Training Evaluated, By Level.

	Rarely	Infrequently	Sometimes	Often	Almost Always	<i>n</i>
Level 1	0.00%	11.86%	8.47%	8.47%	71.19%	59
Level 2	3.64%	21.82%	20.00%	25.45%	29.09%	55
Level 3	28.26%	28.26%	17.39%	13.04%	13.04%	46
Level 4	55.26%	26.32%	5.26%	5.26%	7.89%	38

If the responses for “Sometimes,” “Often,” and “Almost Always” are grouped together as an indication of how respondents' organizations are evaluating on a somewhat regular basis, it appears that Level 1 evaluations are conducted 88.13% of the time, Level 2 evaluations 74.54% of the time, Level 3 evaluations 43.47% of the time, and Level 4 evaluations 18.41% of the time.

Respondents to this survey were required to specify a percentage when estimating frequency of evaluation. However, respondents for the other surveys referenced here were allowed to choose their own meaning for terms such as “sometimes” and “often”, so a direct comparison between survey results cannot be made with accuracy. However, the results of this survey reflect the same general trend of the other surveys, in which high frequency levels for Level 1 evaluations fade to a much lower frequency for Level 4 evaluations.

Perceptions About Evaluation

Alliger and Janak (1989) identified three assumptions about the Kirkpatrick levels that they believed were common among researchers and practitioners in the training field. The first assumption was that the levels were presented in ascending order of importance, in that results from Level 1 evaluations are not as valuable as those from Level 2, and Level 2 results are not as valuable as Level 3 ones. The second and third assumptions are closely related and assume a causal link existing between the levels and a positive correlation between the results at each level. The survey asked respondents their opinions on the importance and usefulness of evaluations at each level and if they believed in a positive correlation between results at the different levels.

To discover how respondents viewed the value of the different levels of evaluation, they were asked to rate how sufficient each level was for judging the effectiveness of training programs. The results are shown in Table 14.

Table 14. Perceived Sufficiency of Evaluation Results for Judging Training Effectiveness.

	Very insufficient	Somewhat insufficient	Neither sufficient nor insufficient	Somewhat sufficient	Very sufficient	<i>n</i>
Level 1	16.18%	16.18%	14.71%	42.65%	10.29%	68
Level 2	10.29%	22.06%	7.35%	47.06%	13.24%	68
Level 3	14.29%	15.87%	7.94%	36.51%	25.40%	63
Level 4	31.67%	16.67%	8.33%	16.67%	26.67%	60

By grouping together the responses for “Somewhat sufficient” and “Very sufficient,” the survey shows that 52.94% of respondents believe that Level 1 results

provide usable data for evaluating training programs, 60.30% believe that of Level 2 results, 61.91% of Level 3, and 43.34% of Level 4.

Respondents were also asked how important they believed it was to conduct evaluations at each level. The results are shown in Table 15.

Table 15. Perceived Importance of Conducting Each Level of Evaluation.

	Not at all important	Very unimportant	Neither important nor unimportant	Very important	Extremely important	<i>n</i>
Level 1	2.99%	2.99%	28.36%	40.30%	25.37%	67
Level 2	0.00%	0.00%	4.48%	47.76%	47.76%	67
Level 3	0.00%	0.00%	1.47%	27.94%	70.59%	68
Level 4	0.00%	0.00%	12.12%	34.85%	53.03%	66

By grouping together the responses for “Very important” and “Extremely important,” the data shows that 65.67% believe in the importance of Level 1, 95.52% in the importance of Level 2, 98.53% in the importance of Level 3, and 87.88% in the importance of Level 4.

The data for the respondents’ opinions of the sufficiency of each level was compared to the respondents’ opinions of the importance of each level. To make a more accurate comparison, a listwise deletion was performed to eliminate any datasets that were missing any of the data for these questions. The remaining responses were grouped to calculate the percentage of respondents who answered “Somewhat sufficient” and “Very sufficient” for the perceived sufficiency and “Very important” and “Extremely important” for the perceived importance of each level. The comparison between respondents’ perception of the sufficiency of each level to evaluate training effectiveness and the importance of evaluating at each level is shown in Table 16.

Table 16. Comparison of Perceived Sufficiency and Importance of Level, $n=59$.

	Perceived sufficiency of evaluation data	Perceived importance of level
Level 1	52.54%	66.10%
Level 2	61.02%	94.92%
Level 3	59.32%	98.31%
Level 4	42.37%	88.14%

Do training professionals assume a positive correlation between the four levels?

Respondents were asked if they believed that positive results at Level 1 or Level 2 were strong indicators of positive results at either or both of the other levels, or if they felt there was no correlation. The results are shown in Table 17.

Table 17. Perception of Positive Correlation Between Kirkpatrick Levels, $n=68$.

Positive results for Level 1 or Level 2 indicate:	
Positive results for Level 3	11.76%
Positive results for Level 4	1.47%
Positive results for Level 3 and Level 4	16.18%
No correlation automatically indicated	69.12%

The survey results do not appear to support the belief that training professionals assume either an ascending importance to the levels or a causal link between them. The respondents rated Level 2 and Level 3 as nearly equal in importance and usefulness of results, and considered Level 4 to be both less important to conduct and less useful as data for examining their training programs. An overwhelming majority of respondents did not accept an automatic positive correlation of results between the levels.

Through answering a single survey question, respondents self-selected themselves into Success Cases, Non-Success Cases, and Non-Evaluators categories for Level 3. Each category received a different set of questions, after which respondents self-selected

themselves into those three categories for Level 4. Respondents who self-selected themselves into a category but did not answer any further questions were eliminated from the data. The number of respondents who provided data in each category is shown in Table 18.

Table 18. Number of Respondents by Evaluation Level and Category.

Category	Level 3	Level 4
Success Cases	33	16
Non-Success Cases	10	4
Non-Evaluators	21	41
Total	64	61

Level 3 Stakeholders

The Level 3 Success Cases and Level 3 Non-Success Cases identified one or more stakeholders who requested the completion of Level 3 evaluations. The stakeholders are shown in Table 19.

Table 19. Stakeholders for Level 3 Evaluations.

Stakeholder requesting Level 3 evaluation	Frequency of request from this stakeholder for Success Cases (n=33)	Frequency of request from this stakeholder for Non-Success Cases (n=10)
Manager/supervisor of workers who attended training program	12.12%	10.00%
Higher-level management in the organization	21.21%	30.00%
Training department	75.76%	30.00%
Other	21.21%	70.00%

For both categories of respondents, other stakeholders included those external to the organization, such as regulatory agencies. Respondents also identified higher-level management and themselves as “other.”

Level 3 Reasons for Evaluation

Level 3 Success Cases and Level 3 Non-Success Cases identified one or more reasons for conducting Level 3 evaluations. The results for both categories of respondents are shown in Table 20.

Table 20. Reasons for Evaluation at Level 3.

Reason for conducting or attempting to conduct evaluation of training	L3 Success Cases (n=33)	L3 Non-Success Cases (n=10)	Both L3 Success Cases and L3 Non-Success Cases (n=43)
To assess what parts of the training were being used, and which were not so that changes to training content and methods could be made, if needed	75.76%	60.00%	70.45%
To help decide what parts of the organization were effectively executing important new behavioral changes (e.g., strategy, new technology), and which were not, so that remediation efforts could be implemented where needed	39.39%	30.00%	36.36%
To determine whether and how much managers of trainees were supporting on-job application of learning	42.42%	0%	31.82%
To assess the efficacy of the training with a small or “pilot” group of trainees so that changes could be made to increase impact of later and larger roll-outs	24.24%	10.00%	20.45%
To help determine what factors were impeding or facilitating on-job application of learning	39.39%	50.00%	40.91%
To discover emerging best-practices of on-job application so that these could be communicated to other trainees and managers, and also might be built into the training content	30.30%	10.00%	25.00%
To demonstrate the value of your department’s contributions to the organization	45.45%	60.00%	47.73%
Because it was required by the organization	21.21%	30.00%	22.73%
Because it just seemed like the right thing to	18.18%	10.00%	15.91%

do			
Other	9.09%	30.00%	13.64%

Level 3 Non-Evaluators were asked why they did not attempt to conduct any Level 3 evaluations of their organization's training programs. The results are shown in Table 21.

Table 21. Reasons for Non-Evaluation of Level 3.

Reason for not evaluating	Frequency of response (n = 21)
The department's lack of expertise/experience in using evaluative methods	47.62%
Issues with access to post-training data such as employee surveys or performance measurements	42.86%
Lack of resources available for conducting evaluation, such as time and budget	71.43%
Lack of support from organizational management	61.90%
The low importance placed on evaluation by your department	18.18%
The low importance placed on evaluation by your organization	33.33%
Other	4.76%

Level 3 Factors Potentially Impacting Evaluation

All respondents were given a list of factors that may impact the performance of Level 3 training evaluations and were asked to score how available each factor was to their training department. The question used a 5-point Likert scale, with 1 indicating the factor is very unavailable and 5 indicating it is very available. The ratings for each factor were calculated by category, with the average scores shown in Table 22.

Table 22. Availability of Level 3 Evaluation Factors, By Respondent Category.

Factor	L3 Success Cases (n=33)	L3 Non-Success Cases (n=10)	L3 Non- Evaluators (n=21)
---------------	--	--	--

The department's expertise/experience in using evaluative methods	3.09	3.20	2.48
Access to employees for post-training surveys or interviews	3.15	2.70	2.52
Access to other post-training data, such as performance measurements or employee evaluations	2.64	2.00	2.00
Resources for conducting evaluation, such as time and budget	2.76	2.10	2.05
Support from organizational management for conducting an evaluation	2.85	2.90	2.05
The importance placed on evaluation by your department	3.27	3.11	2.48
The importance placed on evaluation by your organization	2.76	3.00	2.00

Level 4 Stakeholders

The Level 4 Success Cases and Level 4 Non-Success Cases identified one or more stakeholders who requested the completion of Level 4 evaluations. The stakeholders are shown in Table 23.

Table 23. Stakeholders for Level 4 Evaluations.

Stakeholder requesting Level 4 evaluation	Frequency of request from this stakeholder for Success Cases (n=16)	Frequency of request from this stakeholder for Non-Success Cases (n=4)
Manager/supervisor of workers who attended training program	25.00%	25.00%
Higher-level management in the organization	43.75%	0%
Training department	68.75%	50.00%
Other	12.50%	25.00%

Respondents identified themselves and an external agency as “other”.

Level 4 Reasons for Evaluation

Level 4 Success Cases and Level 4 Non-Success Cases identified one or more reasons for conducting Level 4 evaluations. The results for both categories of respondents are shown in Table 24.

Table 24. Reasons for Evaluation at Level 4.

