DISSERTATION

LEADERSHIP PERCEPTIONS OF RESULTS AND RETURN ON INVESTMENT TRAINING EVALUATIONS

Submitted by

Kevin F. Preston

School of Education

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2010

COLORADO STATE UNIVERSITY

June 3, 2010

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY KEVIN F. PRESTON ENTITLED LEADERSHIP PERCEPTIONS OF RESULTS AND RETURN ON INVESTMENT TRAINING EVALUATIONS BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

_____
James Banning

_____
James Francis

_____
Advisor: Jerry W. Gilley

_____
Co-Advisor: Alina M. Waite

_____
Interim Director: Carole Makela

ABSTRACT OF DISSERTATION

LEADERSHIP PERCEPTIONS OF RESULTS AND RETURN ON INVESTMENT TRAINING EVALUATIONS

This study sought to validate whether the literature on high level training evaluation (level four results and level five return on investment) accurately reflected the expectations of organizational leaders regarding training evaluation reports. The researcher was interested in what high level training evaluation was being conducted at organizations and whether leaders believed the claims found in the high level training evaluation reports they were receiving.

This qualitative study used a multi-site case study method to examine the training evaluation practices of the U.S.'s ten largest Catholic healthcare organizations. The case study began with a survey sent to the lead HRD professional in each of the target organizations to understand training evaluation practices. Follow-up interviews with HRD professionals were held with six of the eight organizations that responded to the survey and who also measured level four (result) or level five (return on investment) training evaluations. An in-depth analysis of the training practices at four of those organizations, including interviews with nine senior leaders, provided the remaining data

regarding training evaluation practices and leadership perceptions of high level training evaluations.

Findings of the study indicated that while leaders did not always request high level training evaluation data, they found that data to be very valuable when given to them. Leaders wanted to see reports that indicated training program success including metrics that were important to the organization as a whole, not just to HRD professionals. Reports that included both quantitative and qualitative metrics woven into compelling stories were perceived to be most beneficial. These metrics needed to reflect organizational goals, not training department goals. Data obtained from sources seen by the organization as a whole as credible made the reports more believable. Claims of training impact provided directly by training participants were also common and were seen as credible as long as those learners were required to report their application of learning back to their own leadership teams for authentication. Findings from this study were consistent with related literature on the topic. This study provided further evidence that leaders were expecting certain high level training evaluation data to be provided in order to demonstrate training value.

Kevin F. Preston
School of Education
Colorado State University
Fort Collins, CO 80523
Summer 2010

ACKNOWLEDGEMENTS

I am deeply grateful to my friends, family, and colleagues for seeing me through the doctoral journey. Without their support, this would not have been possible.

I have been fortunate to have experienced such a positive pursuit of my educational goals at Colorado State University. My professors from the School of Education and the College of Business have stretched my mind. I have appreciated the insights and direction of my dissertation committee, which included Jerry Gilley, Alina Waite, James Francis, and James Banning. I'm particularly grateful to Jerry Gilley, who has formed me into a different learning professional than I would otherwise have been. Additionally, Alina Waite was a pillar of stability and intellect during this process.

My colleagues at the Catholic healthcare organizations included in this study were essential. Without their time to explain their processes and without their willingness to provide access to executive leaders, this study would not have been possible. Special thanks to my supervisor, Colleen Elliott, who supported my educational goals over the years and facilitated access to the organizations included in this study.

Finally, thanks to my friends and family who have encouraged my progress over the years, especially my parents, Clyde and Tillie Preston, who have always believed in me. Perhaps most importantly, thanks to my partner, Christian Bennett, for not only enduring many years of my schooling, but encouraging me to go on. It has been a long road that I could not have travelled alone.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1 – INTRODUCTION

In 2008 healthcare organizations spent $549 per employee on training according to Bersin and Associates (2009 February). In fact, American organizations as a whole spent $56.2B on training expenditures in 2008 (Bersin and Associates, 2009 February). While the cost of training is a major line item in organizations, training budgets in healthcare organizations have decreased by 11% from 2006 to 2008 due in large part to a 59% decrease in training staff (Bersin and Associates, 2009 February). This decrease in staff could be a result of continually growing efficiencies in the training industry through e-learning, which requires fewer instructors to reach larger audiences. Another possible reason for the decrease in expenditures on training is the result of an inability of Human Resource Development (HRD) professionals to prove the value of those expenditures. When HRD professionals cannot explain why their work is important, HRD departments become easy targets for budget cuts during periods of poor financial or operational performance. The belief of HRD practitioners and academics that they need to prove their value has led to substantial literature about measuring the value of training effectiveness to prove value of the HRD function.

The ever increasing scrutiny of the HRD profession by organizational leadership applies to all industry types, including healthcare. Contrary to popular belief, the healthcare industry is not immune to economic conditions that tend to decrease profits and increase expenses, which can lead to decreased training budgets. Healthcare

organizations are experiencing increased charity care, decreased elective surgeries, and decreased investment income, which lead to decreased revenue and increased expenses (Arnst, 2008). This study explores how the largest Catholic non-profit healthcare organizations demonstrate training effectiveness and how leadership perceives these claims of value by HRD professionals. To do so, I will explore training evaluation techniques and how they are used in healthcare organizations.

Statement of Research Problem

Training is a core function of all organizations because it improves employee competency and performance to meet organizational needs. Organizations are encouraged to create cultures of lifelong learning that tap into the strengths of all employees to advance strategic priorities (Senge, 1990). Some believe that intellectual capital is the most important asset for organizations and that we need to train employees not only to improve their performance, but to retain highly trained employees (Florida, 2005). HRD professionals need to respond to these changes in the environment by strategically supporting the learning needs of their organizations rather than focusing on traditional training practices that are often not aligned with strategic priorities (Gilley & Maycunich, 2000). Changes in the HRD profession have created the need for organizations to build corporate universities that formally align learning with a strategy to effectively demonstrate the value of training programs (Meister, 1998). From this has evolved the necessity for additional and better development of employees and, in turn, a need has arisen to better measure the value of those employee development activities.

Evaluation is a core aspect of training development. Kirkpatrick simply stated that "the reason for evaluating is to determine the effectiveness of a training program"

(1998, p. 3) in his training industry cult classic book, *Evaluating Training Programs*. Kirkpatrick explained four levels of evaluation: (1) reaction, (2) learning, (3) behavior, and (4) results. The Reaction level is where the HRD professional evaluates the learner's opinions about the training. For example, the HRD professional might ask if a learner actually learned anything, if the class was enjoyed, if the room was comfortable, or if the training met expectations. Reaction level evaluation is typically gathered at the end of class and reflects the short-term or immediate opinions of learners. The second level (learning) evaluates the knowledge acquisition of the learner and is usually measured by a post-test given to learners. The third level of evaluation (behavior) measures the learner's change in behavior on the job to determine if learning took place at a sufficiently accurate level. For example, if a learner took a class on new medical billing coding procedures, a level three evaluation might observe the learner doing the coding or spot check the work later to determine if the learner is correctly applying the new skills. Finally, the fourth level (results) looks at the impact of education on the organization. In the medical billing coding example, the impact expected would be a reduction in returned insurance claims or increase in payments from the insurance company. In any event, the Results level measures how the learning initiative benefited the organization in terms of revenue, cost savings, reduced errors, increased quality, increased retention, increased satisfaction.

In my experience, these four levels of evaluation are generally accepted as the norm in the corporate education field. However, many scholars believe the fourth level should be expanded to include a fifth level that some call Return on Investment (ROI). This fifth level of training evaluation simply compares the data collected in level four

(results) to the cost of the program to determine if the benefit outweighs the cost. Phillips (1999) was one of the first who pointed out the importance of comparing what is spent on a training program to what is yielded financially. Phillips (1999) indicated that measures of Return on Investment are "the ultimate level of evaluation" of a training program (p. 61), and that "return can be calculated and monitored over specified time periods, providing management with objective, investment-oriented data on the payoff of an HR effort" (p. 61). Phillips' book, *Accountability in Human Resource Management* (1999), provided a detailed approach to calculating ROI, which has spurred considerable application of the model, some of which will be explored in the literature review of this study.

Many other scholars have created methodologies and prescriptions for calculating the value of training initiatives in corporate settings. These methodologies are the application of the five levels of training evaluation ideas from Kirkpatrick and Phillips. Russ-Eft and Preskill (2001) provided a detailed account of the methodologies that HRD professionals might use to evaluate training effectiveness and impact starting with planning and moving toward data collection methods including observation, surveys, and focus group interviews. They also provided detailed sampling and analysis tips to convert raw data into meaningful measures for capturing benefit. Other authors have managed to apply a classic business financial tool, The Balanced Scorecard, to the evaluation of human resource development initiatives (Becker, Huselid, & Ulrich, 2001). The Balanced Scorecard approach attempts to provide a big picture view of a situation by measuring impact (e.g., financial), quality (e.g., operational), efficiency (e.g., time savings), and satisfaction (e.g., customer, employee, vendor) through quantitative

analysis of programs that typically are associated with subjective results. Tesoro and Tootson (2000) provided a cookbook approach to defining measurements, measuring results, and communicating results to senior leadership. Some measurements we would expect to find organizations evaluating include turnover, productivity, employee satisfaction, and profitability (Bernthal, 2005).

Expanding beyond traditional delivery of training to the changing future delivery of training, HRD professionals need to consider how to evaluate the effectiveness of e-learning in addition to traditional classroom based learning. Rosenberg (2001) suggested that e-learning should be justified with evaluation metrics that revolve around efficiency, which may include reduced time from training development to delivery, reduced time and money spent on classroom facilitation, reduced time and money spent on travel and overhead, and reduced time spent in the classroom by learners, resulting in "opportunity costs." Although most e-learning cannot be expected to reduce all of these elements, the inherent benefit of e-learning is related to convenience and reduced administration, so at least some of these elements should be represented in the training benefits. Rosenberg suggested that opportunity costs are a calculation of how much time a learner would spend in a traditional instructor-led training class opposed to how much time is spent in an online environment. This calculation is based on the assumption that online classrooms can deliver content to learners more rapidly than instructor environments through reduced class time and reduced or eliminated travel time. That reduced time of the learner can be converted to a monetary savings based on salary. The money "saved" in this situation can be multiplied by the number of learners, which can constitute a very large ROI for e-learning by enabling employees to spend their time doing more important

things than attending training. Of course, there is almost never validation that the reduced amount of classroom time is actually spent on productive activity.

Although the metrics for evaluating e-learning training may be different from other modalities, the process of training evaluation is the same. The five levels of training evaluation are used to evaluate e-learning just like they are used to evaluate instructor-led or other training methods. This study examines results of training evaluation for all types of training including e-learning and instructor-led.

Significance of Study

Clearly, there is focus on calculating the effectiveness of training in HRD practice and academia. The literature review will further explore these topics and relevant research. Ample research on why training evaluation is important and how to measure training effectiveness is available; however, there is little literature on how leadership within organizations perceives training evaluation claims. The purpose of training is to build knowledge, skills, and attitudes that support organizational needs; therefore, HRD professionals should attempt to prove the impact of training on the organization through training evaluation and ROI measures as they share data with executives who use it to make decisions about future training budgets and programs (Expertus and Training Industry, 2008). Training evaluation and ROI effectively become the data that determine the future and the survival of HRD professionals. With such an important use for this data, no wonder there is such a focus on training evaluation in the training industry literature. But do we know that leadership believes training evaluation and ROI data are accurate enough to be credible? This study will add to the training evaluation and ROI literature by assessing leadership perceptions of training evaluation and ROI results,

which will help HRD professionals better prepare result reports that are meaningful to leadership.

Research Questions

The intent of this research is to understand how leadership in healthcare organizations perceives training evaluation of organizational results and ROI presented by HRD professionals. To do so, we need to know what HRD professionals measure, and their intent for the measures as well as when those things are measured. Knowing what is measured and why will help determine if HRD professionals and organizational leadership have similar perceptions about the value and use of training evaluation metrics. This study will be guided by the following questions:

1. How do HRD professionals measure training effectiveness within healthcare organizations?
2. Why do HRD professionals choose the measures they do in determining training effectiveness?
3. Who receives training effectiveness measurement reports and for what purpose?
4. How credible does leadership find training effectiveness measurement reports that include level four (organizational results) and level five (Return on Investment – ROI) training evaluation claims?

This research is a multiple site case study that gathers leadership perceptions regarding training evaluation from several large Catholic non-profit healthcare systems. The data from these organizations will be collected in two Phases. The first phase will be to survey HRD professionals at each organization to determine what levels of evaluation they measure, who receives the results, and for what purposes. Organizations that measure level four (results) and/or level five (ROI) will be further studied in the second phase where the researcher will conduct interviews of HRD professionals to better

understand their evaluation reports. This will be followed by interviews with organizational leadership to determine perceptions of the evaluation reports.

Definitions of Terms

This research will refer to the following terms:

360 Performance Review: A process of measuring the performance of an individual by asking their supervisors, peers, and direct reports for feedback. Data from 360 performance reviews was a common metric used in high level training evaluation reports.

Asynchronous e-learning: Computer based training delivered via the Internet or CD-ROM where the learners are not required to meet at designated times. This type of e-learning is typically available 24/7 to the learner. See antonym definition for *Synchronous e-learning*.

Believability: The perceived credibility of someone/something. In this study, I will use believability as a measure of how believable training evaluation and ROI data, measures, and reports are to leadership based on criteria, such as credibility, sustainability, consistency.

Bersin and Associates: A private research firm focused on training related business research. Bersin and Associates conducts annual surveys and reports on what HRD professionals do in practice, including what levels of evaluation are commonly measured.

Formative Evaluation: Measures the effectiveness of training with the intent to improve the training for future participants. See related definition for *Summative Evaluation*.

Human Resource Development (HRD) Professional: The field of HRD includes professionals focused on training facilitation, instructional design, performance consulting, organizational development, and curriculum management. All of these specific roles use training as an intervention to address skills, knowledge, and attitudes building needs. Because all of these roles are involved in measuring the value of training, all of these roles are relevant to this study. Therefore, this study will refer to an HRD professional as anyone who creates, delivers, or consults on training as a performance intervention and measures the impact of that intervention. This study will also refer to the leader of the HRD professionals, which indicates the person that manages all of the HRD staff at an organization.

Instructional Design: The systematic process of effectively and efficiently maintaining or improving human performance (Rothwell & Kazanas, 1998). Instructional design uses a standard model to analyze, design, develop, implement, and evaluate training. Courses created through a solid instructional design process should have an evaluative component.

Leadership: The people who are the top leaders of the organization. These people are either the Chief Executive Officer (CEO) or are within a few levels of the reporting hierarchy to the CEO. Leadership is a noun referring to persons in this study, it is not a verb referring to the act of leading others. Individuals performing in the role of leadership are referred to as leaders in this study.

Learner: The participant who attends a training program.

Level 1-4 Evaluation: The foundational measurement system for determining the effectiveness of training developed by Donald Kirkpatrick (1959). The four levels include (1) reaction, (2) learning, (3) behavior, and (4) results. See Table 1.

Level 5 Evaluation: A synonymous term with Return on Investment.

Return on Investment (ROI): The comparison of how much was spent on training versus how much was gained from training. The ROI metric is typically expressed in dollars by a percentage where 100% is the breakeven point and anything over 100% is added benefit. For example, if an organization spends $100,000 on a training program and receives a benefit of $125,000, ROI is 125% (cost/benefit or $100,000/$125,000). See Table 1.