Reason for conducting or attempting to conduct evaluation of training	L4 Success Cases (n=16)	L4 Non-Success Cases (n=4)	Both L4 Success Cases and Non-Success Cases (n=20)
To assess what parts of the training were being used, and which were not so that changes to training content and methods could be made, if needed	56.25%	25.00%	50.00%
To help decide what parts of the organization were effectively executing important new behavioral changes (e.g., strategy, new technology), and which were not, so that remediation efforts could be implemented where needed	50.00%	25.00%	50.00%
To determine whether and how much managers of trainees were supporting on-job application of learning	37.50%	50.00%	40.00%
To assess the efficacy of the training with a small or “pilot” group of trainees so that changes could be made to increase impact of later and larger roll-outs	37.50%	25.00%	35.00%
To help determine what factors were impeding or facilitating on-job application of learning	43.75%	0%	35.00%
To discover emerging best-practices of on-job application so that these could be communicated to other trainees and managers, and also might be built into the training content	43.75%	25.00%	40.00%
To demonstrate the value of your	68.75%	75.00%	70.00%

department's contributions to the organization			
Because it was required by the organization	25.00%	0%	20.00%
Because it just seemed like the right thing to do	18.75%	0%	15.00%
Other	6.25%	0%	5.00%

The single response for “other” would have been properly classified as demonstrating value to the organization.

Level 4 Non-Evaluators were asked why they did not attempt to conduct any Level 4 evaluations of their organization's training programs. The results are shown in Table 25.

Table 25. Reasons for Non-Evaluation at Level 4.

Reason for not evaluating	Frequency of response (n=42)
The department's lack of expertise/experience in using evaluative methods	54.76%
Issues with access to post-training data such as employee surveys or performance measurements	50.00%
Lack of resources available for conducting evaluation, such as time and budget	57.14%
Lack of support from organizational management	38.10%
The low importance placed on evaluation by your department	23.81%
The low importance placed on evaluation by your organization	42.86%
Other	11.90%

Of the five respondents who chose “other,” four identified lack of relevance to the organization as the reason for not conducting a Level 4 evaluation.

Level 4 Factors Potentially Impacting Evaluation

All respondents were given a list of factors that may impact the performance of Level 4 training evaluations, and were asked to score how available each factor was to

their training department. The question used a 5-point Likert scale, with 1 indicating the factor is very unavailable and 5 indicating it is very available. The ratings for each factor were calculated by category, with the average scores shown in Table 26.

Table 26. Availability of Level 4 Evaluation Factors, By Respondent Category.

Factor	L4 Success Cases (n=16)	L4 Non-Success Cases (n=4)	L4 Non-Evaluators (n=42)
The department's expertise/experience in using evaluative methods	3.25	2.50	2.29
Access to employees for post-training surveys or interviews	3.20	2.25	2.49
Access to other post-training data, such as performance measurements or employee evaluations	3.07	2.25	1.90
Resources for conducting evaluation, such as time and budget	2.75	2.25	2.07
Support from organizational management for conducting an evaluation	3.00	2.25	2.10
The importance placed on evaluation by your department	3.31	2.75	2.46
The importance placed on evaluation by your organization	2.88	2.00	2.00

Level 3 and Level 4 Facilitating Factors

All survey respondents, including those who were not eligible to complete the full survey, were asked to name the factors that they perceived as the most critical in helping and hindering training evaluations at Level 3 and Level 4. Their responses were classified using the seven factors that had been included in the survey as having a potential impact

on conducting evaluations. Some of the responses were tagged with more than one classification. The results are shown in Table 27.

Table 27. Facilitating and Obstructing Factors For Both Levels.

Factor	Mentioned as critical facilitating factor (n=45)	Mentioned as critical obstructing barrier (n=44)
The department's expertise/experience in using evaluative methods	15.5%	20.5%
Access to employees for post-training surveys or interviews	6.7%	9.1%
Access to other post-training data, such as performance measurements or employee evaluations	24.4%	4.5%
Resources for conducting evaluation, such as time and budget	8.9%	31.8%
Support from organizational management for conducting an evaluation	44.4%	47.7%
The importance placed on evaluation by your department	11.1%	4.5%
The importance placed on evaluation by your organization	20.0%	29.5%

Support from management for evaluation efforts was the dominant theme to the responses, as it was mentioned far more often as both a facilitating and obstructing factor than other factors that could affect attempts to conduct Level 3 and Level 4 evaluations. At least two respondents mentioned each of the factors listed in the survey as being of critical importance, which may indicate that all of the factors selected for the survey bear some relevance to the real-world conditions that training professionals experience. Of some note are the results for two other factors. Access to post-training data from sources

other than the learners themselves was the second most mentioned facilitating factor (24.4%) but tied for the least mentioned (4.5%) obstructing factor. It may be that an organization that does not provide access to this data also exhibits a clear lack of managerial support which respondents considered to be the primary obstructing factor. Having resources like time and personnel available for evaluation purposes was only mentioned 8.9% of the time as a critical facilitator for evaluations, but 31.8% of the time as a critical obstructing factor. Organizations may support the need for evaluations in theory, but in practice do not or cannot allocate the resources necessary in order to conduct the evaluations.

Part 2: Interview Data

At the end of the online survey, respondents were asked to volunteer to participate in post-survey interviews. Thirty-two survey respondents volunteered, and all were contacted by e-mail to confirm their interest in participating. Twenty-two survey respondents took part in telephone or VOIP interviews conducted between February 14, 2012 and March 5, 2012. All informants consented to permitting the interviews to be audio recorded.

To protect their anonymity, interviewees were identified in the recording transcripts only by number (assigned by the order in which they interviewed) and their self-classification as Success Cases, Non-Success Cases, and Non-Evaluators at both Level 3 and Level 4. No other survey data was connected to the interviewees. The classification of the interviewees is shown in Table 28.

Table 28. Interviewees by Success Categories.

Classification	# of Interviewees
Level 3 Success Case, Level 4 Success Case	5
Level 3 Success Case, Level 4 Non-Success Case	1
Level 3 Success Case, Level 4 Non-Evaluator	6
Level 3 Non-Success Case, Level 4 Non-Evaluator	3
Level 3 Non-Evaluator, Level 4 Non-Evaluator	7

Interviews were scheduled to run between 15 and 30 minutes; actual interview times ranged from 12 minutes to 40 minutes, with most lasting between 20 to 25 minutes.

These interviews were semi-structured, which allowed for the interviewer to focus on specific topics and the subject to provide free-form responses, which opened the possibility for follow-up questions and more open-ended answers. Based on initial answers to each question, the interviewer asked probing questions in an effort to elicit narrative-type answers. Different interview scripts were developed for Success Cases, Non-Success Cases, and Non-Evaluators. Interviews were kept flexible to allow informants to speak at length about topics that had the greatest effect on their evaluation efforts.

The Impact Model developed for this research (see Table 11) was based around Gilbert's (2007) Behavior Engineering Model (BEM). The six components of the BEM served as the primary structure for the qualitative data analysis.

Interview statements were first sorted into two broad domains:

- Factors that facilitated evaluation
- Factors that obstructed evaluation

Within these domains, statements were deductively coded using the BEM structure. A visual representation for one of the domains is shown in Figure 2.

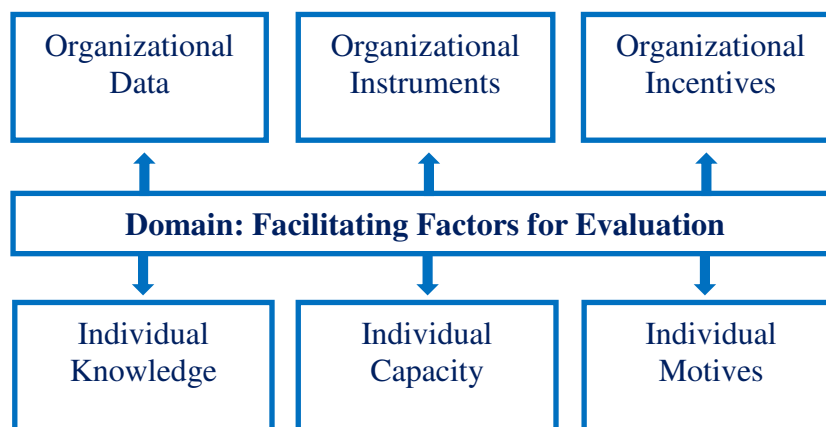


Figure 2. Deductive coding structure sample.

As statements within each domain were divided up into the six BEM components, patterns were identified that signified specific factors most mentioned by interviewees as impacting their efforts to evaluate training at Level 3 and Level 4.

A limitation to this qualitative analysis is the lack of triangulation data, as the only source for information on factors impacting evaluation were the informants who may have been influenced by personal biases, lack of detailed knowledge about organizational priorities and goals, or other issues. The conclusions drawn from the qualitative data are based on the informants' perceptions and interpretations, rather than objective and validated facts.

Facilitating Factors for Level 3 and Level 4 Evaluations

Without exception, every interviewee identified support from organizational leadership as the most critical factor in enabling successful completions of Level 3 and

Level 4 training evaluations. The statements about organizational support were mainly distributed between Organizational Data and Organizational Instruments; few statements were classified specifically as Organizational Incentives.

Organizational Data

In the Impact Model for this research (see Table 11), the critical actions for organizational data are setting clear expectations for the conducting of training evaluation and facilitating the exchange of information (data and feedback) between the training department and the rest of the organization.

Organizations bound by government regulations or accreditation requirements set an expectation of evaluation by necessity; facilities not in compliance with evaluation requirements may lose access to contracts, funding, or permission to operate. Those organizations may be encouraged or required to evaluate some or all of its training programs over a specified period of time; for example, a subject working for a government-regulated organization said that the guidelines strongly recommended that 100% of the training courses offered be evaluated at Level 3 over each five-year period. The regulations or guidelines might not specify how an evaluation should or must be performed, but it would specify the areas of focus and type of output data required from the evaluation.

Other organizations institute policies encouraging or requiring Level 3 or Level 4 evaluation of training interventions for voluntary reasons. Two Level 3 Non-Evaluator/Level 4 Non-Evaluator informants mentioned that their industry best practices included a shift from numbers-based performance assessment to competency-based assessment; they anticipated that as the value of competency becomes more established,

so will the value of training towards competency and the value of evaluating the level of competency produced through training. One noted, “training had never really been tied to job performance before; now the trend is linking it to business goals.” The subject also said, “the competition is who has the smartest and brightest people, so I think training and the evaluation of training is going to get a lot more attention now.” A Level 3 Success Case/Level 4 Success Case described how she believed that her organization strived for maximum efficiency in all areas, based on its strong focus on performance metrics; the organizational culture created by this focus also called for a regular and comprehensive evaluation of its training programs based around measurable outcomes.

When the leaders of an organization accept the need for evaluation, lower-level managers usually fall in line and cooperate in providing data and access to employees. However, this is not always the case. A Level 3 Success Case/Level 4 Success Case described his organization’s leadership as fully aware of the value of training evaluations, but he still faced resistance from the lower-level managers who held a narrower business outlook. “There’s not a blueprint for what it means to them; people compartmentalize, saying ‘it’s a good idea in theory but we don’t need it.’” The subject found that this resistance broke down when he could show them evaluation results that had led to improvements made at the departmental level. A Level 3 Success Case/Level 4 Non-Evaluator reported experiencing a similar increase in interest and cooperation through showing department managers examples of how training evaluation improved performance in other departments.

Some interviewees self-reported as Success Cases in the survey, but in reality had only been successful in completing a very limited number of Level 3 evaluations. Their

organizations were not generally supportive of evaluating training at this level, but in certain instances an executive requested an evaluation. This “squeaky wheel” enabled the training department to gain the temporary access to data and resources it required to conduct the evaluation. When the executive was satisfied, the support for evaluation was removed until the next executive-level request.

Organizational Instruments

The critical actions for organizational instruments are the allocation of resources and tools needed to conduct evaluations. Interviewees mentioned tools such as standardized evaluation processes and learning management system software, but the allocation of resources for evaluation generated most of the discussion.

In this situation, resources are a combination of personnel available for conducting evaluations and the time allowed for evaluation. Monetary resources are also a factor, both in the expense of hiring sufficient personnel and the cost of the materials, expertise, or tools needed for the evaluations. Even when the leadership is eager to evaluate training, the organization’s ability to do so is dictated by the number of personnel it can justify keeping on its training staff. As a Level 3 Success Case/Level 4 Non-Evaluator pointed out, the CEO’s enthusiasm for the value of training evaluations has not saved the training department from being downsized along with the rest of the organization. Another Level 3 Success Case/Level 4 Non-Evaluator observed that despite the guidelines in place for conducting regular evaluations and the personnel officially allocated to conduct evaluations, the reality is that evaluation loses out to priorities that have a more immediate or critical impact on the organization’s performance, and the personnel may be reassigned to one of those priority functions. A Level 3 Success

Case/Level 4 Non-Evaluator stressed the importance of collaborating with management at all levels; if managers make their existing performance data available to the evaluators, this reduces the need for the evaluators to take the time to collect that data themselves.

Organizational Incentives

The critical action for organizational incentives is the highly visible endorsement of the value of evaluation. Only one informant, a Level 3 Non-Evaluator/Level 4 Non-Evaluator, mentioned this as a factor; her division's leader was very vocal to his employees about his goals for his part of the organization. He had discussed with the informant his recent realization that training was a key component in achieving those goals and that he believed evaluation was a necessity towards ensuring the effectiveness of the training. Consequently evaluation efforts have begun, and the informant expects to be successful in getting support from all levels. "There is absolute buy-in from the director now, and everyone will follow his lead."

Individual Knowledge, Capacity, and Motives

The critical actions for these components are knowledge of evaluative methodology and effective communication skills, the cognitive capacity to analyze data, and the intrinsic motivation to perform these abilities at a high level. A Level 3 Non-Evaluator/Level 4 Non-Evaluator and a Level 3 Success Case/Level 4 Non-Evaluator both mentioned that they had done graduate-level work in programs related to instructional design and felt more confident about their future ability to both evaluate training and convince upper management of the value of evaluation. None of the other informants specifically mentioned knowledge or capacity as a facilitating factor. Most of the informants discussed methods, strategies, or theories as a matter of course but did not

single out their knowledge of these topics as important factors for conducting evaluations.

One Level 3 Success Case/Level 4 Non-Evaluator indirectly credited motive as a facilitating factor. His organization did not support Level 3 or Level 4 evaluations, but he felt it was too important to ignore and created a way to collect limited Level 3 data without the organization's cooperation or even knowledge.

Obstructing Factors for Level 3 and Level 4 Evaluations

As with the facilitating factors, the obstructing factors were dominated by organizational-level issue and were mainly distributed between Organizational Data and Organizational Instruments. However, although informants did not directly mention Individual Knowledge as an obstructing factor, their statements about the organizational factors revealed that knowledge appeared to be a key element in obstructing evaluations at Level 3 and Level 4.

Organizational Data

An organization with no expectation for evaluation is not necessarily unsupportive of evaluation efforts. It may simply not be a priority, or not yet a priority. A Level 3 Non-Success/Level 4 Non-Evaluator observed that his organization was growing rapidly yet was still so small that upper management knew the lower-level employees and could directly observe how they performed. His attempts to conduct evaluations had been unsuccessful because management did not yet perceive a need for data-driven evaluation when the employees still worked in a single office location; as the company continues to grow, this attitude is likely to change in the future as employees are split into multiple

offices and employee roles become more specialized. Other interviewees reported a similar lack of priority given to evaluation efforts because the organization did not yet understand the value of evaluation, either because it had never been done or had not been done in such a way that the organization could see the benefits.

Organizational Instruments

The obstructing factor mentioned repeatedly, even by those who had been successful at conducting Level 3 and Level 4 evaluations, was resources. Without enough personnel given enough time, it was difficult to collect and analyze evaluative data.

Those informants working in organizations that followed governmental guidelines were expected to evaluate every course at least once in each five-year period. In reality, the informants reported that they simply did not have enough personnel to achieve this goal; development and implementation of training were higher priorities, so personnel would be re-directed to those tasks. This was echoed by informants in other types of organizations; when one training project was developed, the personnel had to move on to the next project and then the next, without the luxury of time to evaluate the effectiveness of the projects already created. As one Level 3 Non-Evaluator/Level 4 Non-Evaluator noted, “we assemble the training, put it out there, and that’s it, there’s always another project waiting to be done so assessment wasn’t part of it.”

Organizational Incentives

When there is no visible push from upper management to support evaluative efforts, there may be no incentive for employees to cooperate with those efforts. One Level 3 Non- Success/Level 4 Non-Evaluator reported that he had given up entirely on all

levels of evaluation for the time being because upper management did not support evaluation efforts and there were no repercussions for lack of cooperation from lower-level employees. Even the response rate for Level 1 surveys was poor. “I started out with getting 50%-75% response, but it started dwindling down to 10% or 15%... as soon as they figured out there were no repercussions for not filling it out, I think they just put it on the back burner.”

Individual Capacity and Motives

No informants singled out issues with their capability to analyze data or willingness to conduct evaluations, although several noted how their failure to succeed in promoting evaluation had a negative impact on their morale and their respect for the organization.

Individual Knowledge

No informants singled out a lack of knowledge about evaluation methodology or concepts as an obstructing factor for Level 3 and Level 4 evaluation. However, their statements about the relevance of evaluation indicated that individual knowledge may be a critical obstructing factor for conducting Level 4 evaluations.

Kirkpatrick (1996) defined his Level 4 as evaluating training in terms of the desired results that prompted the training in the first place. He included such goals as reduced costs and increased productivity as examples of what would be measured at this level and acknowledged the difficulty in measuring less quantifiable outcomes. While Kirkpatrick did not limit the scope of Level 4 evaluations to such quantifiable data, training professionals who “know of” the Kirkpatrick levels may interpret Level 4 solely

as return on investment (ROI) evaluations, which rate the results of the training against its cost to determine its worth to the organization.

Six interviewees had self-reported as Level 3 Success Cases/Level 4 Non-Evaluators. Of those six, four stated that their organizations did not conduct Level 4 because their organizations did not need to measure ROI for training. Two of them discussed return on expectations (ROE), an evaluation concept based on expected outcomes for the program being evaluated, without considering it to be the same as Kirkpatrick's Level 4. One of these Level 3 Success Cases/Level 4 Non-Evaluators stated, "we're not as concerned with a dollar-for-dollar basis... we're not really in the situation where we have to justify our training; it's really all about return on expectations." This organization created expectations prior to training and then measured its success at achieving those expectations, but this informant did not classify such a measurement as a results-based evaluation.

When one self-reported Success Case for Level 4 evaluation described his method for collecting Level 4 data, it was evident that the data was on behaviors rather than outcomes; the informant called it Level 4 simply because it took place at a later date than the Level 3 evaluation. His organization also collected data solely through surveys sent to a sample of the learners, with no triangulation. "I'm not aware of what I would consider direct measures of performance, nothing like observation or performance review data."

The Organizational Factor

Rummler and Brache (1995) were among those who stressed the importance of focusing on the organizational level when examining performance issues: "If you pit a good performer against a bad system, the system will win almost every time" (p. 64).

Among the informants for this research were two individuals who appeared to have similar levels of training-related education, perception of the importance of behavior-based evaluation, and professional drive. One informant (“Jane”) appeared to be an exemplar Level 3 Success Case, but the other informant (“Joe”) could only achieve a rudimentary degree of evaluation at Level 3. Why was Jane a success at evaluating, but not Joe?

Jane was employed at a large organization, with multiple facilities, in an industry subject to governmental regulations. The organization had developed a culture that emphasized meticulous data collection. “We are so data-driven with quantifiable data; that’s what drives us.” Jane held a leadership position responsible for the functional training of employees in a division with highly quantifiable productivity measurements and a very structured employee environment. Data on productivity was readily available through the organization’s enterprise software, which allowed Jane to easily generate reports on the current productivity levels of individuals or groups. “Typically they’re compared with peers on their team; plus we compare teams across sites.”

Jane’s organization viewed Level 3 evaluations as critical in ensuring that employees are performing with consistent quality and accuracy; Jane had access to both the quantitative data and to qualitative data collected through focus groups with employees and input from their line managers. Any barriers in place were strictly logistical, not cultural.

Despite the structured environment, Jane had the support of management to test out new methods for improving the efficiency of training. For example, classroom training was already available to line employees wishing to move into specialist

positions. Jane recently implemented a program that allowed line employees to engage in self-directed learning and take assessment exams in lieu of attending regular classroom sessions. This, like other aspects of training, will be measured and analyzed on its efficiency and relevance. Jane said, “We want to take the assessment scores of the self-directed learners and compare them to those taking the actual class to see if there’s a significant difference.”

Joe was employed in a non-managerial training role at a medium-sized organization with a single location. Joe’s group developed training both for internal clients and for external vendors who were required to follow certain procedures for ordering products. Joe’s success at conducting Level 3 evaluations was limited to the external vendors; his group was barred from attempting such evaluations of internal training. Joe said, “I wish I could say we did, but we don’t.”

The organization had asked the training department to follow up with external clients, but did not offer guidelines. Joe said, “They’re not showing us what kinds of data they want; they’re not giving us specific questions to ask; they allow us to call the customer using our system; that’s the type of support we get.” Joe’s group developed a post-training interview for follow-up, and they “just turned that follow-up into a level 3 evaluation,” focusing on application usage and comfort level with the procedures. The organization was not interested in receiving results, so they were only shared within the training group unless they indicated a serious problem to be resolved.

As for evaluating internal training, the organization did not allow Level 3 evaluations: “It’s kind of an inconvenience, at least from the feedback I got in the past.” Joe felt this had a negative impact on the quality of internal training by saying, “because

we don't have that feedback saying what we could do better." Joe had tried presenting the case for evaluation to the organization but felt that management viewed his group strictly in an order-taking role. So why did the organization allow Level 3 evaluations of external clients when it prohibited such evaluation of internal training? Joe expressed, "I think they're not aware; they're kind of oblivious to the fact that we're taking that opportunity to get those metrics."

Jane perceived her organization as focused on high levels of performance and strongly supportive of her group's efforts to maintain and improve such performance. Joe named his organization's leadership and vertical organizational structure as the primary barriers to any efforts to improve training efforts. The system will win almost every time.