Reports: Hard or electronic copy of documents that provide evaluative information on training with varied dissemination. These documents are typically shared with organizational leadership either by e-mail or during in-person leadership meetings.

Summative Evaluation: Measures the effectiveness of training with the intent to demonstrate the value of the training program. See related definition for *Formative Evaluation*.

Synchronous e-learning: Computer based training delivered via the Internet where learners virtually meet at designated times for discussion and collaboration. This type of e-learning resembles traditional training delivery methods. See antonym definition for *Asynchronous e-learning*.

Training: The act of imparting skills, knowledge, and attitudes on people to maximize their job performance. While current literature in the HRD field suggests the word training is insufficient because of the broad scope of interventions put into place by

HRD professionals, the term training is used in this study to refer to the specific act of educating employees.

Training Expenditures: The cost of developing, facilitating, and maintaining training including training staff, technology, facilities, cost of learners being away from their job to attend training, and materials.

Value of Education: The core purpose of providing training, as most often measured via the organizational benefit of the training (Kirkpatrick's Level four). This can include performance measures, such as increased revenue, improved safety, decreased turnover, decreased injuries. This does not include transactional metrics, such as number of training classes, number of training hours, trainee satisfaction.

Delimitations

The boundaries of this study include its specific focus on the training evaluation and ROI practices and perceptions for a specific segment of the healthcare industry (non-profit Catholic healthcare organizations). Although the results of this study may be applicable to other healthcare organizations and even to other industries, this study focuses on healthcare and specifically focuses on the faith-based sector of non-profit healthcare. This narrow focus keeps this study within a manageable scope.

This study examines the believability of training from the perspective of leadership. Believability is a perception that can be influenced by a number of factors. Believability of training evaluation and ROI measures may be influenced by a leader's perceptions of an individual in the training department or by current economic strains on the organization. Perceptions can be influenced by micro and macro factors. This study attempts to uncover all factors affecting perceptions of believability.

There are two major uses of training evaluation data: formative and summative evaluation. The intent of formative evaluation is to improve training interventions while they are in a pilot stage by testing the training (Rothwell & Kazanas, 1998). The intent of summative evaluation is to measure whether training met its intended outcomes after the training is implemented (Rothwell & Kazanas). Level four and level five training evaluation is typically provided for summative purposes. With a primary interest in summative evaluation the literature review and this study focuses primarily on references to summative evaluation.

This study uses selection criteria to only include healthcare systems in Phase 2 of the research that already perform some sort of high level evaluation including, level four (organizational results) or level five (ROI). Organizations that do not perform either of these types of activities are not appropriate for Phase 2 of this study and will be noted as such during Phase 1. However, there are many organizations that do not conduct this level of training evaluation and ROI but still have to influence leadership. Although these organizations are not a part of this study, their story may be of interest in future studies as well as providing organizational background for this study.

Assumptions

This study assumes appropriate and accurate applications for measuring impact and ROI of training. This study does not examine the exact methods nor does it validate if training evaluation and ROI were correctly calculated. However, organizations responding to this study will be asked questions in order to determine appropriate application of various training evaluation methods to determine which organizations qualify for further study. This study assesses leadership perceptions about training

evaluation and ROI, not the quality of the training evaluation and ROI. Leadership may perceive valid and reliable training evaluation and ROI results negatively just as leadership may perceive invalid and unreliable training evaluation and ROI results positively. Either way, the quality and accuracy of the data calculation are not of primary relevance to this study because we are looking at the perceptions of leadership based on the information that is presented to them.

Researcher's Perspective

As an HRD professional, the topics of training evaluation and ROI are near and dear to my heart. I've worked in the training field for over 15 years. I have experienced ups and downs in the training field related to the economy and business performance and have come to believe that most leadership believes there is inherent value to employee development and training, but leadership cannot easily quantify that value. Most leaders recognize that a lack of training is detrimental to business, but this opinion tends to falter during difficult economic times. It is common for leadership not to expressly request the higher levels of training evaluation and ROI data until there is a budget shortfall or an organizational crisis that requires a closer look at expenditures that do not directly affect the bottom-line.

I've been involved in many training evaluation and ROI projects over the years and there has always been a common question regarding how the organizational impact can be directly correlated to the training. HRD professionals often indicate a strong belief that training (at least in part) has an impact on related organizational results, but there continues to be skepticism. If HRD professionals and leadership cannot agree on organizational benefit (level four), how can training ROI (level five) be calculated? I

believe this research helped examine the skepticism and nature of that skepticism regarding training evaluation and ROI from our most important stakeholders.

CHAPTER 2 – LITERATURE REVIEW

Training is big business. American organizations spent $56.2B in training in 2008 and American healthcare organizations spent $549 on average per employee in 2008 (Bersin and Associates, 2009, February). Such large expenditures on training beg the question of whether or not money was well spent. Business trends are increasing accountability for HRD professionals to demonstrate the value of their programs monetarily (Phillips & Phillips, 2007). The core purpose of a learning intervention is to help a person who lacks a skill master the new skill and then apply it to the workplace in order to improve results and add value to the organization (Brinkerhoff & Apking, 2001). Organizations want to know if they receive a return on their training investments. The question of return on investment for training has spurred a flurry of literature over the last 50 years starting with Kirkpatrick's original four levels of evaluation model published in his dissertation (Kirkpatrick, 1959). Many scholars and practitioners since have discussed how training should be evaluated and what purposes evaluation serves. The techniques have evolved into more complicated and comprehensive models over the years; however, little has been said about how to use evaluation data to garner support for training initiatives within organizations.

It is common in practitioner training literature to find articles indicating that leadership wants to see training evaluation results to validate continued existence of training departments (Bennett & Griswold, 1984; Bingham & Galagan, 2007; Moore,

2009; Naughton, 2008; Phillips, 1996; Radhakrishnan, 2008). These articles express a sense of urgency and criticality for HRD professionals to prove their value to leadership and that leadership is seeking training evaluation metrics, especially in terms of bottom-line impact. Unfortunately, there is not academic literature to substantiate this claim. Still less has been said about how leadership perceives the credibility of training evaluation data and reports. In this chapter, I will explore the literature about evaluating training effectiveness, how prevalent the practice is, and how training evaluation is used.

Evaluating the Effectiveness of Training

The process of evaluating training evolved from Kirkpatrick's (1959) original work outlining the four levels of evaluation to recommendations for the addition of a fifth level of Return on Investment (Phillips, 1999). HRD professionals follow a standard model for developing training that includes analysis, design, development, implementation, and evaluation. It is critical in this model to link the primary purpose of training identified during analysis to the measurements of how well the training achieves expected outcomes during the evaluation phase (Rothwell & Kazanas, 1998). HRD professionals around the world use this foundation to determine the effectiveness of their training programs. See Table 1 for a summary of Kirkpatrick's (1998) four levels and Phillips' (1999) fifth level.

Table 1 – Levels of Training Evaluation

Level Description	Typical Process	Measure Examples
1 – Reaction: Evaluate learners' opinions about the training	Survey learners at the end of training to gauge their satisfaction, suggested improvements, and expected learning application(s)	▪ Learner satisfaction with the content, facilities, instructor ▪ Learner intent to apply new knowledge, skills, and attitudes to their job
2 – Learning: Evaluate if learners acquire knowledge, skills, and attitudes from training	Test learners at the end of training to measure their retention of the content	▪ Examination scores ▪ Pre- and post-test scores ▪ Simulation assessments
3 – Behavior: Evaluate the extent and accuracy of how learners apply skills to the job	Observe learners in actual working environment or assess work output to determine if performance has positively changed	▪ Observation results ▪ Quality assurance checks ▪ Interview or survey data from peers who observe the work
4 – Results: Evaluate the impact of the training on the organization	Pull data from existing sources that indicate whether the underlying intent of the training was realized	▪ Financial indicators, such as sales, revenue, cost savings ▪ Quality indicators, such as error rate, accuracy rate, returns ▪ Satisfaction indicators, such as customer satisfaction, employee satisfaction, turnover rates ▪ Efficiency indicators, such as administrative time, instructor time, learner time, project duration
5 – ROI: Compare results of training to costs of providing training to determine if the effort was worth the cost	Add up training expenditures and compare them to the cost of providing the training ROI = Cost ÷ Impact	▪ ROI percentage (over 100% indicates positive ROI)

Many scholars have created methodologies and prescriptions for calculating the value of training initiatives in corporate settings for each level of training evaluation devised by Kirkpatrick and Phillips. Russ-Eft and Preskill (2001) provided a detailed account of the methodologies, sampling, and analysis tips that one might use to collect data to evaluate training, including observation, surveys, and focus group interviews. Becker, Huselid, and Ulrich (2001) have managed to apply a classic business financial tool, The Balanced Scorecard, to the evaluation of HRD initiatives for measuring the financial, customer, operational/internal business efficiency, and learning/growth. This approach provides a quantitative analysis of programs that typically are associated with subjective results. Tesoro and Tootson (2000) provided a cookbook approach to defining measurements, measuring results, and communicating results to senior leadership. The remainder of this section further explores the applications of these models to measure the value of training in scholarly and practitioner environments.

Evaluating Training at Levels One through Three

Grammatikopoulos, Papacharisis, Koustelios, Tsigilis, and Theodorakis (2004) applied the Kirkpatrick model by evaluating the effectiveness of an Olympic education training program for HRD professionals. The researchers conducted a factor analysis to develop a level one (reaction) evaluation tool that measured learner satisfaction with the program. The factor analysis indicated high internal consistency among 21 questions on the evaluation survey, which indicates that those 21 items are strong items to use for training evaluation in this situation. This study focused on level one evaluation, which suggests how to best measure learner satisfaction. The study did not progress toward measuring whether learners learned from the course (level two), applied what they

learned (level three), or if behavioral changes led to organizational results (level four), which would have made it possible to calculate ROI (level five).

There seems to be an opinion in the training industry that higher levels of evaluation (Kirkpatrick levels 3-4, behavior and results, and Phillips level 5 (ROI) are more beneficial than lower levels (Kirkpatrick levels 1-2, reaction and learning); however, the fundamental idea behind these models is that learning evolves from positive reactions (level 1) that enables learning (level 2) that leads to change in behavior on the job (level 3) that results in organizational impact (level four) that can provide a positive ROI (level five) assuming that benefits of the impact outweigh costs to deliver the training (Kirkpatrick, 1959; Phillips, 1999). If a learner enjoyed training they are more likely to have learned from that training, which makes application on the job more likely, which should result in organizational results showing a ROI. Alvarez, Salas, and Garofano (2004) found that behavioral change demonstrated in the classroom is an indicator of transfer of skills to the job. However, there are those that believe there is not a correlation between each level of the training evaluation model due to external factors, such as motivation to learn, personality characteristics, and job attitudes (Holton, 1996).

Arguably, level four (results) and level five (ROI) are the most important levels for demonstrating the value of training to leadership. Executives tend to be most concerned with bottom-line data, which these levels produce; therefore, we can say that measuring levels four and five may give HRD professionals a common language with which to communicate to leadership.

Because of the linkage between each training evaluation level, there are studies focused on what leads from learning to results. Kontoghiorghes (2001) used correlational

and multiple regression analyses to determine what factors influence learners' abilities to apply learning to the job. Pre- and post-knowledge testing, manager support, rewards for application, a social organizational culture, and a culture of continuous learning were found to enhance a learners' abilities to apply what they learned in training to the job. Although this study tells us what improves learning transfer, it does not indicate whether the training was valuable in the end. Clemenz and Weaver (2003) got closer to this concept in their research to determine what elements measure training effectiveness. They found that performance was the best measurement of training value from the perspective of the learner, but from the perspective of leadership, just because one can apply what was learned in training does not necessarily mean the training program was valuable.

Another example of how to measure the impact of training through behavior on the job was provided by Rosti and Shipper (1998) in their pre- and post-test control group experimental study seeking to determine the effect of a management training program based on subordinate perceptions of managers' skills measured before and after an intervention. The control group received instructions for interpreting the management skills report after the pre-test while the experimental group received a course on management skills. Management skills scores from the pre-test did not significantly differ among groups, which is to be expected. However, ANOVA showed 5 of the 13 managerial skills were shown to have been impacted by the training for the experimental group at the 0.10 significance level. Even though the skills assessments showed a marginally consistent improvement for the experimental group, direct reports for those managers did not report an improvement in their boss' managerial style. This may be

because employees were reluctant or unable to observe a change in their managers' performance within the timeframe of the study. The study evaluated level three (behavior) at two points in time to determine if extraneous circumstances might be influencing the results, and it was found that there were. Therefore, this study reinforces that training programs can often not be effective in their intent and that extraneous factors have an impact on what effects we can measure from training.

Evaluating Training at Level Four

Evaluating training at level four is a matter of determining whether the training impacted the organization as it was intended to do. Some measurements we would expect to find organizations evaluating training at level four (results) include turnover, productivity, employee satisfaction, and profitability (Bernthal, 2005). The metrics should be selected based on the objectives of the training to link the reason the training was created to measures of how well the training worked (Blanchard, Thacker, & Way, 2000). Most of these metrics are already being measured by organizations for various purposes so there should be little to no extra work to collect them (Dixon, 1996). When evaluating training, it is also important to align what learners believe is important with design of training (McKillip, 2001).

Some scholars suggest training evaluation results should be examined in context of the organizational environment. Alvarez, Salas, and Garofano (2004) examined a decade worth of studies on how training evaluation focused on whether courses met objectives related to training effectiveness in context of individual characteristics (e.g., motivation), training characteristics (e.g., learning objectives), and organizational characteristics (e.g., environment and culture). They found that training aligned with

organizational strategy is more likely to achieve organizational results. Even so, measuring the organizational results from a training intervention is difficult. This study did not indicate leadership perceptions to organizational results or ROI data.

Many scholars and practitioners have applied the training evaluation model over the years to include the organizational results level four. For example, DiPietro (2004) studied the effectiveness of a managerial training program provided to learners in restaurants of a fast-food company (n = 24) via three different methods, including (a) classroom training, (b) on the job training, and (c) interactive media training. There were no statistical differences among the three methods of training and the results of mystery shopper scores, customer complaints as a percentage of total customers, or customer satisfaction using ANOVA. Interestingly, all three training delivery methods were shown to have improved scores on the aforementioned three results, indicating the impact of the course on organizational results (level four evaluation). DiPietro seems to be just shy of saying that the training had a direct impact on the results, nor is management perceptions of the study results described, which is interesting considering the importance of these numbers in the restaurant industry. This is likely because this study was non-experimental and had uncontrollable and latent variables.

The studies demonstrating application of level four evaluation indicated the difficulty in identifying appropriate measures and collecting those measures. Even in cases where the right measures were found, there was hesitancy to correlate organizational results to training programs. Some studies seemed to give an impression that they were conducted for the sake of the researcher or HRD professional rather than for the sake of organizational leadership. Rarely was the use of the training evaluation

reports discussed. level five (ROI) cannot be calculated without level four (results) data. The difficulty in collecting level four data in a credible way spills over into the literature about calculating ROI (level five).