CHAPTER FIVE: DISCUSSIONS AND CONCLUSIONS

Discussion

Frequency of Level 3 and Level 4 Evaluations

The sample of training professionals who participated in this research reported that their organizations conducted Level 1 evaluations for approximately 88% of their training interventions, Level 2 for approximately 74.5%, Level 3 evaluations for approximately 43.5%, and Level 4 evaluations for approximately 18.5%. Although these numbers are somewhat lower than what Pulichino (2007) and ASTD (2009b) found, they follow the same trend of a decreasing percentage of evaluations performed as the level increases.

Survey respondents did not deny the value of conducting Level 3 and Level 4 evaluations. When asked if the data collected at those levels was sufficient to evaluate the effectiveness of their training programs, about 62% of respondents thought Level 3 data was useful for this purpose, and about 43% of respondents found Level 4 data valuable.

Looking at the perceived importance of conducting evaluations at each level, about 98.5% of survey respondents rated Level 3 evaluation as very or extremely important to conduct, with about 88% rating Level 4 evaluation as very or extremely important.

This matched the trend found by ASTD (2009b), wherein Level 3 data was considered more valuable than Level 4. This indicates that the assumption that the evaluations levels have a progressively higher value, as stated by Alliger and Janak (1989), may no longer be a strongly held belief among training professionals.

Some training leaders are pushing to enhance the credibility of training interventions through taking a broader strategic view of how they forward an organization's goals (Berk, 2005; Kirkpatrick & L'Allier, 2004; Vance, 2011). A results-based evaluation of training, with its emphasis on business outcomes, would be a way for training professionals to demonstrate the value they contribute to an organization. Yet Level 4 evaluations are perceived by survey respondents as less important to conduct than behavior-based Level 3 evaluations, and both are conducted with less frequency than reaction- and knowledge-based evaluations.

Stakeholders

The impetus for conducting a Level 3 or Level 4 evaluation appeared to arise from two main sources. The request can come from an organization's upper management in response to either internal needs, such as budgetary or effectiveness studies, or external needs such as agency regulations or guidelines. The real driving force, according to the survey data, is the training department itself; about 65% of both Level 3 and Level 4 evaluations were requested in whole or part by the training department itself.

Reasons for Evaluation

The prevalence of training departments as stakeholders may be explained by the answers selected by survey respondents when asked the reasons for conducting

evaluations. For Level 3, 45% of Success Cases and 60% of Non-Success cases attempted the evaluations in part to demonstrate the value of the department's contribution to the organization; only assessing the content and methods was cited more frequently by respondents. For Level 4, those figures rose to 69% for Success Cases and 75% for Non-Success Cases, with the demonstration of departmental value the most frequently selected reason for attempting evaluation.

The most frequently given reason for conducting Level 3 evaluations was to assess what parts of the knowledge or skills taught in a training intervention were being used on the job and which should be revised or removed; 76% of Level 3 Success Cases and 60% of Level 3 Non-Successes had attempted evaluation for that purpose. Only 21% of Success Cases and 30% of Non-Success Cases reported attempting Level 3 evaluations due to organizational requirements; 25% of Level 4 Success Cases conducted their evaluations in part to meet organizational requirements, although zero Level 4 Non-Success Cases did so. This reason might have been selected more frequently if it had been phrased to include organizational guidelines and cultural expectations; several interviewees discussed these organizational drivers, which were not written policy but nevertheless set expectations for evaluation.

Barriers and Facilitators

The study also looked at reasons why Level 3 and Level 4 evaluations were not done. Level 3 Non-Evaluators most frequently cited lack of resources (71%), lack of support from organizational management (62%), and lack of expertise in evaluation methods (48%) as reasons why they did not attempt Level 3 evaluations. At Level 4, the

lack of resources continued to be the top issue preventing evaluation attempts (57%), with lack of expertise a close second (55%) and access to data third (50%).

In Pulichino's (2007) survey, the greatest perceived barriers for Level 3 evaluations were access to data (86.2%), lack of support from organizational management (85.1%), and lack of time (83.3%). For Level 4 evaluations, the top barriers were access to data (90.5%), lack of time (85.1), and lack of support from organizational management (79.5%).

Pulichino's (2007) survey inquired about lack of expertise; according to his results, it was an important barrier (52.2% for Level 3 and 64.7% for Level 4) but ranked only sixth out of seven at both levels.

The concern about lack of expertise among this study's participants is interesting because about 81% of respondents indicated that they completed at least one university class, workshop, or other type of educational offering specifically on the topic of evaluation. This could be a case of the respondents recognizing what they do not know.

The survey questioned respondents on the availability, or lack thereof, of factors likely to affect their ability to conduct Level 3 and Level 4 evaluations. Each factor was rated on a scale from 1 (no availability) to 5 (definite availability), and scores for each respondent category were averaged. The assumption would be that those respondents identified as Success Cases would report a greater degree of availability for most of those factors than Non-Success Cases would.

Surprisingly, this was not the case for Level 3 evaluations. It was the Level 3 Non-Success Cases, rather than the Level 3 Success Cases, who reported greater

availability of expertise (3.20 for Non-Success Cases versus 3.09 for Success Cases), support from organizational management (2.90 versus 2.85), and importance placed on evaluation by the organization (3.00 versus 2.76). Data from Non-Evaluators rated every factor's availability lower than for either of the other two categories.

Level 4 evaluations, however, met that assumption, with Success Cases rating availability higher for all factors than did Non-Success Cases and Non-Evaluators.

The interview data collected for this study supported the survey in identifying organizational support and availability of resources as the key elements that determine if Level 3 and Level 4 evaluations are completed successfully. It also revealed another element to the issue of lack of expertise. As mentioned in the findings, four of the six Level 4 Non-Evaluators interviewed stated that Level 4 evaluations of training were not relevant to their organizations. The reasons given were that training was so essential to the organization that training costs were not considered an issue or that the organization was not driven by quantifiable financial results and thus could not be evaluated at Level 4.

These interviewees equated Level 4, the results-based evaluation of training, only with ROI and the measurable financial impact. Kirkpatrick (1996) used some easily-quantifiable data as examples of performance metrics for a Level 4 evaluation, but he did not restrict this level of evaluation to cost-related analysis. McLinden and Trochim (1998) developed the concept of return on expectations (ROE), which Kirkpatrick and Kirkpatrick (2010) have championed as a clearer concept for examining the organizational results of training interventions. Interviewees asked about ROE confirmed it was a useful concept, but most did not see it as Level 4 evaluation.

Conclusions

In 1990, two-thirds of the training managers surveyed by ASTD believed they were under increased pressure to demonstrate the value that training brought to their organization, yet only about 10% of their training interventions were evaluated at Level 3 and about 25% at Level 4 (Carnavale, 1990). In 2009b, ASTD conducted another evaluation survey and found that 54.6% of the respondents reported conducting Level 3 evaluations to any extent and 36.9% reported conducting Level 4 evaluations to any extent (ASTD, 2009b).

Over the past two decades, the frequency of Level 3 evaluations has increased markedly but is nevertheless performed for only about half of all training interventions. Despite the continuing need to demonstrate the value of training to organizations in terms of achieving organizational goals, the frequency of Level 4 evaluations has remained low. Training professionals have been aware of the necessity of conducting evaluations at both Level 3 and Level 4 for many years, yet such evaluations are still not regularly done.

Based on the findings of this study, which correspond to the findings from Moller and Mallin (1996), Pulichino (2007), and ASTD (2009b), two of the primary reasons for this appear to be related to organizational support of evaluation activities. Training professionals cited both a lack of resources, including adequate time and personnel, and a lack of managerial support among the top reasons they did not conduct Level 3 and Level 4 evaluations. Both types of evaluations require considerable time to conduct and a high level of cooperation from management at different levels in order to collect accurate evaluative data.

The third primary reason appears to be the need for more education about evaluation and evaluative methods. Training professionals cited a lack of evaluative expertise in their departments as a top reason for not conducting Level 3 and Level 4 evaluations, and the interview data indicated that training professionals might not understand the purpose and potential benefits for Level 4 evaluations for all types of organizations.

Limitations of the Study

One limitation of this study was access to the target population and resulting small sample size. Similar research studies had been conducted on behalf of large professional organizations whose members were solicited for participation by those organizations. This study was conducted independently and in a short timeframe, which limited its reach to the target population of training professionals.

Another limitation is that both the survey and interviews were conducted with no face-to-face communication with any of the participants, so there was no opportunity to observe the informants, their interactions with others in the organization, and the physical work environment. It is possible that more data may have been collected through such observations, which could have added detail to the analysis and allowed for a deeper level of understanding of the facilitators and impediments to conducting evaluations.

The time restrictions on the study as a whole and for the individual interviews did not allow for in-depth descriptive data or purposive sampling, both of which are critical for establishing the transferability of findings in naturalistic inquiry (Guba, 1981). Longer narrative interviews, along with other ethnographic methods such as observation, would be more likely to provide findings applicable in other contexts.

Survey participants may have varying degrees of understanding of the terms and concepts in the survey. Respondents may answer certain questions differently depending on the phrasing of the question and the knowledge base of the respondent; an online survey offers no opportunity for clarification.

Factors considered of critical importance by survey respondents may not have been offered as answer choices on the survey. This could have been addressed by conducting narrative interviews with a small population sample before developing the survey questions.

Recommendations for Future Studies

The interview data revealed some confusion on how training professionals define Kirkpatrick's Level 4 for evaluation. It could be valuable to explore their interpretations of results-based evaluation and whether these differing interpretations impact their ability to get stakeholder buy-in to conduct such extensive evaluations of training interventions. This may be the most critical line of inquiry to follow up from this study, as training professionals may be defining "results" so narrowly that they do not consider results-based evaluations relevant; by doing so, they disregard a critical tool for demonstrating the value they bring to their organizations.

Despite the high percentage of respondents who reported taking one or more courses specifically related to evaluation, the survey data indicates that respondents perceive the lack of expertise as an issue. It may be helpful to examine what training professionals understand about the methodology of evaluation, including terms and models, and how it can be used to structure effective evaluations. For example, training professionals may be relying heavily on self-reported data from learner surveys, without

the benefit of additional quantitative or qualitative data to form a more complete picture of how successfully learners are applying their new work behaviors.

As the training professionals involved in this study pointed to organizational factors as the primary barriers and facilitators for successful evaluation, it may be helpful to study how organizations perceive the role of their training departments. In a highly vertical organization, training departments may be compartmentalized into a restricted role with limited access to information or feedback from other departments. In other organizations, the training department may function as a partner or valued strategic resource. The organizational culture in regards to the role of training may promote self-selection by training professionals into or out of the organization, which in turn affects the level of knowledge, capacity, and motivation within the training department. How does this affect evaluation efforts? Have training departments been successful at using Level 3 and Level 4 evaluation as a tool to change their organizational roles? These are questions future researchers might seek to answer.