Evaluating Training at Level Five

Although the power of ROI is a common theme in the training evaluation literature, there are guidelines of when to measure ROI and when not to measure ROI. Bersin and Associates (2006) believed that ROI should not be measured for every project because it assumes that training is an investment, which is not always the case. This section will review how scholars and practitioners have measured training evaluation and ROI at various levels. Phillips and Phillips (2001) suggested that ROI may be appropriate when a training program has a long shelf-life, is aligned with organizational goals, is expensive, is visible to leadership, and has a large target audience. Although these criteria certainly eliminate the need to track some training programs that are cheap, low-key, small, etc., one would hope that most organizational training programs are well aligned with organizational strategies and visible to leadership.

Mitchell (2001) applied Phillips' ROI model (level five evaluation) by interviewing learners of training courses about their impressions of the training. Although the Phillips model was used, organizational impact and ROI could not be correlated to the training because there was no baseline data of organizational impact metrics. That is, Mitchell did not have the outcome data related to the intended objectives of the course from pre-training intervention. Furthermore, the outcomes could not be attributed to the training because there were other initiatives that potentially affected the outcomes. However, organizational outcomes did improve; but the

improvement could not be isolated as an effect of the training. Because of this, the ROI study indicated the training was a failure. Perhaps this is an inappropriate use of the Phillips model since there was no baseline data to use. The study also indicated that learners felt the training was inherently valuable for reasons outside the course objectives (e.g., camaraderie). The issue of isolating training effects from other potential influences is always troublesome when calculating ROI; however, this could be overcome by designing pre- and post-studies to ascertain the impact. Using multiple sources of baseline data, such as feedback from learners on their experiences and how they applied their learning to the job, would further support data from ROI results.

One study applied the Phillips model with special attention to isolating the effects of training on results by triangulating several data collection methods including learner surveys, manager surveys, and two distinct points of production data intended to measure organizational results (Fusch, 2001). This application of the Phillips model argued that one can isolate the effects of training through the consistency of the multiple data collection methods along with ensuring longitudinal results with the multiple data collection points. In this situation, the results were related to productivity figures collected one month and six months after training. Fusch believed the increases in productivity could be attributed to the training because there was consistency in those two data points and the beliefs of learners and managers that the training would increase productivity. Moreover, there were not any obvious external factors (e.g., new equipment, new procedures) that would affect the results. A major component of this study was comparing the benefits to the costs of providing the training in order to calculate ROI. To do this, Fusch calculated fixed administrative costs (e.g., facilities),

course costs (e.g., course development, trainer time, subject matter expert time, materials), and learner time (e.g., learner salary for time in class). This total cost was then compared to the benefit and converted to a rate of return that showed a positive impact for providing the course. The method for calculating ROI in this study was solid and logical, but there was no mention of who received these numbers or what the recipients thought of the ROI calculation.

A particularly strong quasi-experimental study examined the effect of a new sales employee orientation program on retention and revenue generation (Mattox & Jinkerson, 2005). The study focused on the cost of orienting new sales people to how orientation (or lack thereof) impacted retention on the job and therefore revenue generation opportunity based on average sales made per employee per day on the job. Factorial ANOVA indicated a significant difference among groups who attended training versus those who did not ($F = 151.29$, $p < 0.001$). Employees who attended orientation stayed with the organization 250 days (over 8 months) longer than employees who did not attend the orientation, which equated to an estimated $80.2M in additional revenue generation opportunity. ROI was calculated at 21:1 where employers received $21 in return for every $1 spent on the orientation program, which can also be said that there was a 2,100% ROI. These impressive results were shared with organizational stakeholders, but Mattox and Jinkerson did not indicate management's reaction or use of these figures.

Evaluating New Methods of Training Delivery

There is a change occurring in the Human Resources field as a whole. Gilley and Maycunich (2000) suggested that Human Resource professionals are changing from historically transactional functions to more strategic functions. This is evident in the

change of HRD professionals who historically focused on designing and delivering training without having a solid connection to the underlying intent or organizational strategy intended from training. Gilley and Maycunich suggested that HRD professionals need to become organizational development consultants positioned to identify and solve performance problems through a variety of formal and informal learning interventions. Strategic focus and thorough analysis by HRD professionals leads to appropriate interventions that address performance problems, which lead to organizational results. They suggest that HRD professionals should plan what and how to measure business impact early in the training project and use various methods for collecting information including interviews, focus groups, surveys. Those measures should then be positioned in reports for use by leadership. Connection of training impact via evaluation to organizational strategy is a key component to HRD professionals' success.

Other authors suggested that high levels of training evaluation produce leadership buy-in to training programs. Brinkerhoff and Apking (2001) presented a model for setting goals with learners and managers before training that are re-visited after training to establish results from training. This process of establishing the impact of training is then translated into a success case using additional feedback from training learners and stakeholders. The uses of success case method reports is to demonstrate training value, bolster quantitative claims about training value, market the value of training, and identify areas to improve training (Brinkerhoff, 1983). Marketing the value of training infers that success case reports are valued by leadership; however, there is not any direct indication of this. The success case method is a time consuming and expensive task for HRD

professionals. If leadership doesn't value success case method reports, HRD professionals should not undertake them.

The evolution of HRD professionals' job functions has spilled over into an evolution of methods to deliver training including formal learning, such as instructor-led training, web-based training, and tuition reimbursement for academic learning. HRD professionals are increasingly implementing informal learning, such as social networking, communities of practice, blogs, and wikis. Many of these new formal and informal ways to train employees involve new technologies, which lead to further challenges when it comes to evaluating the effectiveness of those interventions.

Clarke (2004) examined aspects of organizations' training and development systems to differentially determine the extent to which either formal or informal learning is assessed. Clarke indicated there were significant positive correlations between organizations that conduct training evaluation for informal learning based on organization size ($R^2 = 0.051$, $p < 0.05$), senior management in learning ($R^2 = 0.159$, $p < 0.001$), personal development plans ($R^2 = 0.238$, $p < 0.05$), and amount of informal learning ($R^2 = 0.238$, $p < 0.05$); however, the effect size of all of these was medium with the exception of personal development plans and amount of informal learning, which both approached a larger than typical effect size (Morgan, Leech, Gloeckner, & Barrett, 2007). The prevalence of formal learning assessment was found to be positively correlated to organizational size ($R^2 = 0.039$, $p < 0.05$), training strategy ($R^2 = 0.108$, $p < 0.01$), and availability of paid study leave ($R^2 = 0.148$, $p < 0.05$). Although these correlations are not strong, this tells us that larger organizations are more likely to assess both informal and formal learning. Organizations with training strategies and paid study leave measure

formal learning, but not necessarily informal learning. This is probably due to the fact that training strategies and paid study leaves are typically in support of formal learning. This indicates that organizations with large and comprehensive learning programs are more likely to conduct evaluation activities to measure the benefit of the training programs. However, this still does not indicate why some organizations choose to measure training effectiveness while others do not, nor does it indicate the purpose of evaluating training for the organizations that do.

Although there are alternative ways to deliver training to employees, those training methods are evaluated in the same way as traditional training methods. The four levels of evaluation apply to web-based training just as they apply to instructor-led training. However, there are additional efficiency measures that likely apply to e-learning including reduced time and money in training development and delivery, classroom facilitation, travel and overhead, and classroom time for learners, which results in "opportunity costs" (Rosenberg, 2001). These metrics, although unique to e-learning modalities, are collected in support of level five (ROI) to demonstrate how e-learning is less expensive compared to traditional training delivery methods and can be included in Balanced Scorecards (Becker, Huselid, & Ulrich, 2001). Coupling level five ROI e-learning efficiency measures with levels one through four effectiveness measures produce a powerful story for executives regarding training effectiveness and efficiency. Executives would likely be impressed when an HRD professional can demonstrate that organizational outcomes (training effectiveness) were achieved as a result of training that was delivered in a timely manner at minimum cost (training efficiency). Whether executives believe this type of claim is the question examined in this study.

Measuring the value of e-learning is the same process as measuring the value of traditional training; however, e-learning tends to measure most at levels one (reaction) and two (learning) because these levels are simple to measure with surveys and tests (Roffe, 2002). Roffe suggests that e-learning measurements should have a special focus on efficiency measures to demonstrate how e-learning saves time and money over traditional training methods. The 24/7 accessibility of asynchronous e-learning and the ability for introverted learners to actively participate in training are additional benefits of e-learning that can be evaluated (Billings, 2000). Learners tend to be more satisfied with synchronous e-learning that allows virtual collaboration with instructors and peers (Thurmond, Wambach, Connors, & Frey, 2002). Interestingly, most of the literature focused on e-learning evaluation is from the early 2000's. There seems to be a shift in the evaluation field to consider training evaluation a generic activity regardless of delivery method, which I believe is reducing the amount of literature on this specific topic.

Prevalence of Training Evaluation

A common shortcoming in the literature describing applications of training evaluation models is that HRD professionals only measure at level one (reactions). Numerous studies have been conducted over the years to determine what percentage of organizations evaluate training at each level. Table 2 shows the results of seven of these studies. These numbers tell us that it is much more common for an organization to measure reactions than results. When Expertus and Training Industry Inc. (2008) asked organizations why they measured level one, the typical response was because it is faster and easier than measuring levels two through five. These findings are substantiated by

Bersin and Associates (2006) who found that although 72% of organizations think that measuring business impact (level four) is extremely valuable, only 10% actually do. Yadapadithaya (2001) found that failure to evaluate training effectiveness at level four was a major difficulty for 14% of private organizations and 19% of public organizations. In part this may be because 94% of private, 89% of public, and 65% of multi-national organizations think creating a valid and reliable training evaluation system is a major challenge for HRD professionals (Yadapadithaya).

Table 2 – Prevalence of Training Evaluation Study Summary

Source	Organization Type	(n)	Level of Evaluation (%)				
			1	2	3	4	5
Bersin (2006)	Various	1,700+	81	35	14	10	5
Bersin (2009)	Various	350+	81	33	9	7	4
Blanchard, Thacker, & Way (2000)	Canadian Various	202	68	31	47	36	--
Brewer (2007)	Non-Profit	74	82	42	24	15	7
Expertus (2008)	Various	84	80	56	33	17	--
Jones (2008)	Illinois Hospitals	25	100	100	64	52	--
Yadapadithaya (2001)	Indian Private	109	100	79	31	8	--
	Indian Public	80	100	61	26	5	--
	Indian Multinational	26	100	92	89	65	--
		Mean:	**88**	**59**	**37**	**24**	**5**

The research indicates that not every organization is measuring every level of evaluation. The research also indicates that not every program is measured at all levels of evaluation. Scholars tend to agree that not every program requires all levels of evaluation. Phillips and Phillips (2007) indicated that HRD professionals should strive to

measure 80-100% of programs at level one (reaction), 50-60% at level two (learning), 15-25% at level three (behavior), 10% at level four (results), and 5% at level five (ROI). These recommendations indicated that 100% of organizations should be measuring ROI for 5% of their programs. Even though these recommendations suggested that not all programs should be measured at all levels, we still observe that many organizations are not conducting certain levels of evaluation at all, which is well below the recommended guidelines of Phillips and Phillips (2007).

Reasons Organizations Do Not Conduct High Level Evaluation

The small number of organizations conducting level four results evaluation (20%) and level five ROI (5%) begged the question of why organizations shy away from higher level training evaluation (see Table 2). In addition to being difficult, high level evaluation (levels 4 and 5) is often not undertaken because HRD professionals do not have the skills or time to calculate this level of evaluation (Parry, 1996). Best practice organizations had a closer connection between business and learning strategies that resulted in training program evaluation conducted with metrics that were meaningful to business units according to a study of five best practice training organizations (Dixon, 1996). Sometimes it is difficult to measure ROI for training programs that are intended to impact non-monetary values, such as government mandated training (Fusch, 2001).

When HRD professionals undertake this work, they often have to reconcile the fact that although training costs are immediate, training benefits are long-term and often not expressed in terms of dollars (Parry, 1996). Moreover, HRD professionals are fearful that negative high level training evaluation results will damage the training group through budget and staffing cuts (Parry). Best practice training organizations believed that

training was a portion of the impact on performance and that it was not important to isolate training's exact impact on the results (Dixon, 1996). Although isolating training results can be achieved through control groups, complex statistical methods, or triangulation of data collection methods, Holton (1996) cautioned that poor organizational results may indicate that training was an inappropriate intervention rather than poorly designed and delivered.

Another possible reason for the reluctance to measure high levels of evaluation may be the academic preparation of the HRD professional. Phillips (2003) found a positive correlation between higher levels of HRD professionals' academic preparation and the amount of level five evaluation conducted by those HRD professionals ($F = 4.11$, $p < .007$). Brewer (2007) replicated the question of academic preparation compared to levels of evaluation conducted but found no correlation using Kruskal-Wallis H test except for level three evaluation ($\chi^2 = 12.82$, $p < .05$). Higher levels of academic preparation were not correlated with prevalence of level four results evaluation ($\chi^2 = 0.75$, $p = .95$) or level five ROI ($\chi^2 = 2.47$, $p = .65$). There did not seem to be supporting research as to exactly why organizations are reluctant to measure training at high levels of evaluation other than the ease and speed of measuring low levels of training evaluation.

In the end, most of the literature indicated that high levels of training evaluation should be conducted selectively for programs that have the largest potential for demonstrating results, programs that cost the most, and programs that have accessible measures (Becker, Huselid, & Ulrich, 2001; Blanchard, Thacker, & Way, 2000; Dixon, 1996). The practitioner literature continually bombards HRD professionals with urgent

recommendations to conduct high level evaluation. HRD professionals need to take this advice in context of their program's situation.

Accuracy of Training Evaluation

HRD professionals often incorrectly conduct training evaluation because there is considerable confusion in the learning industry about evaluation. Plant and Ryan (1994) studied 620 large companies in Southern England to determine what types of training evaluation were conducted. The results were typical indicating that most organizations primarily measured levels one through three (reaction, learning, and behavior). When further probed about how organizations measure levels three through four (behavior and results), responses were inconsistent with expectations. For example, when asked how skills application is measured (level 3) 28% of the organizations did so by evaluating the syllabus, 36% by reviewing a business case, and 10% by using examinations. These are not appropriate ways to measure change in performance on the job because they do not directly measure job performance or output. An examination is a measure of learning (level 2), not of skills demonstration (level 3). When participants were asked to explain how they measure the benefit of training (level four), 34% did so by reviewing the syllabus with the course sponsor and 22% by conducting a skills-gap analysis of employees. These are not appropriate or accurate measures of the benefit of the training because they do not measure organizational results. Reviewing a syllabus with the course sponsor is a level one evaluation, not a level four evaluation, because it measures reaction and perception of training. Conducting a skills-gap analysis is at best a level two evaluation because it measures what is known versus what needs to be known. The findings were further validated by examining narrative responses to the survey, which

indicated respondents generally had a subjective approach to evaluation. This tells us that even when HRD professionals purport to be measuring the effectiveness of training, they often misunderstand or incorrectly identify what level of evaluation they are conducting. This may indicate inflated percentages of organizations who claim to conduct various levels of training evaluation.

This confusion continues today as demonstrated in a recent training industry research organization study about the use of new learning technologies. The report stated that learning technology initiatives were likely to pay off (Wexler et al., 2008). This claim was supported by survey data from HRD professionals who indicated they believe in the value of these new learning technologies. A HRD professional's belief about new learning technology is not an indicator of its potential value. Belief is not a measurement of organizational benefit, much less ROI. Even the learning research professionals misinterpret the levels of evaluation.

Use of Training Evaluation Data and Reports

Leadership is already concerned about the low utility of training due to poor learning talent to develop and deliver training, lack of time for learners to attend, and a belief that management training is the responsibility of the learner's manager (Longenecker & Fink, 2005). A survey of 96 Fortune 500 CEOs found that the most important learning evaluation data for executives is business impact followed by ROI (Phillips & Phillips, 2009). The amount of literature regarding how and when to evaluate training and conduct ROI calculations infers that the practice of doing so is important. However, one may wonder what should be done with these reports. Who uses them and why? Holton (1996) professed the danger of using high level evaluation results to

indicate if a training program is effective due to the issue of intervening variables. However, others believed that you can isolate the effects of training on organizational results through data triangulation (Fusch, 2001; Phillips, 1997). Even if you can isolate the effects of training, you have to find an influential audience with whom to share the results. The studies described in the previous sections about how training evaluation is conducted rarely mention how data were used to demonstrate the value of training to organizational leadership. If the data have no use, the data have no purpose.

Cousins and Leithwood (1986) examined 65 studies conducted from 1971 to 1986 that examined the use of evaluation results. They found that the two major uses could be categorized as (a) evaluation implementation and (b) decision/policy setting. This tells us that HRD professionals use training evaluation to better their programs and substantiate the value of HRD professionals to leadership while leadership uses training evaluation to make decisions about what programs are most beneficial. A significant limitation in this study was that it looked at the use of training evaluation from the perspective of HRD professionals, not from the perspective of leaders. Knowing what experts and practitioners in the HRD field think is important, but only part of the broader picture of the value of training evaluation.

Bober and Bartlett (2004) built on Cousins and Leithwood's (1986) work by conducting a qualitative study of the uses of training evaluation data in corporate universities. Bober and Bartlett conducted interviews and document analysis at four highly effective and diverse corporate universities that indicated use of either all four levels of Kirkpatrick's training evaluation model or all five levels of Phillips training evaluation model. Their goal was to understand what types of training programs are

evaluated, who uses the data, how the users use the data, which portions users find most useful in the data, and which factors (as defined by Cousins and Leithwood) are most influential in whether the data are used. The study did not focus on how the training was evaluated, but rather on how the training evaluation reports were used. Findings indicated that members of the corporate university (including HRD professionals, instructional designers, deans) typically use evaluation data in a formative way to improve training while leadership uses the data in a summative way to make decisions about program continuation and ultimately the continuation of the corporate university. Seven of twelve factors were found to be influential at all sites studied for the use of evaluation data (see Table 3). Interestingly, the most important factors were communication quality and timeliness while the factors of credibility, relevance, and findings fell to the bottom of the list. It is telling that leadership and HRD professionals find a quality presentation about a timely topic more beneficial than a credible and reliable presentation. Because HRD professionals' responses were not separated from leaderships' responses, it is difficult to determine if there is alignment between the two groups.

Table 3 – Influential Factors Ranked for Training Evaluation Data Use

Factor	Mean	Rank
Communication quality	1.25	1
Timeliness	2.75	2
Commitment or receptiveness to evaluation	3.00	3
Evaluation quality	5.00	4
Credibility	5.50	5
Relevance	5.75	6
Findings	8.25	7

The healthcare industry appears to mirror other industries in training evaluation practices and uses. Jones (2008) found that the primary purpose for training evaluation in large Illinois hospitals was program effectiveness (88%), assessment of trainee performance (44%), and determination of whether the needs of the organization were met (36%). Even though a third of HRD professionals in Illinois hospitals (n = 25) indicated the primary purpose of training evaluation to be determination of whether organizational needs were met, 52% of respondents indicated they measured level four (results). When asked why results (level four) were not measured more frequently, responses were equally divided among the following causes: too difficult, too costly, not required by organization/supervisor, or that there is a lack of training in evaluation methods. Hospitals that measure results did so to trend training program impact on organizational metrics, benchmark their training programs to others in the industry, ensure correct alignment of training programs to balanced scorecard elements, and to ensure training met organizational needs in an efficient way. When training results are measured, the results are primarily used by HRD professionals; however, it was mentioned that

stakeholders (e.g., leadership) also used the data to make decisions. Again there was no mention of how credible those stakeholders found the data to be.

Believability of Training Evaluation Data and Reports

The training literature touched on the believability of training evaluation reports, but did not clearly articulate what makes data believable. The topic of believability is critical in the computer data and business intelligence fields. Prat and Madnick (2007) defined three dimensions of believability including (a) trustworthiness of source, (b) reasonableness of data, and (c) temporality of data. Trustworthiness is established when data has a provenance or lineage to an origin people already believe is accurate. Reasonableness is established when the data is viewed as being possible and consistent with other like data. Temporality is established when there is consistency between estimated data and actual data. The learning industry could apply Prat and Madnick's definitions to high level training evaluation calculations to increase the believability of data by ensuring that (a) the data sources used to calculate the training evaluation come from a source leadership already trusts, (b) the data expresses results that are not overinflated and seem plausible, and (c) the summative evaluation data resembles original expectations and does not drastically overextend results.

The data computing literature on believability seemed to roughly correlate with some of the training evaluation literature. Phillips (2007, April) suggested that data collected for ROI purposes must be of the right quality and quantity by coming from credible sources and being compiled using only the most conservative figures in order to ensure management believability. Believability of ROI results is created by (a) executives providing the data, which eliminates arguments of the data; (b), data being

auditable; (c) conservative data collection and analysis including assuming negative results for non-response to surveys; (d) credit for organizational results being applied to numerous causes, not isolated to training alone; (e) error rate adjustments included in final reports; (f) only using partial results, not conducting longitudinal studies; (g) calculating fully loaded training program costs; and (h) inferring correlations between level one, two, three, four, and five results (Phillips, 2007, December). Summarizing evaluation results should be reinforced by explanations of the evaluation strategy, data collection and analysis processes, program costs explanation, program results summary, and recommendations (Phillips, 1999). Triangulating multiple data sources to ensure consistency and sustainability builds a case for isolating training effects, which in turn builds credibility (Fusch, 2001; Phillips, 1997). But what if a HRD practitioner does all of this and leadership still does not believe the results? This section explores the literature around leadership perceptions of training evaluation data and reports.

Figgis (2001) shared the training evaluation practices of 10 small and medium sized organizations that highly valued training and development with 19 other organizations to determine how the other organizations perceived the training evaluation practices. Both groups in the study expressed interest in understanding what value was received by training. Leadership in the 10 organizations that highly valued training were interviewed to determine their perceptions of the training evaluation data they received. Factors that increased leadership believability in the data included HRD professionals (a) explaining both positive and negative results, (b) giving concrete examples, (c) providing varied data sources, and (d) providing well-written stories. The leadership tended to comment about training evaluation stories that resonated with their own experiences or

work within the organization. Although the leaders had to be prodded to read the reports, once they did, they were positive about the results. The leaders in the second group of 19 organizations all felt that the training evaluation practices of the 10 organizations that highly valued training were of high-quality. A third of the 19 organizations indicated they planned to dramatically revise their own training evaluation programs based on the practices of the 10 best practice organizations. Another third said they did not plan to change any of their practices with the remaining third planning to adopt a fraction of the ideas. This indicates that although training evaluation results can seem positive within an organization, as they were in the ten organizations, those results are not necessarily transferrable to other organizations' practices.

Connecting organizational priorities to learning program objectives and therefore to learning evaluation is imperative to producing training evaluation reports that are positively perceived by leadership (Brinkerhoff & Apking, 2001; Gilley & Maycunich, 2000; Holton, 1996; Kirkpatrick, 1998; Phillips, 1999). If leadership does not see how training programs support achievement of organizational priorities, they will not believe or care about training evaluation results. This core tenet is the foundation of training evaluation success, but it is not enough to ensure success.

The Corporate University Xchange (2008) found that 4% of managers *ask for business impact in dollars but don't really believe the numbers*. Regardless of whether business impact is requested or not, 36% of leaders who receive this information believe the training impact measures provided to them. Table 4 summarizes the data from this study, which highlights the fact that leadership is not particularly interested or in support of the training evaluation metrics provided to them. If this is the case, one would wonder

why HRD professionals spend so much time gathering that information. Moore (2009) used the Corporate University Xchange study to reinforce the importance of training evaluation even though there is a strong presence of skepticism from leadership on the training evaluation results.

Table 4 – Leadership perceptions of Training Evaluation

% of Leaders	Perception
4	Ask for business impact in dollars but don't really believe numbers
5	Think the act of measuring to prove value is not worth the time and effort
8	Believe measures are not valuable because it's impossible to isolate other factors
14	Are satisfied with value if most people give high course evaluations (level 1)
14	Are satisfied with anecdotal evidence of value and do not press for numbers
27	Believe training programs add value and do not ask for quantifiable measures
28	Accept estimates of value if conservative estimates and a degree of confidence are used

Literature Summary

There seems to be a disconnect between what leadership believes is important and what HRD professionals are providing. Part of this is likely the result of minimal academic study about leadership perceptions of training evaluation data and reports. The literature was abundant with studies about how to measure the effectiveness of learning, but those studies rarely indicated how the training evaluation reports were received by clients and leadership. Without this critical piece of knowledge, HRD professionals do not know how and when to evaluate the effectiveness of training in an efficient manner

that meets the needs of their primary customers and leadership. This study attempts to better understand those leadership perceptions to assist HRD professionals in effective use of their time.

CHAPTER 3 – RESEARCH DESIGN

The purpose of this study was to determine how level four (results) and level five (ROI) training evaluation reports from HRD professionals were perceived by leadership within large non-profit Catholic healthcare systems. Training evaluation has long been studied in the academic environment and has been applied in practice across all industries, including Catholic healthcare. Recent literature on the topic has expressed a heightened need for level four (results) and level five (ROI) to demonstrate the value of training to leadership; however, little is known about what leadership thinks about the training evaluation reports. This study examines leadership perceptions about this topic.

Study Design

This study examined the perceptions of leaders in relationship to their experiences with training evaluation. An in-depth analysis of what is done in a bounded system at a particular time or place through multiple data sources, such as field notes, interview, observation, and survey data is known as a case study model (Creswell, 1998). This study used multiple data sources to understand how ten large Catholic healthcare organizations measure each of the five levels of training evaluation, how training evaluation reports are used, and how leadership perceives the value and believability of these reports. Because this study used a bounded system (ten organizations in the same sector of a single industry) to examine an issue (training evaluation perceptions), this study was an instrumental case study (Creswell).

The primary purpose of this study was to determine leadership perceptions of level four (results) and level five (ROI) training evaluation reports provided by HRD professionals. Although the topic of training evaluation has been studied in the recent past, the specific arena of leadership perceptions of training evaluation has not. Therefore, this study sought to examine a specific element of a familiar topic area in the HRD field. Because there was little research on leadership perceptions regarding training evaluation, a qualitative approach was selected because qualitative studies allow for extensive exploration of perceptions and situations in their natural settings (Creswell, 1998). That is, the researcher needed the flexibility of a qualitative approach to explore perceptions. This exploratory concept is not possible with quantitative research.

To examine this topic, a short survey was initially conducted to identify organizations that provide leadership with training evaluation reports. This survey was followed by interviews with HRD professionals who created level four or level five training evaluation reports. Those HRD professionals then in turn provided contact information for the researcher to interview learners who received the reports. The use of survey and interview data collection methods made this study mixed methods, although the emphasis was qualitative. A mixed methods approach enables a combination of research methods to gather all of the information needed to answer the research questions (Krathwohl, 1998). A hallmark of qualitative research is the openness of the researcher in terms of design, while quantitative research typically has a rigid outcome and process in mind (Krathwohl). This study balanced the two in that the research questions were specific and require multiple methods to answer, but a certain level of flexibility was required to extract the variance in perceptions and lived experiences related to training

evaluation. Many characteristics of qualitative research were involved in this examination of participant perspectives and their meaning. The research was performed in a natural setting using the researcher as a key instrument of data collection (Creswell, 1998, p. 16).

This study was divided into two phases. The first phase sought to determine which organizations report level four (results) and level five (ROI) training evaluation reports to leadership using appropriate methods. The correctness of those evaluation reports was not validated, rather, the researcher merely validated that the appropriate methods were used to extract the information. The second phase further examined the level four and level five training evaluation reports, their use in the organization, and the perception of leadership regarding the content of these reports. These two phases required different research methodologies.

Phase 1: Understanding Training Evaluation Practices

Phase 1 used a survey to extract data from a large population. The purpose of a sample survey is to determine how often something occurs within an organization and what the relationship is to other variables (Krathwohl, 1998). In this study, the survey identified which of the ten largest U.S. Catholic healthcare systems conducted level four (results) and level five (ROI) training evaluation within the previous two years to the survey (2007-2009). The survey also sought to determine the appropriateness of how measures were used to demonstrate value.

Phase 2: Exploring Leadership Perceptions of Evaluation Practices

Phase 2 used a qualitative foundation to understand perspectives in the organizations identified during Phase 1. The HRD professionals surveyed in Phase 1

identified the leaders at their organizations who receive and use the training evaluation reports. HRD professionals who indicated in the Phase 1 survey that level four (results) and level five (ROI) training evaluation reports were in use at their organization during the last two years were asked to provide additional information about these results. During that process, HRD professionals identified the names of one to three leaders in their organization who received training evaluation data. Those names comprised the interview list for Phase 2 of the research. Document analysis and interviews were the primary data collection methods. Interviews are appropriate methods for extrapolating detailed information from participants while being flexible to drive the interview in multiple directions (Creswell, 1998). This study required multiple directions during the interviews to talk about what leadership expects, receives, experiences, and does regarding training evaluation reports. Interview techniques provide the flexibility to address those situations.

Study Setting and Sample

A hallmark of case study research is that it is bound by a system (Creswell, 1998). As previously mentioned, training evaluation is conducted by all types of industries. In this case, the bounded system was a collection of the ten largest Catholic healthcare providers in the U.S. Healthcare was chosen because of the researcher's familiarity with the industry. A single segment of this industry was identified to enable in-depth examination of training evaluation practices and perceptions in like organizations. Case studies rely on triangulation to ensure validity. It is easier to gather multiple data sources for triangulation within a bounded system, which is why a multiple site case study was chosen for this research.

The subset of Catholic healthcare providers was chosen also to narrow the sample participants to leaders who may have had similar experiences with training evaluation. Leaders of large Catholic healthcare systems were likely to have received similar high level training evaluation data from their HRD professionals. This similarity in experiences was important to developing a deep understanding of their lived experiences. The sample was further narrowed to include only non-profit healthcare systems because the financial underpinnings of for-profit and non-profit organizations are different, which leads to different types of level four training evaluation metrics. This section describes the organizations in which this research was conducted.

The Catholic Healthcare Systems

As of 2007, there were 5,708 registered hospitals in the United States (American Hospital Association, 2007). Many of these hospitals are grouped into healthcare systems that own and operate the individual hospitals. These systems function as corporate parent organizations by providing supporting services to the member hospitals including centralized information technology, finance, supply chain, clinical engineering, and billing. Catholic healthcare systems operate 12.7% (624) of all hospitals in the U.S. and admit 15.7% of all patients (American Hospital Association). Although Catholic healthcare systems operate as single entities, there is a focus on training at each hospital within the system. Each Catholic healthcare system manages the training needs of the entire organization at varying levels of governance. Although the training topics may not be exactly the same for each healthcare system, they are typically similar in nature. For example, Catholic healthcare systems are likely to provide training on using centralized

technology, compliance, practice standards, leadership, mission and ministry, and any other topic that relates to centralization or standardization across the system.