REFERENCES

- Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology*, 42(2), 331–342.
- American Society for Training & Development. (2009a) *Training budgets under the microscope*. Retrieved from <http://www1.astd.org/TDimages/dec08/InfoGraphLG.png>
- American Society for Training & Development. (2009b). *The value of evaluation: Making training evaluations more effective*. Alexandria, VA: ASTD Research.
- American Society for Training & Development. (2010). *2010 ASTD state of the industry*. Alexandria, VA: ASTD Research.
- Berk, J. A. (2005). Identifying tangible business results. *Chief Learning Officer*, 4(3), 30-35.
- Brinkerhoff, R. O. (2003). *The success case method: Find out quickly what's working and what's not*. San Francisco, CA: Berrett-Koehler Publishers, Inc.
- Carnavale, A. (1990). Evaluation practices. *Training & Development Journal*, 44(7), 63-69.
- Davidson, E. J. (2005). *Evaluation methodology basics*. Thousand Oaks, CA: Sage Publications, Inc.
- Dessinger, J. C., Moseley, J. L., & Van Tiem, D. M. (2011). Performance improvement/HPT model: Guiding the process. *Performance Improvement*, 51(3), 10-17.
- Expertus, Inc. & Training Industry, Inc. (2008). *Training efficiency: Operating costs*. Mountain View, CA: Expertus, Inc. Retrieved from http://www.expertus.com/sites/expertus.com/files/public-files/Survey-Results_Learning-Cost-Management_Expertus-08.pdf
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327-358.
- Gilbert, T. F. (2007). *Human competence: Engineering worthy performance* (Tribute ed.). San Francisco, CA: Pfeiffer.

- Guba, E. G. (1981). ERIC/ECTJ annual review paper: Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology*, 29(2), 75-91.
- Harless, J. (1976). Speaking from experience. *Training & Development Journal*, 30(3), 10.
- Holton III, E. F. (1996). The flawed four level evaluation model. *Human Resource Development Quarterly*, 7(1), 5–21.
- International Society for Performance Improvement. (2012). ISPI Performance Standards. Retrieved from http://www.ispi.org/uploadedFiles/ISPI_Site/About_ISPI/About/Standards.pdf
- International Society for Performance Improvement. (2012). What is HPT? Retrieved from <http://www.ispi.org/content.aspx?id=54>
- Kirkpatrick, D. (1996). Revisiting Kirkpatrick's four-level model. *Training and Development*, 50(1), 54-59.
- Kirkpatrick, D. (2010). 50 years of evaluation. *T+D*, 64(1), 14.
- Kirkpatrick, D. L., & L'Allier, J. J. (2004). Evaluation as a strategic tool. *Chief Learning Officer*, 3(6), 30-33.
- Kirkpatrick, J. D., & Kirkpatrick, W. (2010). ROE's rising star: Why return on expectations is getting so much attention. *T+D*, 64(8), 34-38.
- LeCompte, M.D., & Schensul, J.J. (1999). *Analyzing and interpreting ethnographic data*. Walnut Creek, CA: Alta Mira Press.
- McLinden, D. J. (1995). Proof, evidence, and complexity: Understanding the impact of training and development in business. *Performance Improvement Quarterly*, 8(3), 3-18.
- McLinden, D. J., & Trochim, W. M. K. (1998). Getting to parallel: Assessing the return on expectations of training. *Performance Improvement*, 37(8), 21–26.
- Moller, L., & Mallin, P. (1996) Evaluation practices of instructional designers and organizational supports and barriers. *Performance Improvement Quarterly*, 9(4), 82-92.
- Newstrom, J. W. (1978). Catch-22: The problems of incomplete evaluation of training. *Training & Development Journal*, 32(11), 22.

- Pershing, J. A. (2006). Human performance technology fundamentals. In Pershing, J.A. (Ed.), *Handbook of human performance technology: Principles, practices, and potential* (3rd ed.) (pp. 5-34). San Francisco, CA: Pfeiffer.
- Pulichino, J. P. (2007). Usage and value of Kirkpatrick's four levels of training evaluation. (Unpublished doctoral dissertation). Pepperdine University, Malibu, CA.
- Rummler, G. A., & Brache, A. P. (1995). *Improving performance: How to manage the white space on the organization chart*. San Francisco, CA: Jossey-Bass Inc.
- Seawright, J. & Gerring, J. (2008) Case selection techniques in case study research: A menu of qualitative and quantitative options. *Political Research Quarterly*, 61(2), 294-308.
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: SAGE Publications, Inc.
- Stufflebeam, D. L., & Shinkfield, A. J. (2007). Evaluation theory, models, & applications (Chapter 1). San Francisco: Jossey-Bass.
- Tyler, R.W. (1942). General statement on evaluation. *Journal of Educational Research*, 35(7), 492-501.
- Tyler, R.W. (1949). *Basic principles of curriculum and instruction*. Chicago, IL: University of Chicago Press.
- United States Department of Labor Bureau of Labor Statistics. (2012). Economic news release: Occupational employment and wages - May 2011. Retrieved from http://www.bls.gov/news.release/archives/ocwage_03272012.htm
- Van Tiem, D. M., Moseley, J. L., & Dessinger J. C. (2004). *Fundamentals of performance technology: A guide to improving people, process, and performance* (2nd ed.). Silver Spring, MD: International Society for Performance Improvement.
- Vance, D. (2011). Running learning like a business. *Chief Learning Officer*, 10(1), 50.

APPENDIX A

Gilbert's Behavior Engineering Model

Gilbert's Behavior Engineering Model

In 1978, Thomas Gilbert (2007) presented his Leisurely Theorems on human competence, which he defined as the state of creating valuable results without excessive cost. He asserted that behavior itself was only relevant in how it produced or prevented accomplishments or desired outcomes; it was essential to identify the relevant behaviors in order to increase performance. His Third Leisurely Theorem stated that a performance deficiency can be traced back to a deficiency in individual behavior, environmental (organizational) behavior, or a combination of both.

Gilbert (2007) developed a framework for examining a performance deficiency through identifying and classifying these individual and organizational behaviors. His Behavior Engineering Model (BEM) categorized behaviors as Information, Instrumentation, and Motivation, with each behavior type occurring at both the individual and organizational level.

See Table A.1 for the Behavior Engineering Model.

Table A.1. Gilbert's Behavior Engineering Model Including Definitions.

	Information	Instrumentation	Motivation
Environmental Factors (Organizational)	<u>Data</u> Job descriptions and performance expectations Clear guidance on how to perform to expectations Feedback on performance	<u>Resources</u> Tools, materials, and other resources required to meet performance expectations	<u>Incentives</u> Financial and non-monetary incentives for meeting performance expectations Opportunities for growth within the organization
Personal Repertoire (Individual)	<u>Knowledge</u> Employees with the knowledge and skills that meet the requirements for performance Employee skills matched to position or duties	<u>Capacity</u> Employees hired to meet the realities of the job expectations Employees capable of doing or learning what is required to meet performance expectations	<u>Motives</u> Employees hired to meet the realities of the work and work environment Employees with intrinsic motivation to meet performance expectations

APPENDIX B

Boise State University Institutional Review Board Notification of Approval



Office of Research Compliance

1910 University Drive
 Boise, Idaho 83725-1138
 HumanSubjects@boisestate.edu | 208.426.5401

DATE: December 19, 2011

TO: Perri Kennedy (PI)
 Seung Chyung (co-PI)

FROM: Institutional Review Board (IRB)
 Office of Research Compliance

SUBJECT: IRB Notification of Approval
 Project Title: *Training Professionals' Understanding and Usage of Kirkpatrick's Level 3 and Level 4 Evaluations*

The Boise State University IRB has approved your protocol application. Your protocol is in compliance with this institution's Federal Wide Assurance (#0000097) and the DHHS Regulations for the Protection of Human Subjects (45 CFR 46).

Review Type: Expedited	Approval Number: 217-SB11-108
Date of Approval: December 19, 2011	Expiration Date: December 18, 2012

Your approval is effective for 12 months. If your research is not finished within the allotted year, the protocol must be renewed before expiration date indicated above. The Office of Research Compliance will send a reminder notice approximately 30 days prior to the expiration date. The principal investigator has the primary responsibility to ensure a RENEWAL FORM is submitted in a timely manner. If the protocol is not renewed before the expiration date, a new protocol application must be submitted for IRB review and approval.

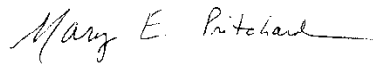
Under BSU regulations, each protocol has a three-year life cycle and is allowed two annual renewals. If your research is not complete by December 18, 2014, a new protocol application must be submitted.

All additions or changes to your approved protocol must also be brought to the attention of the IRB for review and approval before they occur. Complete and submit a MODIFICATION/AMENDMENT FORM indicating any changes to your project. When your research is complete or

discontinued, please submit a FINAL REPORT FORM. An executive summary or other documents with the results of the research may be included.

All relevant forms are available online. If you have any questions or concerns, please contact the Office of Research Compliance, 426-5401 or HumanSubjects@boisestate.edu.

Thank you and good luck with your research.

A handwritten signature in cursive script that reads "Mary E. Pritchard".

Dr. Mary E. Pritchard

Chairperson

Boise State University Institutional Review Board

APPENDIX C

Solicitation for Survey Participants



RECRUITMENT E-MAILS

Research Survey: Training Professionals' Understanding and Usage of Kirkpatrick's Level 3 and Level 4 Evaluations

Text for LinkedIn posts soliciting participants:

My name is Perri Kennedy and I am a graduate student in Instructional & Performance Technology at Boise State University. As part of my thesis research, I am surveying training and development professionals to learn about their usage of behavior-based (Kirkpatrick Level 3) and outcome-based (Kirkpatrick Level 4) training evaluations in their organizations. The goal of my research is to examine the organizational and individual factors which act as barriers or facilitators when performing such evaluations, and how training professionals can work with or around them.

If you are a training professional, I would appreciate your participation in this survey. It will take you 30 minutes or less to complete it, and your answers will be kept strictly confidential.

If you are willing to participate, please click the link below to access the survey.

https://boisestate.qualtrics.com/SE/?SID=SV_bQQLVrOdRA8DHYU

APPENDIX D

Survey Cover Letter



SURVEY COVER LETTER OF INFORMED CONSENT

Training Professionals' Understanding and Usage of Kirkpatrick's Level 3 and Level 4 Evaluations

PURPOSE AND BACKGROUND: Perri Kennedy, a graduate student in Instructional & Performance Technology at Boise State University, is conducting a research survey to learn more about training professionals' experiences with performing behavior-based (Kirkpatrick Level 3) and outcome-based (Kirkpatrick Level 4) training evaluations. The information gathered will be used to better understand if and when training professionals perform such evaluations, and to learn which individual and organizational factors may impact performing the evaluations. You are being asked to participate because you are a participant in an online or face-to-face community for training professionals, and are currently employed as a training professional.

PROCEDURES: If you agree to participate in this study, you will be connected to an online survey. The survey includes demographic questions, multiple choice questions, and open-ended questions. The survey will take 30 minutes or less to complete. We ask that you try to answer all questions; however, you may skip any questions that you prefer to not answer. At the end of the survey, you may accept or decline the opportunity to participate in a post-survey interview which will be conducted in early February 2012.

RISKS: The risks to you, personally or professionally, are minimal. The researcher will make all possible effort to maintain the confidentiality of your identity and responses.

BENEFITS: You will not be paid for participating in this survey. Although you will not receive any direct benefit from participation, this research may benefit training professionals who seek to be more successful in evaluating training programs.

EXTENT OF CONFIDENTIALITY: Your survey responses will be handled as confidentially as possible. Your name or other identifying information will not be used in any written reports or publications which result from this research. Only the principal researcher and co-investigators will have access to the research data. Data will be kept for three years (per federal regulations) after the study is complete and then destroyed.

For this research project, we are requesting demographic information. The combined answers to these questions may make an individual person identifiable. We will make every effort to protect participants' confidentiality. However, if you are uncomfortable answering any of these questions, you may leave them blank.

PARTICIPATION IS VOLUNTARY: You may decline to participate in this survey. You may also refuse to answer any questions you do not want to answer. If you volunteer to participate in this survey, you may withdraw from it at any time without consequences by closing your browser window; by doing so, you end the session and your answers will not be recorded.

QUESTIONS: If you have any questions or concerns feel free to contact the principal researcher, Perri Kennedy, or her faculty advisor, Dr. Yonnie Chyung:

Perri Kennedy, graduate student

Instructional & Performance Technology

Boise State University

Phone: (301) 787-1449

Email: perrikennedy@u.boisestate.edu

Dr. Yonnie Chyung, Professor

Instructional & Performance Technology

Boise State University

(208) 426-3091

ychyung@boisestate.edu

If you have questions about your rights as a research participant, you may contact the Boise State University Institutional Review Board (IRB), which is concerned with the protection of volunteers in research projects. You may reach the board office between 8:00 AM and 5:00 PM, Monday through Friday, by calling (208) 426-5401 or by writing: Institutional Review Board, Office of Research Compliance, Boise State University, 1910 University Dr., Boise, ID 83725-1138.

If you voluntarily agree to participate, please click the Continue button to continue to the survey. By completing the survey, you are consenting to voluntarily participate in this research.

If you would prefer not to participate, please click the Decline button, or simply close your browser to end the session.

APPENDIX E

Full Survey

[NOTE: Any text in RED does not appear in the actual survey.]

Training Professionals' Understanding and Usage of Kirkpatrick's Level 3 and Level 4 Evaluations

Introduction

You are invited to participate in a research project which examines how organizations conduct behavior-based and outcome-based evaluations of training programs. Your completion of this survey will provide valuable information for this research, which in turn will contribute valuable information to the training & development professional community.

Statement of Anonymity

The data collected through this survey will be used by Perri Kennedy, a graduate student at Boise State University, as part of a thesis research project. Ms. Kennedy is required by the university to maintain strictest data confidentiality. Your participation in this survey is voluntary, and your responses will be kept completely anonymous. If you choose to not complete the survey, you may exit at any time by closing your browser window.

Eligibility Screening

Please select the one that best describes your work:

- As an internal employee of my organization, my primary job tasks involve training & development, human resources, organizational development, or related fields for the internal clients of my organization. **[send to SURVEY START]**
- As an internal employee of my organization, my primary job tasks do not involve training & development, human resources, organizational development, or related fields. **[send to FREE RESPONSE]**
- I work as an external consultant for my client organization(s). **[send to FREE RESPONSE]**

SURVEY START

Demographics

Select your gender:

- Male
- Female

Select your age range:

- in the 20s
- in the 30s
- in the 40s
- in the 50s
- in the 60s or older

Which of the following best describes your primary job function?

- Instructional designer
- Trainer
- Human resources professional in training & development
- Human resources professional (other)
- Organizational development specialist
- Manager in training & development, organizational development, or human resources
- Manager (other)
- Other – Describe: _____ **[free response field]**

How many total years of experience do you have in training & development, human resources, organizational development, and/or related fields?

- Less than 2 years
- 2 to 5 years
- 6 to 15 years
- More than 15 years

For approximately how many employees do you provide training & development, human resources, organizational development, and/or related services?

- Fewer than 100
- 100-500
- 501-1500
- More than 1500
- Not sure

Please indicate the type of organization you work for:

- Consulting
- Education
- Finance

- Government
- Health Care
- Insurance
- Manufacturing
- Military
- Retail
- Telecommunications
- Other - Please specify: _____ [free response]

What is the highest level of education which you have completed?

- High school diploma
- Associate's degree
- Bachelor's degree
- Master's degree
- Doctoral degree

Do you have formal education in **instructional design** or **educational technology** or **related fields**? Check all that apply.

- Associate degree in instructional design, educational technology, or related fields
- Bachelor's degree in instructional design, educational technology, or related fields
- Master's degree in instructional design, educational technology, or related fields
- Doctoral degree in instructional design, educational technology, or related fields
- Certificate in instructional design, educational technology, or related fields

Have you taken courses specifically on **training evaluation** or **program evaluation**? Select one.

- a credit-based university course
- a non-credit based workshop or seminar
- both credit and non-credit courses
- no, I have not taken any courses specifically on training evaluation or program evaluation

Section 1

Kirkpatrick's 4 Levels of Evaluation model classifies types of training evaluations.

- Level 1 (Reaction) looks at how learners feel about the training itself.
- Level 2 (Knowledge) gauges how well learners acquire new information during the training.
- Level 3 (Behavior) examines whether learners actually use this new information on the job.
- Level 4 (Results) measures how well the training achieved the organizational goals it was designed to meet.

Questions in this section will ask you about your experiences with conducting Kirkpatrick's four levels of evaluations of training programs implemented in your organization. Please answer the questions to the best of your knowledge.

Approximately, on what percentage of the total number of training programs does your department conduct Level 1, Level 2, Level 3 and Level 4 evaluations?

Level 1	• 0% -----100% of the total training programs
Level 2	• 0% -----100% of the total training programs
Level 3	• 0% -----100% of the total training programs
Level 4	• 0% -----100% of the total training programs

How sufficient do you think Level 1, Level 2, Level 3, and Level 4 evaluations of training are able to judge the effectiveness of your organization's training programs?

Level 1	• Not sufficient at all 1 2 3 4 5 Sufficient
Level 2	• Not sufficient at all 1 2 3 4 5 Sufficient
Level 3	• Not sufficient at all 1 2 3 4 5 Sufficient
Level 4	• Not sufficient at all 1 2 3 4 5 Sufficient

How important do you think it is to conduct Level 1, Level 2, Level 3, and Level 4 evaluations of training programs?

Level 1	• Not important at all 1 2 3 4 5 Extremely important
Level 2	• Not important at all 1 2 3 4 5 Extremely important
Level 3	• Not important at all 1 2 3 4 5 Extremely important
Level 4	• Not important at all 1 2 3 4 5 Extremely important

Rank the four levels of Kirkpatrick's evaluation model in terms of their importance:

- Level 1
- Level 2
- Level 3
- Level 4

Please explain briefly your rationale for ranking the levels in this order:

In your opinion, positive evaluation results at Level 1 or Level 2 are a strong indicator of:
(Select one):

- positive evaluation results at Level 3
- positive evaluation results at Level 4
- positive evaluation results at both Level 3 and Level 4
- no relationship between the levels

Please explain briefly your rationale for choosing this answer: -

Which of the following best describes your department's efforts to conduct Kirkpatrick's Level 3 (Behavior) evaluations of training programs implemented in your organization?

- My department has completed at least one Level 3 evaluation of a training program. [\[send to L3 SUCCESS CASES\]](#)
- My department has attempted to conduct at least one Level 3 evaluation of a training program, but was unable to complete the evaluation. [\[send to L3 NON-SUCCESS CASES\]](#)
- My department has not attempted to conduct Level 3 evaluations of training programs. [\[send to L3 NON-EVALUATORS\]](#)

L3 SUCCESS CASES

Who in your organization requested that a Level 3 evaluation be conducted for a training program? Select all that apply.

- Manager of the department/workgroup which attended training
- Management above the level of the manager of the department/workgroup which attended training
- The training & development, organizational development, or human resources department
- Others - Please explain: _____

For what reasons did your department conduct Level 3 evaluation of training? Select all that apply.

- To assess what parts of the training were being used, and which were not so that changes to training content and methods could be made, if needed
- To help decide what parts of the organization were effectively executing important new behavioral changes (e.g., strategy, new technology), and which were not, so that remediation efforts could be implemented where needed.
- To determine whether and how much managers of trainees were supporting on-job application of learning.
- To assess the efficacy of the training with a small or “pilot” group of trainees so that changes could be made to increase impact of later and larger roll-outs.
- To help determine what factors were impeding or facilitating on-job application of learning
- To discover emerging best-practices of on-job application so that these could be communicated to other trainees and managers, and also might be built into the training content.
- To demonstrate the value of your department’s contributions to the organization
- Because it was required by the organization
- Because it just seemed like the right thing to do
- Others - Please explain: _____

Who in your organization received the results of a Level 3 evaluation? Select all that apply.

- Manager of the department/workgroup which attended training
- Management above the level of the manager of the department/workgroup which attended training
- The training & development, organizational development, or human resources department
- Others - Please explain: _____

Below is a list of factors which may affect your department's ability to conduct Level 3 evaluation of training. In your opinion, **how available** to your department is each of the following factors?

Factor	Availability						
The department's expertise/experience in using evaluative methods	Totally lacking	1	2	3	4	5	Sufficient
Access to employees for post-training surveys or interviews	Totally lacking	1	2	3	4	5	Sufficient
Access to other post-training data, such as performance measurements or employee evaluations	Totally lacking	1	2	3	4	5	Sufficient
Resources for conducting evaluation, such as time and budget	Totally lacking	1	2	3	4	5	Sufficient
Support from organizational management for conducting an evaluation	Totally lacking	1	2	3	4	5	Sufficient
The importance placed on evaluation by your department	Totally lacking	1	2	3	4	5	Sufficient
The importance placed on evaluation by your organization	Totally lacking	1	2	3	4	5	Sufficient

[send to L4 SCREENING]

L3 NON-SUCCESS CASES

Who in your organization requested that a Level 3 evaluation be conducted for a training program? Select all that apply.

- Manager of the department/workgroup which attended training
- Management above the level of the manager of the department/workgroup which attended training
- The training & development, organizational development, or human resources department
- Others - Please explain: _____

For what reasons did your department **attempt to conduct** Level 3 evaluation of training? Select all that apply.

- To assess what parts of the training were being used, and which were not so that changes to training content and methods could be made, if needed
- To help decide what parts of the organization were effectively executing important new behavioral changes (e.g., strategy, new technology), and which were not, so that remediation efforts could be implemented where needed.
- To determine whether and how much managers of trainees were supporting on-job application of learning.
- To assess the efficacy of the training with a small or “pilot” group of trainees so that changes could be made to increase impact of later and larger roll-outs.
- To help determine what factors were impeding or facilitating on-job application of learning
- To discover emerging best-practices of on-job application so that these could be communicated to other trainees and managers, and also might be built into the training content.
- To demonstrate the value of your department’s contributions to the organization
- Because it was required by the organization
- Because it just seemed like the right thing to do
- Others - Please explain: _____

Who in your organization **would have received** the results of a Level 3 evaluation? Select all that apply.