In order to limit the sample size of this study and to ensure the case study organizations were as similar as possible, it was determined that only the ten largest Catholic healthcare systems in the U.S. would be studied. The ten largest systems were identified using data from Modern Healthcare magazine (Carolson & Galloro, 2009). These 10 systems operate 366 of the 624 (59%) Catholic hospitals in the U.S. (American Hospital Association, 2007; Carlson & Galloro). Organization size was determined by the number of beds each system had, which is a standard metric in the healthcare industry to gauge size. Table 5 lists information about each of the ten organizations included in this case study. Although differences between beds, hospitals, and revenue are broad across the ten organizations, the healthcare systems are similar in that they operate their businesses in similar fashions and provide similar training programs for topics, such as leadership development, compliance, clinical education, and new employee orientation.

Although the organizations in this study are non-profit healthcare providers, they function much like other for-profit businesses. Revenue is still important for these non-profit organizations because revenue funds expansion of community services. Culturally these organizations are unique from for-profit organizations because most leaders in non-profit healthcare have a deep passion and belief in the work they do. However, these non-profit organizations have visions, missions, and values just like any other company. Even though these non-profit organizations have a noticeable cultural difference from other for-profit organizations, they still tend to function largely the same as any other corporation in America.

Table 5 – Ten Largest U.S. Catholic Healthcare Systems by Number of Beds

System	Number of Beds	Number of Hospitals	Revenue *(In Millions)*
1. Ascension Health St. Louis, MO	15,296	67	13,489
2. Catholic Health Initiatives Denver, CO	8,267	77	8,245
3. Catholic Healthcare West San Francisco, CA	7,249	41	8,401
4. Catholic Health East Newtown Square, PA	6,371	28	4,364
5. Christus Health Irving, TX	5,463	30	3,167
6. Trinity Health Novi, MI	5,401	32	6,383
7. Providence Health & Services Seattle, WA	4,938	27	7,026
8. Catholic Healthcare Partners Cincinnati, OH	3,963	30	4,045
9. Sisters of Mercy Health System Chesterfield, MO	3,363	19	3,748
10. SSM Health Care St. Louis, MO	2,999	15	2,644

The Target Population and Sample

The population for this study was those who either created and distributed training evaluation results (Phase 1) or used evaluation results (Phase 2). Although the core purpose of this study was to examine leadership perceptions about training evaluation reports, it was necessary to first understand the reports leadership received. Because of

that, there were two main populations for this study including (a) HRD professionals and (b) organizational leaders.

Participants in this study were carefully selected from the population. Qualitative studies tend to employ purposive samples rather than random samples because the target of qualitative research tends to be focused and a random sample would likely produce unintelligible data (Miles & Huberman, 1994). Boundaries of the case study model allow the targeted selection of participants. The process of identifying the participants for this study based on their understanding and interaction with the issue at hand (training evaluation), made the sampling method purposive sampling (Krathwohl, 1998). Within-case sampling was also employed in this study because the leaders to be interviewed in Phase 2 were selected by those sampled in Phase 1 (Miles & Huberman).

HRD professionals at each of the large Catholic healthcare systems were the first population of interest and were most familiar with the creation and communication of training evaluation results within their respective organizations. The HRD professionals were typically responsible for developing, facilitating, and tracking training for technical/clinical skills as well as soft skills, such as leadership. The HRD professionals in this study tended to be the top leader of HRD professionals within their organizations. Each leader of the HRD professionals at the top ten Catholic healthcare systems belonged to a community of practice focused on sharing best practices related to employee development. The roster for this community provided the sample population for Phase 1 ($n = 10$).

The second population of interest was the leaders who used the training evaluation reports. These leaders included the top leadership roles at the participating

organizations responsible for driving the direction of their healthcare systems. This study focused on the C-Level executives including Chief Executive Officers (CEO), Chief Operating Officers (COO), Chief Human Resource Officers (CHRO), Chief Nursing Officers (CNO), Senior Vice Presidents of Mission, or other chief roles. In some systems, these top level executives had vice president titles. Regardless of the title, the target population for Phase 2 leaders included executives who either were the CEO or who reported directly to the CEO.

All leaders identified by the HRD professionals in Phase 1 were interviewed. The researcher reached saturation of data near the end of interviewing all of the leaders identified by the HRD professionals. Saturation is the process of examining qualitative data throughout data collection to determine when new concepts stop arising (Richards, 2005). The researcher developed the coding structure simultaneously to conducting the interviews, which informed when saturation was reached. Saturation was indicated when new codes stopped arising during data analysis. Coghlan and Brannick (2001) suggested that data collection should be stopped when there is a failure to yield new ideas and information. However, saturation used on its own only indicates that breadth, not depth, of data was covered where the latter provides a helpful lens for cycling between a narrow focus on things that are interesting and a wider focus on why those things are interesting (Richards).

The leadership target audience of Phase 2 of this study would be very difficult to access in other organizations because of the nature of gate-keeping around these individuals in large organizations. To manage the political relationships of conducting research within the researcher's own and sister organizations, the researcher followed

Coghlan and Brannick's (2001) recommendation of conducting a stakeholder analysis by identifying all of the leaders across the target organizations who had a stake in how this research was conducted, how the information was to be used, the communication those leaders required, and the relationship between leaders. Regular communication with these leaders was required not only to keep them informed of progress, but also to maintain their support of research activities so that participants were encouraged to fully participate throughout the study.

Managing Researcher Bias

The researcher worked in one of the participating organizations in this study. Creswell (1998) advised against conducting research within one's own organization due to potential bias. Although the researcher did not have direct authority over any of the study participants, I knew some of the interviewees, and therefore their responses, may have been more candid than if I was not familiar with the interviewee population (Krathwohl, 1998). However, the intent of the research was to add to the collective body of training evaluation knowledge and to improve the practices related to training evaluation within my own organization. To that end, this study included the researcher's organization as well as several other similar organizations to lessen the potential for bias.

The researcher encouraged organizations to participate by promoting the idea that the organization and the participants would benefit from identifying best practices to implement in the future. This dual focus on academic research and practical application is the hallmark of action research (Coghlan & Brannick, 2001). When both the researcher and the organization intend to take action on a study, the research becomes large-scale transformational change (Coghlan & Brannick). Because of this intense focus

from the organization, there is a focus on reflection of the findings and learning from both the organization and leadership. This in itself helps the researcher stay objective in the interest of shared findings that can be applied to the everyday work of the researcher.

The researcher's relationship to the subject organizations placed the researcher in close proximity to the everyday activities being examined in this study. The researcher attempted to acknowledge the different realities experienced versus those of the HRD professionals and organizational leaders. Understanding this ontological approach enabled the researcher to second-guess findings by seeking alternative reasons why something may have been true and identifying data triangulation opportunities. However, the researcher had to be willing to accept that the realities of different people vary and sometimes conflict. These conflicting realities were identified and reported during this study as much as possible.

Qualitative researchers have strong opinions about their topic areas (Richards, 2005). I admit that I felt there was too much emphasis in the learning industry on training evaluation results being imperative to the success of training departments. I questioned whether the extensive time and effort to conduct level four and level five training evaluation produced enough benefit. I also believed that HRD professionals needed to be able to explain their value to leadership in terms that leadership understands and values. Believing that the literature puts too much emphasis on training evaluation while also believing that HRD professionals need to prove their self value as well as the value of their training programs is somewhat conflicting. However, I believed that HRD professionals need to better understand what measures leadership finds valuable and believable. Understanding what leadership finds believable can help HRD professionals

produce appropriate reports to demonstrate value. As an impartial researcher my goal was to be open to the opinions supplied by leadership on the value claims provided to them.

Confidentiality and Consent

Confidentiality was of utmost importance for this study because the researcher was closely associated with the organizations studied. It was imperative to not deceive participants by ensuring they understood the general purpose of the study and how the study results were to be used (Creswell, 1998). Participants were allowed to share information "off the record" that was not transcribed, but that was sometimes used in general analysis. Participants were informed that their responses were confidential and that individual names, organization names, program names, and other identifying healthcare system demographics (e.g., number of hospitals, number of beds, revenue) would not be reported in conjunction with interview results (see Appendix A – Human Subjects Approval).

Part of this study included a survey sent to HRD professionals to identify which organizations engaged in level four (results) and level five (ROI) training evaluation. Because follow-up and clarification were needed for surveys that indicated the organizations were measuring levels four and five, the responses were not anonymous and the names of the respondents were required. The researcher guaranteed confidentiality of responses and that results would not be tied back to any specific organization or person so participants felt more comfortable providing honest data (see Appendix B – Survey Cover Letter).

Data Collection

Data collection occurred in two phases. The primary intent of Phase 1 was to identify leaders for participation in Phase 2 interviews. Phase 1 used a survey to determine which training evaluation levels were undertaken within each organization. Phase 2 used interviews to gather data about the perceptions of leadership regarding the training evaluation results. This section further describes how data were collected for both Phases.

Phase 1: Survey and Clarifications with HRD Professionals

Phase 1 of this study focused on the first three research questions:

> Research Question 1: How do HRD professionals measure training effectiveness within healthcare organizations?
>
> Research Question 2: Why do HRD professionals choose the measures they do in determining training effectiveness?
>
> Research Question 3: Who receives training effectiveness measurement reports and for what purpose?

To fulfill the core purpose of this study to interview leaders who received level four (results) or level five (ROI) data (Research Question four), it was essential to identify the leaders who received this information. A survey was used to make this identification followed by interviews with HRD professionals to better understand processes related to level four and level five training evaluation. A summary of response rates and what types of individuals were interviewed can be found in Table 6.

Table 6 – Participation Summary

Organization	Phase 1	Phase 2		
	Survey	*HR Interview*	*Business Interview*	*Total*
1	Returned – Qualified for Phase 2	1	2	3
2	Returned – Qualified for Phase 2	1	2	3
3	Returned – Qualified for Phase 2	1	1	2
4	Returned – Qualified for Phase 2	1	0	1
5	Returned – Qualified for Phase 2	Does not isolate training impact		
6	Returned – Qualified for Phase 2	No recent levels 4 or 5 evaluation		
7	Returned – Qualified for Phase 2	Chose not to participate in Phase 2		
8	Returned – Unqualified for Phase 2	Does not measure levels 4 or 5		
9	Not Returned	N/A		
10	Not Returned	N/A		

The researcher surveyed one HRD professional in each of the ten largest Catholic healthcare systems to extrapolate the levels of evaluation conducted and if appropriate methods were used for those levels of evaluation based on Plant and Ryan (1994). The survey (see Appendix C - Survey) was divided into six sections, the first being demographics and the following five asking about each of the five levels of evaluation. The last five sections asked the same questions about the prevalence of, opinion toward, process for, and use of the level of evaluation being addressed in that section. To understand the value each question contributed and how each question was used, see Table 7. Questions 3 and 4 in sections V (level four) and VI (level five) are the most important questions in the survey because these identify potential interview candidates

for Phase 2 by determining which large Catholic healthcare systems were appropriate measures for levels 4 or 5 training evaluation and how the data were being used by various positions, respectively. Although much of the data collected in the other questions and sections about training evaluation levels one through three (Sections II-IV) did not help to identify which organizations were to be interviewed in Phase 2, the data from those questions were necessary to understand the overall landscape of how training was evaluated at each organization. Additionally, HRD professionals are familiar with the five levels of evaluation and they would have been confused if the survey skipped levels one through three, which is another reason why the survey included those sections.

The internet-based survey was e-mailed to the lead HRD professional at each of the ten large Catholic healthcare organizations requesting a response within three weeks (see Appendix B – Survey Cover Letter and Appendix C – Survey). The goal was to receive one Phase 1 survey response per organization, which is why only the lead HRD professional at each organization was sent the survey. Each survey recipient was asked to forward the survey to another person if someone else was better equipped to respond. An e-mail reminder to all survey recipients was sent after week one to spur completion rates (see Appendix D – Survey Response Reminder A). Eight of the ten organizations completed the survey by the end of the second week after distributing the survey. HRD professionals from the two organizations who had not responded by the end of the second week were sent personal e-mails asking for their completed survey by the deadline (see Appendix E – Survey Response Reminder B). These two organizations (see organizations 9 and 10 in Table 6) never responded; therefore, the response rate for Phase 1 was 80% (8 of the 10 target organizations).

Table 7 – Survey Purpose Matrix

Question	Purpose	Data Use per Section					
		I	II	III	IV	V	VI
Name, phone, organization	Follow-up contact information.	C	n/a	n/a	n/a	C	C
1. For what percentage of your training do you measure…?	Identifies the prevalence of training evaluations at this level.	D	D	D	D	D	D
2. Select the response that best describes your attitude toward this level of training evaluation.	Identifies Phase opinion of the importance of this level of evaluation.	D	D	D	D	D	D
3. How do you measure this level of training evaluation?	Tests if the HRD professional is using the correct methods to measure this level of evaluation.	D	D	D	D	D I	D I
4. Who uses this level of training evaluation and for what purpose?	Identifies the recipients of the training reports as well as the HRD professional's perception of what is done with the reports.	D	D	D	D	D I	D I

Table Key:
I = This section asks for HRD professional (respondent) contact information
II = This section asks about Level 1 training evaluation (reaction)
III = This section asks about Level 2 training evaluation (learning)
IV = This section asks about Level 3 training evaluation (behavior)
V = This section asks about Level 4 training evaluation (results)
VI = This section asks about Level 5 training evaluation (ROI)
C = Use data to contact HRD professionals for follow-up in Phase 2
D = Use data to describe the population
I = Identify organizations that qualify for Phase 2

Using the data provided from the survey, the researcher identified seven of the eight responding organizations as having (a) measured level four (results) and/or level five (ROI) and (b) conducted those measures using appropriate methodologies (see organizations 1-7 in Table 6. Only one of the eight organizations did not measure either

level four or level five (See organization 8 in Table 6). Question 3 from sections II through VI was based on Plant and Ryan's (1994) research focused on the accuracy of HRD professionals' practice in training evaluation. If the response to Question 3 indicated that the HRD professional was reasonably performing level four and level five evaluation, that HRD professional was contacted for a follow-up conversation. See the data analysis section for more information on how responses to question three were judged.

Follow-up conversations were then held with six of the seven qualified HRD professionals to better understand their training evaluation reports because one of the seven qualified organizations did not respond to requests to participate further in the study (see organizations 1-6 and 7, respectively, in Table 6). Before the follow-up conversations, HRD professionals were asked to send copies of their training evaluation reports through e-mail or in hard copy format to the researcher. The researcher received five of the six organizations' high level training evaluation reports – all of which were level four results reports. Data sources from each organization involved in Phase 2 were filed together with the survey results from that organization. The researcher reviewed these reports and then interviewed the six HRD professionals via telephone to better understand how the reports were created, why they were created, and how credible HRD professionals felt the reports to be (see Appendix F – HRD Professional Interview Framework). The goal of these interviews was to better understand who received the evaluation reports and what the HRD professionals expected to be done with these reports. Another goal was to gain information to be able to assess if the leaders in Phase 2 would be good informants, which means they are willing to share their experiences and

knowledge about the topic (Krathwohl, 1998). The Informants need to be knowledgeable about the topic, but must not have an alternative agenda that may color their responses.