- Manager of the department/workgroup which attended training
- Management above the level of the manager of the department/workgroup which attended training
- The training & development, organizational development, or human resources department
- Others - Please explain: _____

Below is a list of factors which may affect your department's ability to conduct Level 3 evaluation of training. In your opinion, **how available** to your department is each of the following factors?

Factor	Availability						
	Totally lacking	1	2	3	4	5	Sufficient
The department's expertise/experience in using evaluative methods							
Access to employees for post-training surveys or interviews							
Access to other post-training data, such as performance measurements or employee evaluations							
Resources for conducting evaluation, such as time and budget							
Support from organizational management for conducting evaluations							
The importance placed on evaluation by your department							
The importance placed on evaluation by your organization							

[send to L4 SCREENING]

L3 NON-EVALUATORS

What were the main reasons for not attempting to conduct Level 3 evaluations of training programs? (Select all that apply)

- The department's lack of expertise/experience in using evaluative methods
- Issues with access to employees for post-training surveys or interviews
- Issues with access to other post-training data, such as performance measurements or employee evaluations
- Lack of resources available for conducting evaluation, such as time and budget
- Lack of support from organizational management
- The low importance placed on evaluation by your department
- The low importance placed on evaluation by your organization
- Other - Please explain: _____

Below is a list of factors which may affect your department's ability to conduct Level 3 evaluation of training. In your opinion, **how available** to your department is each of the following factors?

Factor	Availability						
The department's expertise/experience in using evaluative methods	Totally lacking	1	2	3	4	5	Sufficient
Access to employees for post-training surveys or interviews	Totally lacking	1	2	3	4	5	Sufficient
Access to other post-training data, such as performance measurements or employee evaluations	Totally lacking	1	2	3	4	5	Sufficient
Resources for conducting evaluation, such as time and budget	Totally lacking	1	2	3	4	5	Sufficient
Support from organizational management for conducting an evaluation	Totally lacking	1	2	3	4	5	Sufficient
The importance placed on	Totally lacking	1	2	3	4	5	Sufficient

evaluation by your department							
The importance placed on evaluation by your organization	Totally lacking	1	2	3	4	5	Sufficient

[send to L4 SCREENING]

L4 SCREENING

Which of the following best describes your department's efforts to conduct Kirkpatrick's Level 4 (Results) evaluations of training programs implemented in your organization?

- My department has completed at least one Level 4 evaluation of a training program. [\[send to L4 SUCCESS CASES\]](#)
- My department has attempted to conduct at least one Level 4 evaluation of a training program, but was unable to complete the evaluation. [\[send to L4 NON-SUCCESS CASES\]](#)
- My department has not attempted to conduct Level 4 evaluations of training programs. [\[send to L4 NON-EVALUATORS\]](#)

L4 SUCCESS CASES

Who in your organization requested that a Level 4 evaluation be conducted for a training program? Select all that apply.

- Manager of the department/workgroup which attended training
- Management above the level of the manager of the department/workgroup which attended training
- The training & development, organizational development, or human resources department
- Others - Please explain: _____

For what reasons did your department conduct Level 4 evaluation of training? Select all that apply.

- To assess what parts of the training were being used, and which were not so that changes to training content and methods could be made, if needed
- To help decide what parts of the organization were effectively executing important new behavioral changes (e.g., strategy, new technology), and which were not, so that remediation efforts could be implemented where needed.
- To determine whether and how much managers of trainees were supporting on-job application of learning.
- To assess the efficacy of the training with a small or “pilot” group of trainees so that changes could be made to increase impact of later and larger roll-outs.
- To help determine what factors were impeding or facilitating on-job application of learning
- To discover emerging best-practices of on-job application so that these could be communicated to other trainees and managers, and also might be built into the training content.
- To demonstrate the value of your department’s contributions to the organization
- Because it was required by the organization
- Because it just seemed like the right thing to do
- Others - Please explain: _____

Who in your organization received the results of a Level 4 evaluation? Select all that apply.

- Manager of the department/workgroup which attended training
- Management above the level of the manager of the department/workgroup which attended training
- The training & development, organizational development, or human resources department
- Others - Please explain: _____

Below is a list of factors which may affect your department's ability to conduct Level 4 evaluation of training. In your opinion, **how available** to your department is each of the following factors?

Factor	Availability						
The department's expertise/experience in using evaluative methods	Totally lacking	1	2	3	4	5	Sufficient
Access to employees for post-training surveys or interviews	Totally lacking	1	2	3	4	5	Sufficient
Access to other post-training data, such as performance measurements or employee evaluations	Totally lacking	1	2	3	4	5	Sufficient
Resources for conducting evaluation, such as time and budget	Totally lacking	1	2	3	4	5	Sufficient
Support from organizational management for conducting an evaluation	Totally lacking	1	2	3	4	5	Sufficient
The importance placed on evaluation by your department	Totally lacking	1	2	3	4	5	Sufficient
The importance placed on evaluation by your organization	Totally lacking	1	2	3	4	5	Sufficient

[send to FREE RESPONSE]

L4 NON-SUCCESS CASES

Who in your organization requested that a Level 4 evaluation be conducted for a training program? Select all that apply.

- Manager of the department/workgroup which attended training
- Management above the level of the manager of the department/workgroup which attended training
- The training & development, organizational development, or human resources department
- Others - Please explain: _____

For what reasons did your department **attempt to conduct** Level 4 evaluation of training? Select all that apply.

- To assess what parts of the training were being used, and which were not so that changes to training content and methods could be made, if needed
- To help decide what parts of the organization were effectively executing important new behavioral changes (e.g., strategy, new technology), and which were not, so that remediation efforts could be implemented where needed.
- To determine whether and how much managers of trainees were supporting on-job application of learning.
- To assess the efficacy of the training with a small or “pilot” group of trainees so that changes could be made to increase impact of later and larger roll-outs.
- To help determine what factors were impeding or facilitating on-job application of learning
- To discover emerging best-practices of on-job application so that these could be communicated to other trainees and managers, and also might be built into the training content.
- To demonstrate the value of your department’s contributions to the organization
- Because it was required by the organization
- Because it just seemed like the right thing to do
- Others - Please explain: _____

Who in your organization **would have received** the results of a Level 4 evaluation? Select all that apply.

- Manager of the department/workgroup which attended training
- Management above the level of the manager of the department/workgroup which attended training
- The training & development, organizational development, or human resources department
- Others - Please explain: _____

Below is a list of factors which may affect your department's ability to conduct Level 4 evaluation of training. In your opinion, **how available** to your department is each of the following factors?

Factor	Availability						
	Totally lacking	1	2	3	4	5	Sufficient
The department's expertise/experience in using evaluative methods							
Access to employees for post-training surveys or interviews							
Access to other post-training data, such as performance measurements or employee evaluations							
Resources for conducting evaluation, such as time and budget							
Support from organizational management for conducting evaluations							
The importance placed on evaluation by your department							
The importance placed on evaluation by your organization							

[send to FREE RESPONSE]

L4 NON-EVALUATORS

What were the main reasons for not attempting to conduct Level 4 evaluations of training programs? (Select all that apply)

- The department's lack of expertise/experience in using evaluative methods
- Issues with access to employees for post-training surveys or interviews
- Issues with access to other post-training data, such as performance measurements or employee evaluations
- Lack of resources available for conducting evaluation, such as time and budget
- Lack of support from organizational management
- The low importance placed on evaluation by your department
- The low importance placed on evaluation by your organization
- Other - Please explain: _____

Below is a list of factors which may affect your department's ability to conduct Level 4 evaluation of training. In your opinion, **how available** to your department is each of the following factors?

Factor	Availability						
The department's expertise/experience in using evaluative methods	Totally lacking	1	2	3	4	5	Sufficient
Access to employees for post-training surveys or interviews	Totally lacking	1	2	3	4	5	Sufficient
Access to other post-training data, such as performance measurements or employee evaluations	Totally lacking	1	2	3	4	5	Sufficient
Resources for conducting evaluation, such as time and budget	Totally lacking	1	2	3	4	5	Sufficient
Support from organizational management for conducting an evaluation	Totally lacking	1	2	3	4	5	Sufficient
The importance placed on evaluation by your department	Totally lacking	1	2	3	4	5	Sufficient
The importance placed on evaluation by your organization	Totally lacking	1	2	3	4	5	Sufficient

[send to FREE RESPONSE]

FREE RESPONSE

Below are two **OPTIONAL** questions. When you have completed your answers, click on the Continue button at the bottom of the page to go to the final section of the survey. If you prefer to not answer any of these questions, click Continue to go to the final section of the survey without entering a response to either question.

What do you think is the most critical facilitating factor that organizations should maintain in order to successfully conduct Level 3 and Level 4 evaluations, and why do you think so?

What do you think is the most obstructing barrier that organizations should reduce or remove in order to successfully conduct Level 3 and Level 4 evaluations, and why do you think so?

[send to INTERVIEW SOLICIT]

INTERVIEW SOLICIT

Thank you for participating in this research survey. As an additional component of this research project, I will conduct telephone interviews with selected survey respondents. If you are selected and agree to participate, your interview is expected to take between 15 and 30 minutes.

If you are willing to be interviewed, please click the button below.

[button text: Yes, I am willing to participate in a follow up interview]

[send to CONFIDENTIALITY STATEMENT]

If you are not willing to be interviewed, please click the button below to end the survey.

[button text: End survey now]

[goes to SURVEY END]

CONFIDENTIALITY STATEMENT

Information for Potential Interview Participants

Thank you for participating in this research survey.

As an additional component of this research project, I would like to conduct telephone interviews with selected survey respondents during the first couple of weeks of February, 2012. If you are selected and agree to participate, your interview is expected to take between 15 and 30 minutes.

You are invited to participate in the interview! Your participation is voluntary, and later you may decline to participate if you change your mind. By participating in the interview, you will no longer be anonymous; however, your identity will not be revealed to anyone but the researcher.

If you are selected as an interview participant, you will be contacted via e-mail on January 30, 2012 so we can schedule a telephone interview at a time convenient to you.

Select one:

- Yes, I agree to be contacted by the researcher to schedule a post-survey interview. The researcher may contact me at the following e-mail address: **[fill-in box for e-mail]**
- No, I do not wish to participate in a post-survey interview.

[button text: End Survey Now]
[send to SURVEY END]

SURVEY END

If you would like to be notified when the final research results are available, please enter your e-mail address in the box below. It will be stored in a separate database with no connection to the survey, so your anonymity will remain protected.

E-mail address: **[fill-in box for e-mail address]**

To exit the survey, close your browser window.

Thank you again for your participation!

Perri Kennedy
Candidate, Master of Science in Instructional & Performance Technology
Boise State University

APPENDIX F

Written Informed Consent Letter and Verbal Informed Consent Script



WRITTEN INFORMED CONSENT LETTER
Training Professionals' Understanding and Usage of Kirkpatrick's Level 3 and Level 4 Evaluations

PURPOSE AND BACKGROUND: Perri Kennedy, a graduate student in Instructional & Performance Technology at Boise State University, is conducting a research project to learn more about training professionals' experiences with performing behavior-based (Kirkpatrick Level 3) and outcome-based (Kirkpatrick Level 4) training evaluations. The information gathered will be used to better understand if and when training professionals perform such evaluations, and to learn which individual and organizational factors may impact performing the evaluations.