At the end of this process, it was determined that four of the six organizations were appropriate to move to Phase 2 of the study based on the organization's appropriate use of level four evaluation in 2007 through 2009. One of the two organizations that did not progress to Phase 2 hadn't conducted level four or level five evaluation within the past two years, so leaders would not have remembered receiving the reports (see organization 6 in Table 6). The other organization that did not progress to Phase 2 functioned under more of an organizational effectiveness methodology as opposed to a classic training methodology, which made it impossible to extract evaluation reports that were specific to training interventions (see organization 5 in Table 6). Evaluations at this organization were blended with many interventions and there was no attempt to isolate the effects of training nor were reports specific to training developed.

Phase 2: Interviews with Leaders

Phase 2 of this study focused on the core research question:

> Research Question 4: How credible does leadership find training effectiveness measurement reports that include level four (organizational results) and level five (Return on Investment – ROI) training evaluation claims?

After understanding evaluation practices at each organization in Phase 1, the next step was to interview leaders at the four organizations that conducted level four (results) and level five (ROI) training evaluation to understand their perceptions. These interviews were the most important in terms of gathering data for the core research question of leadership perceptions of training evaluation data. Leaders identified as receiving level four (results) or level five (ROI) training evaluation reports during Phase

1 were interviewed to explore their perceptions of the reports they received. Nine leaders were interviewed during Phase 2 (see Table 6). Four of these individuals led the human resources function and therefore, had a vested interest in training and were typically aware of training evaluation activities. Five of the individuals interviewed were leaders with business and operational functions including Chief Executive Officers, Chief Operating Officers, and Senior Vice Presidents of Mission.

Participating leaders were reassured of confidentiality and the purpose of the study through a consent form e-mailed to them (see Appendix G – Leadership Interview Consent). Consent was granted by the leader responding to the content e-mail and connecting the researcher with their assistant to schedule the interview appointment. Consent e-mails ensured participants understood how much time was required, how interviews would be used to gather data, that participation was voluntary, how the researcher was conducting and communicating findings of the study, how confidentiality would be maintained, and how this research would benefit their organization (Miles & Huberman, 1994).

For the telephone interview a funnel-sequenced model was used to start the conversation broadly and narrow to specific questions about perception (Krathwohl, 1998). The purpose of this narrowing was to give the interviewees the flexibility to provide information they thought was most important first and then move toward more specific elements. The interview started with questions asking broadly about leadership perceptions of training evaluation and then narrowed to more evaluative opinions about the training evaluation reports in question (see Appendix H – Leadership Interview

Framework). The interview framework provided the researcher with direction, but the questions were not sent to the interviewee ahead of time.

Because it is important to identify the purpose of each question in an interview, the interview framework groups questions into areas of intent (Richards, 2005). The preplanned and well defined fourth research question was used to create an interview structure that ensured all topics were reviewed and data compilation was easy (Krathwohl, 1998). The researcher used the structured interview to guide the conversations but also was flexible enough to probe for understanding of unexpected comments. Survey data and HRD professional interviews were reviewed in preparation for each leader's interview to customize the interview questions to that organization. This ensured the most depth and breadth was achieved in each interview. Another reason structured interview was selected was because the sample was carefully predetermined.

The interviews were conducted via telephone due to geographic diversity. This was not seen as a detriment because the organizations involved already function in virtual work environments due to their internal geographic diversity. This makes phone calls part of the norm for doing business. Given the high managerial level of interview participants and the related lack of availability of these leaders, 20 minute interviews were conducted. All nine interviews were recorded and transcribed by the researcher, totaling approximately three hours of leader interviews. Leaders were not provided with transcriptions of their interviews to validate because their responses to such a request were unlikely.

Pilot Testing

The Phase 1 survey was given to five people familiar with the HRD professionals to ensure it was clear and proper information was returned. Pilot testers were asked to identify unclear questions and provide feedback on information they would have liked to have provided but could not because there was not a question on the topic. Completion of the pilot survey was timed to determine how long the survey would take participants to complete. Expected duration was added to the survey cover letter.

Interview questions were reviewed with training evaluation subject matter experts at the researcher's organization who were then ineligible to participate in the actual study. Pilot testers determined if questions were clear and suggested additional questions to ask during the interview. Mock interviews were conducted with three pilot testers to practice the final set of interview questions in a real-life scenario. Two of the mock interviews were held over the telephone to test telephone interview techniques and recording capability. The researcher practiced paraphrasing responses to validate understanding because that technique was used during actual interviews to ensure correct capture of ideas. The interviewer also practiced a nondirective approach, which is the process of rephrasing questions to uncover the underlying perceptions of the interviewee (Krathwohl, 1998). Pilot testers were sent interview notes and the transcription to validate correct interpretation of responses. The interviewer debriefed with pilot testers for any interpretation disagreements between the interviewer notes, transcription, and the interviewees' original intent.

Data Analysis

The two-phased approach in this study divided data analysis into multiple steps. First, survey results were analyzed to determine which facilities measure training at level four (results) or level five (ROI) and therefore would be eligible to participate in Phase 2. Survey results for levels one through three training evaluation were also examined to describe the organizations' practices. Then, documents and HRD professional interview notes were analyzed to determine what training evaluation reports were provided to leadership. Finally, leadership interview transcripts were analyzed to determine leadership perceptions.

Phase 1: Survey Analysis

The surveys were short and the same questions were repeated in each of the five sections, which accelerated survey analysis. Every section, with the exception of demographics, included identical items about the prevalence and methodology for the level of evaluation being examined. Although this study focused on the use of levels four and five training evaluation data, questions were included for all levels one through five because HRD professionals are typically familiar with these levels and would find it difficult to answer a survey that skipped the first three levels. Data regarding levels one through three training evaluation were used to set the stage for the prevalence of training evaluation across the organization, but were not used to determine which facilities required follow-up interviews. See Table 7 for details on the purpose and use of each question.

The first question in each section of the survey, apart from demographics, asked for the percentage of training courses that the respective level of evaluation was

conducted. These data provided the researcher with a perspective of how prevalent the five evaluation levels were at the organization. The second question asked the HRD professional about his or her perception of importance for the evaluation level. These data provided the researcher with a perspective on why the training evaluation levels were conducted at the rates indicated in the first question. Descriptive statistics were calculated to determine prevalence and perception of importance in these organizations and to compare those findings to other studies in the literature review. Responses for these questions at each level of evaluation were used to paint a basic picture of training evaluation practices across the organizations.

The third question on the survey was intended to validate if the evaluation level was being conducted using appropriate methods according to Plant and Ryan (1994). Responses to this question were analyzed to determine the accuracy and understanding of the evaluation levels (see Table 8). The intent of doing this was not to validate the accuracy of training evaluation methods or reports, rather, the intent was to ensure the HRD professional was knowledgeable enough about training evaluation to indicate use of the most appropriate methods for measuring each level of evaluation. The researcher assumed that training evaluation conducted with appropriate methods also conducted correctly. This may or may not be the case in reality, but the researcher chose not to validate the accuracy of training evaluation reports at each organization. Each evaluation measure listed in this item was associated with an evaluation level based on Plant and Ryan. Because some measures are stronger than others, additional weighting was given to the most correct responses. Write-in responses were individually assessed. If an organization did not use at least one primary or secondary method to measure level four

or level five, it was determined not to be a candidate for Phase 2 of the study. All organizations that indicated they measured levels four and five used at least one primary or secondary method. Many of these organizations also used inappropriate methods to measure levels four and five, but this was considered acceptable because these methods were merely augmenting the primary or secondary methods.

The fourth question in each section asked about how the evaluation level data were used. Responses to this question were analyzed to describe the overall use of evaluation data. Organizations where levels four or five were measured were further analyzed to determine how the data were used. These responses identified which HRD professionals were eligible for follow-up interviews for clarification and validation on the responses. The sole intent of these follow-up interviews with HRD professionals was to better understand their level four and level five training evaluation practices, obtain copies of reports, and create a list of organizational leaders to include in Phase 2. Data for these interviews were correlated with data from Phase 2 interviews to analyze whether HRD professionals' perceptions of leadership believability in training evaluation were aligned to their own. Data analysis techniques for this are described in the Phase 2 Data Analysis section.

Table 8 – Appropriate Measures of Each Training Evaluation Level

Data Collection Method	Evaluation Level Appropriate for Use				
	1 Reaction	2 Learning	3 Behavior	4 Results	5 ROI
Course sponsor reviews syllabus	●				
Learner responds to satisfaction survey	●				
Manager observes learner on the job			●		
Learner identifies financial benefit to the organization				○	
Manager identifies financial benefit to the organization				○	
Learner takes examination or test		●			
External body accredits course	○	○	○		
Educator calculates Cost/Benefit analysis					●
Course sponsor calculates organizational metrics (e.g., revenue, safety)				●	

Table Key:
● = Strong evidence/method for this evaluation level
○ = Weak evidence/method for this evaluation level

Descriptive statistics were used to summarize the data from the survey to identify and explain the overall training evaluation practices in play across the target organizations and indicate which organizations' evaluation practices were appropriate for Phase 2 follow-up.

Validity and Reliability

Items one (percentage of use per evaluation level) and three (assessment of appropriate measures per evaluation level) of the survey in conjunction with item four (use of training evaluation reports) were used to ensure reliability of the respondents by checking to see if responses indicating "None" were consistent with those responding indicating no use of that level of evaluation, respectively. Reliability measures the consistency in responses, which in this case uses convergent evidence to assess the relationships among items (Krathwohl, 1998). While reliability measures consistency, validity ensures measuring what we are purporting to measure (Krathwohl). Phase 1 of this study ensures validity by basing item three to the Plant and Ryan (1994) study and item four to the Bober and Bartlett (2004) study. Comparing responses for these items to these studies provided construct validity, which is the overall assessment of whether the measure behaves as one would expect (Krathwohl). We expected to see consistency among results of this study and the aforementioned studies, which we did as is explained in Chapters 4 and 5.

Phase 2: Document and Interview Analyses

The core research question of this study was related to understanding how leadership perceived the training evaluation reports they receive. The qualitative data for interviews were examined using Creswell's (1998) data analysis spiral including collecting data, managing data, reading and reflecting on data, classifying data, interpreting data, and representing data. The data sources analyzed (documents and interview transcripts) were intended to build an understanding of perceptions of training evaluation at the target organizations.

The researcher transcribed the interviews to increase initial familiarity with the content. After transcription, the researcher read the leadership interviews multiple times. The first few readings focused on noting potential themes or ideas about the data thereby expanding the data. Purposive reading was conducted to identify themes and to comment on each piece of data (Richards, 2005). Each time the data were read, comments were documented for later use. Each time something was noted as "interesting" the question of why that was interesting was explored by the researcher to further analyze the data in context of conditions under which the comment was stated, the consequences of the idea, and why someone might think this (Richards). These elements of interest were related to use and believability of training evaluation.

Coding

The purpose of coding qualitative data is to generate categories, tag original data related to categories, and to assist in the analytic process (Richards, 2005). Three coding lenses were used when reading the transcripts for coding purposes including descriptive, topic, and analytic. Descriptive coding tags the data with basic demographics (i.e., who, what, where), topic coding tags the data with categories, and analytic coding begins to interpret the meaning of the data (Richards). Descriptive coding was tagged by the researcher in NVivo as cases. These case codes included data by respondent type (e.g., CEO, CNO, COO), delivery method (e.g., instructor-led training, web-based training), and topics (e.g., leadership training, clinical training). Topic coding identified purposes, uses, value, credibility, accuracy, relevancy to business, usability for decision making, and alignment with expectations of the training evaluation reports. Analytic coding required the researcher to think about passages in terms of which were interesting and

why. The three types of coding were purposefully done in three separate readings of the transcripts so as not to confuse the intent of each reading. Themes were then finalized into a list of categories for coding the data. The data were read again with the intent to classify each relevant portion of the text. See Appendix I – Coding Framework, for a list of codes collected into categories arranged by research question.

As themes emerged from the data, especially in the topic and analytic coding, relationships among codes were organized into a hierarchical catalog structure that grouped codes into themes and subthemes. The catalog structure not only shows the relationships among codes, it begins to establish patterns, provide an overall view of the data, and enables easier data searching (Creswell, 1998; Richards, 2005). To create the catalog the researcher organized categories by topics (what people said), interpretations (what the researcher believed the data represented), ideas (what categories emerged from the data), and demographics (what level and department the leader was in) (Richards). As suggested by Creswell, no more than six top-level themes were identified.

Because of the multiple coding lenses, NVivo software was used to code the data. Coding software also enables faster searching of the data to ask additional questions. After coding themes, additional questions about the data were identified. The data were then searched by code or key word to find data related to the question. These questions sometimes created new codes and themes and other times simply answered the question at hand.

The themes were then interpreted. Categorical aggregation was conducted to look for patterns related to issue-relevant meanings (Creswell, 1998). Multiple occurrences of certain perceptions indicated a prevalence of that perception. However, the researcher

challenged multiple occurrences to ask if the comments were the same and why the respondents would have the same opinion to establish congruence of perceptions (Richards, 2005). Training evaluation perceptions were examined in context of the reality in which the participants lived by examining the potential themes from different angles and with different data sources. Further analysis was completed by qualifying those themes to determine when the multiple occurrences held up and when the theme didn't apply (Miles & Huberman, 1994). Direct interpretation was used to draw meanings for single occurrences that were particularly relevant to training evaluation (Creswell). Finally, naturalistic generalizations were made to describe how this case might be applicable to other cases (Creswell). This interpretation was organized into a narrative augmented by direct quotes.

Validity and Reliability

Triangulation was a central theme to prove validity of this study through multiple data collection sources including the survey, documents, and interviews. Certain survey items were compared with interview content to validate understanding and agreement of terms, which indicated reliability of the survey, as is described in the Phase 1 Validity and Reliability section. Reliability is partially demonstrated through the coding process when themes and patterns are identified based on multiple occurrences that indicate shared opinions between interviewees. Richards (2005) suggested member checking in addition to triangulation to demonstrate validity. Member checking is the process of reviewing the initial findings with participants to determine if they see the situation in the same light. Some member checking occurred during the interviews by the interviewer validating understanding by re-stating interviewees' responses through statements like, "I

heard you say…is that right?" The interviewees in Phase 2 were unlikely to review findings of the study, so the interview member checking was conducted with the HRD professionals interviewed during Phase 1. This was done by the researcher sending the initial findings to the HRD professionals to determine if the findings seemed reasonable. Reliability is shown in part by the level of believability the study has to the target audience of the findings (which in this case were the HRD professionals), which is why member checking with the HRD professionals was possible (Richards).

Miles and Huberman (1994) suggested that findings in a study be critically examined to increase validity by ensuring the target population was representative, validating that multiple occurrences completely fit into the identified theme, and searching for outliers or conflicting data to a theme. If these methods uncover a possible threat to validity the researcher can increase the number of interviews to try and even out the data or validate that the theme or finding was correctly identified. With this in mind, coding was performed regularly during the interview schedule to constantly assess the need for more information and to identify the point of saturation.

Integrating the Data into the Case Study

This study involved collecting data about training evaluation perceptions by exploring training evaluation practices at large Catholic healthcare systems (Phase 1) and discussing leadership perceptions of the deliverables of those practices (Phase 2). Miles and Huberman (1994) suggested there are levels of analytic abstraction that start with the basic coding levels and move toward the top level of interpreting and explaining the data through the organization of themes. They suggested that hypotheses should be tested during this final level of analysis and that the data should be searched in context of these

hypotheses. This process was used by the researcher to create a descriptive picture of the situation. The picture included unique instances for individual organizations as well as similarities that apply to the overall case study. Looking at the training evaluation in this way shed light on how HRD professionals evaluate training and how believable those evaluation reports are to the leaders who receive them. Major findings from this analysis were compiled into themes related to each of the four research questions, which served as the organization foundation for Chapters 4 and 5.

CHAPTER 4 – RESULTS

The purpose of this study was to better understand what leadership thought of high level training evaluation reports given to them. While the process of evaluating the effectiveness of training is prevalent in training industry literature, an understanding of leadership perceptions of these reports was rare. Understanding these perceptions was the core intent of this study so that HRD professionals could better conduct meaningful training evaluations in the future. To understand what leadership thought of training evaluation, the researcher first had to understand exactly what reports leaders were given. Because of this, the research was divided into two phases. Phase 1 focused on how training was evaluated, why certain things were measured, and who received the final reports. Phase 1 was driven by the first three research questions:

> Research Question 1: How do HRD professionals measure training effectiveness within healthcare organizations?
>
> Research Question 2: Why do HRD professionals choose the measures they do in determining training effectiveness?
>
> Research Question 3: Who receives training effectiveness measurement reports and for what purpose?

To answer these questions, HRD professionals from the U.S.'s ten largest Catholic healthcare providers were surveyed to better understand their training evaluation practices. The survey was followed by HRD professional interviews to uncover the nuances of training evaluation practices within each participating organization. Once training evaluation practices were understood, the researcher moved to Phase 2, which

focused on the perceptions of these leaders in interviews. Phase 2 was guided by the fourth, and most important, research question:

> Research Question 4: How credible does leadership find training effectiveness measurement reports that include level four (organizational results) and level five (Return on Investment – ROI) training evaluation claims?

The researcher used survey, document analysis, and interview techniques to answer the four research questions. This multiple site case study sought to understand the training evaluation practices and leadership perceptions of several like organizations so that a better overall understanding of this situation could be gathered. The findings of this research are organized within the four research questions.

Research Question 1: How Training is Evaluated

The first part of this research involved gaining a better understanding of exactly how training is evaluated in organizations. A survey was sent to one HRD professional at each of the ten largest Catholic healthcare organizations in the U.S. The survey asked HRD professionals to indicate what levels of training evaluation they measured, how often they measured each level, and what they did with that information. HRD professionals who responded to the survey were contacted for follow-up clarification interviews. Several findings related to methods, prevalence, and accuracy were uncovered during the survey and follow-up conversations with HRD professionals. These findings were further explored during interviews with leaders at each organization. While discussing methods to evaluate training, HRD professionals and leaders continually expressed an interest in correlating training impact to the organizational results those programs were intended to support. This theme came up continually while discussing how training evaluation is measured in these organizations.

Prevalence of Training Evaluation

The eight organizations that participated in Phase 1 were asked to indicate the percentage of the training programs they measure at each level of evaluation. See Table 9 for a summary of these percentages. This table also indicates the recommended percentage of training that should be measured at each level with an asterisk (*) based on Phillips and Phillips (2007). The prevalence of measuring each level of training evaluation is largely consistent with the Phillips and Phillips (2007) recommendations because the majority of responses fell within the recommended levels as you can see by the highest frequencies being in the same percentage as recommended by Phillips and Phillips. Although the majority of organizations in this study indicated they measured levels four and five at the recommended frequency, there were several who indicated that they did not measure at these levels at all. In fact, only 75% of organizations in this study measured level four and only 63% measured level five, which is inconsistent with the Phillips and Phillips (2007) recommendation that suggests that all organizations should measure at least some programs at these levels. There were two organizations in this study that indicated they did not measure level four at all and three organizations that did not measure level five at all.

Table 9 – Frequency of Measuring Each Level of Evaluation

Percentage of Training Measured at this Level	Level 1 *Reaction*	Level 2 *Learning*	Level 3 *Behavior*	Level 4 *Results*	Level 5 *ROI*
100%	2 *	0	0	0	0
75-99%	4 *	3	1	1	0
50-74%	1	4 *	2	0	0
25-49%	0	1	2	1	1
1-24%	0	0	3 *	4 *	4 *
None	1	0	0	2	3
Total percentage of organizations that measure this level from this study	***88%***	***100%***	***100%***	***75%***	***63%***
Total percentage of organizations that measure this level from past studies	***88%***	***59%***	***37%***	***24%***	***5%***

* Recommended percentage of training to evaluate at this level (Phillips & Phillips, 2007)

Although Phillips and Phillips (2007) recommended that 100% of organizations should measure at least some of their training programs at each level, other studies on the prevalence of training evaluation in organizations indicated that the actual prevalence is much less than those recommended levels (see Table 2 from Chapter 2). When examining what percentage of organizations in this study measured training at each level, these findings were sporadically consistent with those in the other studies. The last row of Table 9 summarizes the average prevalence of each level of training evaluation in organizations based on the seven studies examined in Chapter 2. Although the prevalence of level one evaluation in this study was consistent with the other seven studies, the higher level evaluations were inconsistent when compared to the other studies

examined in Chapter 2. The organizations in this study were much more likely to conduct levels four and five evaluation than seemed to be the norm based on the seven studies cited in Table 2.

Even though the organizations in this study were more likely to conduct high levels of training evaluation, they still did not reach recommended frequency for each level (as is described above). This is interesting because although only 75% of organizations measured level four (results) for any training programs, all of the organizations indicated that level four training evaluations were important. In fact, the two organizations who indicated that they did not conduct level four evaluations, indicated they felt level four evaluations were *important* or *really important.* Of even greater interest is that all three of the organizations who indicated that they did not conduct level five evaluations also indicated level five evaluations were *absolutely critical*. One would wonder why something that someone felt was absolutely critical was not done.

Reasons Organizations Didn't Conduct High Level Evaluations

One interest area of the researcher was to understand why organizations seemed to shy away from conducting high level training evaluation. The literature suggested that organizations did not conduct high level training evaluation because it was difficult, time consuming, and that HRD professionals did not have the skills to do so. Concerns about having the knowledge and skills to conduct high level training evaluation were common in the HRD and leader responses of this study. One leader interviewed, who was an HR executive, said,

> We are learning disabled in both of these ways of determining ROI. We don't have adequate understanding and methodologies to prove it and we don't have the resources to do the work to prove it.

Another leader said, "Development looks at the multiple domains of leadership, which is harder to measure." The concept of difficulty stemmed from a lack of competence in this area. Both HRD professionals and leaders felt that there was a lack of understanding related to conducting levels four and five evaluations, which was a cause for those evaluations not being conducted.

Although a lack of competence was a common element as to why high levels of evaluation were not conducted more frequently, there were also multiple comments related to the time and effort required to do this. One leader who was familiar with high level training evaluation expressed concerned about the competence and time required to conduct high levels of training evaluation by saying,

> We don't do a lot of level four. It's unusual that we're able to demonstrate what people say it is. We don't have vehicles to help us prove level four. We don't have the resources to go out and do this.

Another leader said, "It takes a lot of work to run the statistics and create the metrics." It was clear that in addition to a lack of competency on high level training evaluation, HRD professionals also did not have the time it takes to conduct this level of evaluation.

Return on Investment

Although five of the eight organizations who responded to the Phase 1 survey indicated they conducted level five (ROI) training evaluation, no one offered ROI data or reports as part of their response to Phase 2. One HRD professional said, "I couldn't conduct a return on investment study if I wanted to because I don't know how much this training program costs in the first place." Organizations were given leeway to provide

the researcher with any level four or five training evaluation report they had, but all organizations who participated in Phase 2 provided level four (results) reports, not level five (ROI) reports. It is possible that these organizations either did not understand what ROI evaluations were or that they were not confident in ROI reports they had and therefore, chose not to share this type of information.

Methods Used to Evaluate Training

HRD professional follow-up interviews were focused on how organizations were measuring levels four and five evaluation. HRD professionals conducted high level training evaluation based on the core objectives of a training program. Once they understood what was being achieved, they gathered data to support the training's contribution to achieving those organizational objectives. HRD professionals were asked to provide samples of the output of these training evaluation projects including the level four and level five training evaluation reports they created.

These reports were typically PowerPoint presentations provided to various stakeholders, including leadership. A few of the reports provided were Word documents designed as handouts or pamphlets. The reports ranged in size from 6 to 99 pages with an average length of 30 pages. The reports generally included activity data about the program (e.g., number of learners complete, number of learners registered, number of hours complete) in addition to evaluative data. However, only some of this outcome data would be appropriate representations for levels four and five training evaluation reports. Most reports included level one evaluation data (learner satisfaction). No report included level two evaluation data (learner learning measured by testing). Most reports included both level three (behavior measured by observation) or level four (results measured by

outcomes). All reports at least included level four data since this was the criteria for inclusion in Phase 2. As mentioned earlier, no report included level five (ROI) data.

Comparing High and Low Performers

One of Phillips' (2007) recommendations was to compare high and low performers in order to better understand what performance is sought. Several interview participants suggested comparing the performance of high and low performers before and after training to gain this understanding. One HRD professional said,

> You use quality indicators to look at how a manager does across the board. Then we can compare high and low performers and what the low performers are not getting.

The concept of comparing high and low performers to determine how to develop low performers into high performers is also recommended during the instructional design process (Rothwell & Kazanas, 1998). This was an important aspect of gathering data before and after a training event because it tells HRD professionals more about what behavior they should expect to see, which is a prerequisite to effective observations.

Observations

Several HRD professionals indicated that firsthand observations of learners applying the skills they learned were critical to demonstrating the value of any kind of training. This inferred a strong need to observe a change on the job that should be the result of acquiring new skills. Comments related to the need for observations included,

> If you wanted to look at basic kinds of training for let's say, clinical topics, you can test whether people are doing something better through observation, time motion studies, errors. You have to be at the moment observing them.

> One context of looking at it is after you've been exposed to the learning environment, what's the difference back on the job? And then, in actual practice, what's the difference?

> The most important part [of leadership development] is to see leadership performing better than they did in the past. Those leaders see the change in themselves, which is just as important as the outcome data collected.

These comments all had a consistent message that we must be able to identify the change we expect of the learner and then see if they actually act in that way back on the job.

Pre- and Post-Measures

Both HRD professionals and leaders mentioned the need to collect pre- and post-information in order to understand if training had an impact. One leader suggested that two things would make training evaluation reports more beneficial by saying,

> One, performance evaluations before and after program participation. Two, the results of interviews with their supervisors about whether the person is a better or worse leader after attending the program.

Both of these recommendations require a comparison of learner performance before and after participation in the training. Before and after measures require that baseline data is available. Unfortunately, baseline data is often not collected, as evidenced in this comment from a leader,

> What are the skills of managers? What is their overall performance level? Are they better now than they were a year or two ago? We don't have a single number we can use to compare 2 years ago to now, so we probably could do a better job trying to identify that.

This infers that leaders want to see an improvement of performance, but without baseline data, that is impossible.

Impact Map and Success Case Method

Three of the four organizations who participated in Phase 2 of the research used Impact Mapping (Brinkerhoff & Apking, 2001) and Success Case Method (Brinkerhoff, 1983) to measure the value of training and report that value back out to the organization. These models seem to be particularly of interest to the organizations in this study,

possibly due to the high level of collaboration between Catholic healthcare providers that facilitates the sharing of best practices. There were mixed leader responses regarding the viability of these reports. One leader said,

> The Success Case didn't help to determine what to do next. The report was irrelevant to how we ended up changing the program.

However, there were other leaders who believed in the concept of Impact Mapping and Success Case Method by indicating there wouldn't be any other way to really know what the value of a program was. Leaders tended to believe that data examined in Success Case Methods was more believable when the learners had to present it back to their own leaders. This is discussed more in the findings and discussion for research question four.

Accuracy of Methods Used to Evaluate Training

The survey not only asked what levels of training evaluation were measured, but also how those levels were measured. The intent of asking this was to determine if organizations were using the correct methods to measure each level of evaluation. Plant and Ryan's (1994) study was the basis for determining the accuracy of training evaluation methods. Using the indicators of accurate training evaluation methods from Table 8, the researcher found that most organizations were using appropriate methods to measure each level. An organization was deemed to be using appropriate methods to measure a level of training evaluation if they used one or more of the methods indicated as "strong evidence/practice" from Table 8. A summary of how many organizations used appropriate methods for each level of training evaluation can be found in Table 10. All six organizations who measured level four training evaluation did so using organizational metrics (strong evidence) in addition to a variety of other methods. All five organizations who measured level five training evaluation did so using cost/benefit analysis (strong

evidence) in addition to a variety of other methods. In general, HRD professionals used appropriate methods to measure each training evaluation level, especially when it came to level four and level five training evaluation.

Table 10 – Using Appropriate Training Evaluation Methods Summary

Appropriateness of Method	**Level 1** ***Reaction***	**Level 2** ***Learning***	**Level 3** ***Behavior***	**Level 4** ***Results***	**Level 5** ***ROI***
Used appropriate methods	7	7	7	6	5
Used inappropriate methods	0	1	1	0	0
Total percentage of organizations using appropriate methods	***100%***	***88%***	***88%***	***100%***	***100%***

Measuring the Effectiveness of Leadership Development

Although Phase 1 indicated that HRD professionals should think about all training evaluation conducted during the past two years, the four organizations that participated in Phase 2 all provided leadership development examples. The researcher asked HRD professionals to explain how they measured level four and level five training evaluation during the follow-up interviews to the survey in Phase 1. During these conversations, HRD professionals focused on what they did for leadership development. The researcher did not specifically ask HRD professionals to focus on a single topic of learning, but they all chose to focus on the same topic (leadership development) in the end.

Research Question 2: Why Measures are Chosen

After understanding how organizations conducted level four and level five training evaluation, the researcher focused on exactly what measures were used. The metrics used by each of the organizations who participated in Phase 2 of the study were similar, due in large part to the learning topic of each of those studied being leadership development. The metrics identified during this study were either objective or subjective, which is how this section is outlined. There were also comments from leaders during Phase 2 interviews that indicated some metrics were less meaningful than others. Those metrics were discussed in the last part of this section.

Objective Metrics

All level four training evaluation reports included objective metrics that were generally pulled from a third party source outside of the training department. These outside departments included HR, Finance, Operations, and Quality departments. These objective metrics were typically quantitative and including such things as:

- Performance appraisal data including 360 reviews, upward appraisals, and annual performance reviews
- Financial metrics including net patient revenue, total revenue, census, and cost savings
- Voluntary turnover
- Internal promotion rates
- Internal transfer rates
- Clinical quality metrics including error rate, infection rate, and through-put
- Employee interest in training program participation
- Employee satisfaction
- Physician satisfaction

- Patient satisfaction

The most common metric referred to by leaders was performance review data including 360 performance review appraisals, which are performance reviews conducted not only by the employees' supervisors, but also by the employees' peers and direct reports. Leadership development can be a very subjective subject, so it was not surprising that so much emphasis was put on 360 reviews because they attempt to quantify human performance. Leaders used aggregate 360 data to measure training effectiveness as evidenced by these leader remarks,

> We look at 360 scores to determine who's good and who needs improvement. We expect to see higher performance from those who attend our leadership development program.

> The most powerful connection is when you connect the 360 assessment with the individual development plan. This lets the employee drive their development based on their needs.