You have been asked to participate in this interview because you completed the initial survey conducted for this project, you volunteered to be a potential post-survey interview candidate, and you were selected as an interview candidate based on your survey responses.

PROCEDURES: If you agree to participate in this interview, you will be interviewed via telephone or Skype at a mutually convenient time. The interviews for this project will be conducted in early February of 2012. This interview will last approximately 15 to 30 minutes. You will be asked at the beginning of the interview to give verbal consent to participate in the interview. We ask that you answer all interview questions; however, you may skip any questions that you prefer to not answer.

In order to ensure accuracy, we would like to audio record the interview. However, you may choose to deny this request, in which case only written notes will be collected.

RISKS: The risks to you, personally or professionally, are minimal. The researcher will make all possible effort to maintain the confidentiality of your identity and responses. The interview will be recorded for accuracy of data collection, but you will never be addressed by name during the recorded interview to protect your identity.

BENEFITS: You will not be paid for participating in this survey. Although you will not receive any direct benefit from participation, this research may benefit training professionals who seek to be more successful in evaluating training programs.

EXTENT OF CONFIDENTIALITY: Your interview responses will be handled as confidentially as possible. Your name or other identifying information will not be used in any written reports or publications which result from this research. Only the principal researcher and co-investigators will have access to the research data. Data will be kept for three years (per federal regulations) after the study is complete and then destroyed. You will be identified in the records only by code.

PARTICIPATION IS VOLUNTARY: You may decline to participate in this interview. You may also refuse to answer any questions you do not want to answer. If you volunteer to participate in this interview, you may withdraw consent at any time; any recording of the interview to that point will be destroyed.

QUESTIONS: If you have any questions or concerns feel free to contact the principal researcher, Perri Kennedy, or her faculty advisor, Dr. Yonnie Chyung:

Perri Kennedy, graduate student
Instructional & Performance Technology
Boise State University
Phone: (301) 787-1449
Email: perrikennedy@u.boisestate.edu

Dr. Yonnie Chyung, Professor
Instructional & Performance Technology
Boise State University
(208) 426-3091
ychyung@boisestate.edu

If you have questions about your rights as a research participant, you may contact the Boise State University Institutional Review Board (IRB), which is concerned with the protection of volunteers in research projects. You may reach the board office between 8:00 AM and 5:00 PM, Monday through Friday, by calling (208) 426-5401 or by writing: Institutional Review Board, Office of Research Compliance, Boise State University, 1910 University Dr., Boise, ID 83725-1138.

VERBAL INFORMED CONSENT SCRIPT

Training Professionals' Understanding and Usage of Kirkpatrick's Level 3 and Level 4 Evaluations

My name is Perri Kennedy, and I am a graduate student in Instructional & Performance Technology at Boise State University. I am conducting a research project to learn more about training professionals' experiences with performing behavior-based and outcome-based training evaluations.

You have been asked to participate in this interview because you completed the initial survey conducted for this project, you volunteered to be a potential post-survey interview candidate, and you were selected as an interview candidate based on your survey responses.

You were e-mailed a Written Informed Consent letter prior to the interview. I will now briefly re-state the information in that letter.

This interview will last approximately 15 to 30 minutes. I ask that you answer all interview questions. However, you may skip any questions that you prefer to not answer.

In order to ensure accuracy, we would like to audio record the interview. However, you may choose to deny this request, in which case only written notes will be collected.

This interview will be recorded for accuracy of data collection. To protect your identity, I will not address you by name during the recording, and you will be identified by a code in any transcripts or other documents. I ask that during the interview, you avoid using the actual names of individuals or organizations when answering my questions.

The recording and transcripts will be retained for three years after the project is complete in order to comply with federal regulations for research. I will destroy the records when the three years are over.

The data collected for this project will be used as the basis for a Master's Thesis. I may also submit articles for publication based on this data. No identifying information will be included in the thesis or any other public discussion of this data.

You will not be paid for participating in this interview.

You may decline to participate in this interview. You may also refuse to answer any questions you do not want to answer. If you volunteer to participate in this interview, you may withdraw consent at any time; any recording of the interview to that point will be destroyed.

If you have questions about your rights as a research participant, you may contact the Boise State University Institutional Review Board at (208) 426-5401.

If you have any questions before agreeing or declining to participate, please ask them now.

1. Are you willing to voluntarily participate in this interview?
2. Do you agree to allow this interview to be audio recorded?

APPENDIX G

Interview Scripts

**Training Professionals' Understanding and Usage of Kirkpatrick's Level 3 and Level 4
Evaluations**

Interview Script for Success Cases, Non-Success Cases, and Non-Evaluators

Part 1: Informed Consent

My name is Perri Kennedy, and I am a graduate student in Instructional & Performance Technology at Boise State University. I am conducting a research project to learn more about training professionals' experiences with performing behavior-based and outcome-based training evaluations.

You have been asked to participate in this interview because you completed the initial survey conducted for this project, you volunteered to be a potential post-survey interview candidate, and you were selected as an interview candidate based on your survey responses.

You were e-mailed a Written Informed Consent letter prior to the interview. I will now briefly restate the information in that letter.

This interview will last approximately 15 to 30 minutes. In order to ensure accuracy, we would like to audio record the interview. However, you may choose to deny this request, in which case only written notes will be collected.

I ask that you answer all interview questions. However, you may skip any questions that you prefer to not answer.

This interview will be recorded for accuracy of data collection. To protect your identity, I will not address you by name during the recording, and you will be identified by a code in any transcripts or other documents. I ask that during the interview, you avoid using the actual names of individuals or organizations when answering my questions.

The recording and transcripts will be retained for three years after the project is complete in order to comply with federal regulations for research. I will destroy the records when the three years are over.

The data collected for this project will be used as the basis for a Master's Thesis. I may also submit articles for publication based on this data. No identifying information will be included in the thesis or any other public discussion of this data.

You will not be paid for participating in this interview.

You may decline to participate in this interview. You may also refuse to answer any questions you do not want to answer. If you volunteer to participate in this interview, you may withdraw consent at any time; any recording of the interview to that point will be destroyed.

If you have questions about your rights as a research participant, you may contact the Boise State University Institutional Review Board at (208) 426-5401.

If you have any questions before agreeing or declining to participate, please ask them now.

Question 1: Are you willing to voluntarily participate in this interview?

Question 2: Do you agree to allow this interview to be audio recorded?

Part 2A: Semi-structured Interview for Success Cases

I am now going to ask you to discuss your experiences with conducting training evaluations at your organization. You may also talk about your experiences at previous organizations. To better protect your privacy, please do not use the actual names of any of the organizations or individuals involved.

1. According to your survey responses, you were able to conduct a successful behavior-based or outcome-based evaluation of at least one training program at your organization. Is that correct?
2. How often does your organization conduct behavior-based or outcome-based evaluations of its training programs?
3. Who asks for the evaluations to be performed, and who receives the results of the evaluations?
4. What methods do you use for evaluating training? (i.e. surveys, performance reviews, performance metrics, interviews, and so forth)
5. What level of support do you get from the organization to conduct these evaluations, and how have you used this support to your advantage?
6. What kind of barriers have you encountered while conducting these evaluations, and how did you eliminate or bypass them?
7. How does your organization see your department's role, and do you think that conducting behavior-based/outcome-based evaluations affects how it sees you?

Part 2B: Semi-structured Interview for L3/L4 Non-Success Cases

I am now going to ask you to discuss your experiences with conducting training evaluations at your organization. You may also talk about your experiences at previous organizations. To better protect your privacy, please do not use the actual names of any of the organizations or individuals involved.

1. According to your survey responses, you were NOT able to conduct a successful behavior-based or outcome-based evaluation of at least one training program at your organization. Is that correct?
2. Why did you attempt to conduct this kind of evaluation? Did someone elsewhere in the organization ask for it, or did you/your department want to do it?
3. What level of positive support did you get from the organization to conduct these evaluations, and how did you attempt to work with this support?
6. What kind of barriers did you encounter while attempting these evaluations, and how did those barriers prevent you from completing the evaluations?
7. How does your organization see your department's role, and do you think that conducting behavior-based/outcome-based evaluations would affect how it sees that role?

Part 2C: Semi-structured Interview for Non-Evaluators

I am now going to ask you to discuss your experiences with conducting training evaluations at your organization. You may also talk about your experiences at previous organizations. To better protect your privacy, please do not use the actual names of any of the organizations or individuals involved.

1. According to your survey responses, you did NOT attempt to conduct a behavior-based or outcome-based evaluation of any training program at your organization. Is that correct?
2. Do you conduct reaction-based or knowledge-based evaluations of the training programs? How and when are they conducted?
3. What are the reasons why you/your department do not attempt to evaluate your training programs beyond knowledge-based evaluations?
4. Does another group in the organization evaluate the on-the-job results of your training programs? If so, what kind of results do they share with you?
5. How does your organization see your department's role, and do you think that conducting behavior-based/outcome-based evaluations would affect how it sees that role?

Part 3: Conclusion

That's all the questions I have. Thank you very much for your time!

APPENDIX H

Reprint Permission from John Wiley & Sons for Use of Copyrighted Figures

Campbell, Brenton - Hoboken

From: Safdar, Sheik - Hoboken
Sent: Thursday, March 22, 2012 11:29 AM
To: Campbell, Brenton - Hoboken
Cc: perrikennedy@u.boisestate.edu
Subject: FW: NON-RIGHTSLINK

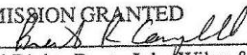
Categories: Permissions

Hi Brent,

Please find attached a thesis request for you to process.

Thanks

PERMISSION GRANTED

BY: 
 Global Rights Dept., John Wiley & Sons, Inc.

NOTE: No rights are granted to use content that appears in the work with credit to another source

Sheik Safdar | Permissions Coordinator
 Phone: 201-748-6512 | Fax: 201-748-6008 | Email: ssafdar@wiley.com John Wiley & Sons, Inc. | 111 River Street | Hoboken, NJ | 07030

-----Original Message-----

From: Perri Kennedy [<mailto:perrikennedy@u.boisestate.edu>]
Sent: Thursday, March 22, 2012 11:01 AM
To: Krouser, Gale - Hoboken
Subject: NON-RIGHTSLINK

I am a graduate student in Boise State University's Instructional & Performance Technology program, and am writing a thesis to fulfill part of the requirements for the Master of Science degree in IPT. My topic includes a discussion of the role of evaluation in the Human Performance Technology process.

I am requesting permission to reprint two illustrations to which Wiley and Sons owns the copyright:

1.

Authors: D.M. Van Tiem, J.L. Moseley, & J.C. Dessinger Book Title: Fundamentals of Performance Technology (2nd Ed.) published 2004 by the International Society for Performance Improvement
 ISBN: 1890289175
 page 3: Human Performance Technology model

2.

Author: Gilbert, Thomas
 Book Title: Human competence: engineering worthy performance (Tribute Ed.) published 2007 by John Wiley and Sons
 ISBN: 0787996157
 page 88, Table 3.4, "The Behavior Engineering Model"