Leaders in this study expressed a strong interest in seeing specific, quantitative, performance-based data to support the effectiveness of learning. Some leaders didn't know exactly what that meant, but knew they wanted this type of data. One leader said,

> I really look to any performance oriented data. Anything that will tell me their performance as an employee of the organization before and after training would be pertinent and relevant. I'm at a loss about what that might be, but that is what is important to me.

This highlights the fact that leaders wanted to see measurable and objective data, even for training topics that may have been more subjective in nature. However, it was unclear to leaders what those data elements were. Leaders had a difficult time articulating exactly what they wanted to see. Leaders were more easily able to explain what they saw but did not like, opposed to explaining what they wanted to see in the future.

Subjective Metrics

HRD professionals and leaders seemed to agree that quantitative figures were not enough to paint a picture of training effectiveness. Telling a meaningful story was just as impactful as providing financial savings. In fact, there seemed to be distrust of quantitative data without accompanying qualitative stories. These subjective metrics fell into two categories: anecdotal stories and self-identified data.

Anecdotal Stories

When leaders were asked how they knew that training was effective, they tended to pause before answering. This question was asked before any questions specific to the training evaluation report provided by the HRD professional in Phase 1. The intent of asking this question was to understand the leaders' gut feelings for determining the value of training. One leader's response to how he knew training was working was,

> [I know training is working] anecdotally. But also in terms of training that I've been engaged in through asking questions [of learners] if this training session, this experience, has prepared them to be more effective.

This leader clearly heavily relied on his own experience talking with learners of the training program. This sentiment was held by many leaders. Another leader said, "[my belief that training is valuable] comes more from personal interaction and observation as opposed to getting a report that tells me so." A third leader said:

> My comments are not based on this report. My comments are based on interactions with former program graduates.

These leaders all saw the value of training through their own lens. They all had experiences with learners and others in the organization that influenced the way they saw training's effectiveness.

Self-Identified Metrics

All level four evaluation methods discussed in this study included (at least in part) data that was self-identified by learners. These data sometimes came from after class evaluations, post-surveys, interviews with past learners, or learner report-outs. One leader simply said, "we ask what they're going to do" when asked how the leader knew training was effective. It was not always evident that this information was self-identified because the leaders tended to believe in this data, as can be evidenced in this leader statement,

> The greatest majority of the comments that we received was that they finally now understand how to apply the technical and administrative skills in the context of ministry. Why it's important to do that, not just the how [it's important to do that]. To me, that's a breakthrough.

This leader put a great deal of faith in the learner's claim of applying what they learned to their job. Another leader speaking about the credibility of self-identified data said, "I can trust that people we're honest with us." There was a common sentiment with leaders that the learners were one reliable source of information for determining the effectiveness of training.

Even organizations that used best practice models like Success Case Methods relied heavily on learner self-identified data. One leader said,

> We ask participants to tell us what they like. But the Success Case analysis was more about what they did with what they learned. Once they said what was meaningful, then the question was, what did you do with it, how did you apply it, what were the results?

The Success Case method was almost entirely based on learner self-identified data. While self-identified data were considered reliable by most leaders, there were a few who were skeptical of data accuracy. One leader said,

> When you really question the participants if they're feeling better at this or feeling better at that, that's all good and fine, but it doesn't really tell me if they are better. It tells me what they feel.

Clearly this leader did not think that self-identified data could be relied upon on its own. Leader perceptions about the reliability of learner identified data is explored more in research question four.

The interview results also indicated overlap between the theme of self-identified data and the best practices explored in the methods section of the findings related to research question one. Specifically, several organizations used self-identified data to measure pre- and post-performance. One leader explained how they measure this data by saying,

> A lot of the evidence is anecdotal. For example, seeing people perform at higher levels or having more confidence in public speaking. We also get reports from corporate headquarters that help us see improvements that are occurring. Some of this is self-assessment by the individual ranking their skills before and after a program.

HRD professionals and leaders were both comfortable with the idea of learners self identifying their performance before and after training. However, self-identified data along these lines was not the preference. Again, this is discussed more in research question four.

One major type of self-identified data mentioned during the study was learner satisfaction with the training. Leaders tended to continually use learner interest in participating in the training as an indicator of program success. One leader said, “the learners were jazzed about it” in context of describing program success. Another leader said, “we ask participants to tell us what they like.” Another leader indicated that learner

satisfaction was an indicator that the program did not need to be changed. Leaders put a lot of credence on what learners thought about the program.

Less Meaningful Metrics

Leaders also indicated which metrics meant less to them than others. Some metrics were found to have both positive and negative comments from different leaders. Those varied metrics are described in the objective and subjective sections above. There were also some metrics that received universally negative responses in terms of demonstrating the value of a training program. These metrics were based on training program activity and efficiency, such as the number of people who went through training, learner hours, instructor hours, and e-learning course usage. Some leaders specifically called out these metrics as being unhelpful. Other leaders did not mention them at all even though the metrics were called out in the reports the leader received. These training activity and efficiency metrics were not included in all training evaluation reports from Phase 2, but where they were included the leaders either indicated the metrics were less valuable or did not mention the metrics at all.

Research Question 3: Who Receives Reports and Why

The researcher asked both HRD professionals and leaders who they thought cared about training evaluation reports and why. The intent was to better understand who was requesting this data. The responses to these questions from the HRD professionals informed the researcher of who the leaders should be in Phase 2 of the study. The responses to these questions from the leaders in Phase 2 informed the researcher of what leaders really felt about these topics. These responses were strong lead-ins to research question four.

Training Evaluation Data Requesters

Almost all of the HRD professionals indicated that training evaluation reports were given to senior leadership. Senior leadership definitions varied between the organizations, but included groups such as the CEO and his or her executive team, the board of trustees, senior HR executives, hospital CEOs, and sometimes down to Director level leaders. However, only half of the HRD professionals indicated that they were asked to provide this data. One such HRD professional said,

> The first senior leader who asked for this is no longer with the organization. This person told me one of the most critical things I needed to face was to find a way to demonstrate the value of this kind of work. He said he understood it intuitively, he gets it, understands it, and will fight for it, but that he wouldn't be here forever. When I got hired, my boss told me the same thing and another senior leader was saying the same thing to a partner of mine. It became clear that we needed to find a way to do the ROI.

This response indicated an inconsistency in the drive to provide this data. The HRD professional was guided toward collecting the data, but did not seem to be held accountable for doing so. This unclear direction could be a reason more HRD professionals do not collect level four and level five training evaluation data in the first place.

Another HRD professional indicated that training evaluation requests were dependant on the topic,

> We definitely get questions from leaders in the field about the value of programs and whether they should continue. The questions only come on the more selective topics like leadership development, not on regulatory or compliance topics.

Although not all HRD professionals indicated such a strong desire for this type of work from leadership, they all indicated that they thought providing the data was critical in order to demonstrate their value. Clearly, there was a deeper desire to collect this

information for topics that were perceived to be of greater importance and expense. This was evident both in the comment above and in the fact that all HRD professionals participating in Phase 2 of this study chose to submit a leadership training example instead of another topic area. This may have resulted from the fact that leadership training was more expensive to provide than other topics and was more visible to others in the organization.

Other HRD professionals indicated that while they were not asked to demonstrate the value of training, they were asked to provide other evaluative data. One HRD professional said,

> We're being asked questions by our organization on how much education is appropriate in terms of hours. We're asked about what best practice companies are doing. We're looking for a variety of metrics that we could then share with our leaders.

These data points are not indicative of level four (results) and level five (ROI) data; rather, they are more closely aligned to activity statistics. Another HRD professional said, "We were not asked [to provide training evaluation reports]. We wanted to test the Success Case methodology." This comment indicated the HRD professional was interested in applying the newest learning industry techniques to their workplace. In this situation, leadership had not specifically asked for a Success Case method, nor had leadership directly asked for any measure.

When the leaders were asked about why they were receiving this data, several of them indicated they had not asked for it. One leader said,

> I don't recall if we specifically asked for [training evaluation reports], but it's always been supplied to us. The CEO has probably asked for the data, but I haven't.

None of the leaders felt the data were unhelpful; but rather that the training evaluation reports were generated by the HRD professionals' own volition. The proactive stance of the HRD professionals was not viewed negatively by leaders. On the contrary, the leaders inferred that the training evaluation reports were important. When one leader was asked whether the leader had requested the training evaluation report, the response was,

> I didn't ask for it specifically, but [the program sponsors] talked about it. The concern I had was, how effective was the leadership training program and how do we know what folks are getting out of this. The facts were suggestive. Good camaraderie but it had to be more than that.

Although this leader did not ask for the report directly, there was a realization of not having a clear understanding of the value of the training without additional data. This leader had a desire to know what the effectiveness of the training program was even though the leader had not asked for any specific report. Perhaps the lack of leader request for this data stemmed from a lack of time to focus on the topic of training. The reason could also be that leaders did not put a high priority on training in the first place.

Training Evaluation Data Use

Phase 1 of this study asked HRD professionals to indicate who used what levels of training evaluation data for what purposes. Table 11 summarizes what HRD professionals did with levels four and five evaluation as well as what HRD professionals thought leaders did with this same information. These percentages only reflect the six organizations that indicated they measured level four (results) and the five organizations that indicated they measured level five (ROI). The majority of HRD professionals indicated that they used the data for formative purposes, including improving and adding courseware, in addition to determining the need for the course moving forward. However, less than half of the HRD professionals indicated that leaders did anything with

this information. If HRD professionals really believed that leaders did not use high level training evaluation data, it is curious to know why they collected it in the first place.

Table 11 – HRD Professional Perception of Training Evaluation Data Use

Use	Level 4 (Results)		Level 5 (ROI)	
	HRD	Leaders	HRD	Leaders
Modify the course	83%	0%	60%	0%
Improve instructor facilitation	83%	0%	60%	0%
Identify additional course needs	83%	0%	40%	0%
Determine learner job placement	17%	0%	0%	0%
Market the program	50%	17%	80%	0%
Identify barriers to learning transfer	83%	0%	40%	0%
Determine continuation of vendor relationships	33%	0%	60%	0%
Determine continuation of course	83%	17%	100%	20%
Justify existence of course	83%	33%	60%	40%
Justify existence of education department	33%	33%	60%	40%

When HRD professionals and leaders in Phase 2 were asked to indicate what the intent of the training evaluation reports was, a wide variety of responses arose. The first major use of training evaluation data was a formative purpose of improving instruction for future classes. The second major area was demonstrating value. The third was the collection and dissemination of best practices. Both HRD professionals and leaders discussed all three of these areas as being key intents for training evaluation practices. This seemed in direct conflict to the high number of HRD professionals in Phase 1 who

said that only some leaders used high level training evaluation report data to justify the existence of courses and training departments. Clearly, leadership saw more of a purpose to this data than HRD professionals felt they actually had. The rest of this section refers to interview data from HRD professionals and leaders in Phase 2 grouped into these three major areas of training evaluation data use.

Formative Data Use

Both HRD professionals and leaders interviewed in Phase 2 indicated formative evaluation as a significant purpose of training evaluation reports. One leader said, "[The training evaluation report] was a big a-ha that enabled us to make even bigger changes to the program." One would assume that HRD professionals would want to use training evaluation for formative purposes, but one would probably not assume that leaders would find formative evaluation as valuable as the HRD professionals. However, it was a common thread in leader interviews that formative evaluation was important. One leader said, "We use [training evaluation reports] as feedback to improve on the course." Another leader cited the importance of using training evaluations for program improvement by saying,

> I started critiquing these programs. I talked to others and said we need to fix this. Others said it was fine – I said prove it to me. I'm talking about real organizational performance – help me understand why we're not connected. In the end, we developed a series of changes in response to the evaluation.

One area that both HRD professionals and leaders felt was important regarding the training evaluation data, was to better identify future participants for the training. One leader said,

> We thought about the results and sought more information about the results. We put in some prescriptive processes to address the issues. For example, we are now more disciplined about who comes into the program now.

This use of the data was not just to improve the course in the future, but also to ensure the right people attended the course in the future. Several leaders indicated concern with the right people attending the right courses at the right time. It seemed like poor training evaluation outcomes were as much an indicator of poor training quality as they were of poor learner selection for a training program.

The bulk of comments regarding formative training evaluation were about making changes to curriculum based on results and impact from past participants. This feedback was used to sometimes dramatically alter training programs for the future. One HRD professional explained the impact of changing a course based on training evaluation data by saying,

> We changed the [leadership training] content to focus on the "How" from the "Leadership 101" model we used before. [The course] didn't have application to the work people were doing. We changed the content so that the "Leadership 101" information was now 30-40% of the program and 60% of the program was application to our culture.

This HRD professional indicated that without the training evaluation work, the drastic changes to content in the leadership curriculum would not have been possible. This HRD professional went on to say,

> When we sat down to start looking at what does this program need to contain and why, the first thing we did was look at the [training evaluation] data to help us know here's how we need to plug this stuff in, which led us to change the program. We planned to make changes based on this data, which we did. Now, we want to make more changes, so we need more data.

There appeared to be a cyclical pattern for using training evaluation data to improve programs. The data were used to make initial changes, but there were always more changes desired in the future. The never ending cycle may be one deterrent to HRD professionals for conducting more frequent high level training evaluations in the future.

Demonstrating Value Data Use

The most frequently indicated use of level four and level five training evaluation data by both HRD professionals and leaders was related to demonstrating the value of training. Several leaders indicated that financial resources in their organizations were limited and that those who wanted to use financial resources needed to demonstrate a return on investment. One HRD professional indicated intended use of training evaluation data as evidence for continued investment in training programs by saying,

> The Success Case analysis was one way for us to be able to say to our senior leaders who are grappling with these finite resources "Here is evidence to help you understand the difference this kind of experience makes in the effectiveness of our leaders."

The evidence gathered by this HRD professional enabled the training department to increase the number of course offerings because the value of the course was successfully demonstrated to leadership. Another HRD professional explained this same experience by saying,

> We've made a lot of cuts in the past two years. We're still in murky waters. There is an extremely limited pool of funds. We're being told to budget for [leadership training]. If [leadership] didn't believe in this, they would not be telling us to budget for more courses. They would find another way to use this money.

There is clearly a need to demonstrate value in order to secure funds for training department continuation. Still another HRD professional said, "If we [conduct training evaluation] we can tell executives that training can make a difference to the bottom line. They are looking for proof of this." Although organizational leaders did not always directly state that financial resources would be increased as a result of positive training evaluations, HRD professionals inferred that meaning when budgets remained static or increased. Leaders often indicated that financial resources for training were at stake, but

they did not indicate a guarantee that more resources would be granted for training programs that demonstrated high value to the organization.

Leaders interviewed in Phase 2 shared the belief that demonstrating value of training sustains those training programs in the future; although they did not often directly state this position to HRD professionals. One leader said,

> I believe that people value [leadership training] if it is provided in a carefully constructed manner so that it meets business needs and meets the challenge of enhancing competence or capability of developing capacity. So that there becomes an acceptance of the investment needed to be made in developing human capital assets. So much as we would do unique things to enhance the physical capital effectiveness and efficiency in our organization; likewise leaders in the business would be interested in enhancing human capital assets to achieve business results or living out our mission.

This leader was linking the impact of training to the core mission of the organization. This link created the foundation necessary to obtain continued funding for training programs. Simply indicating that a program has value is not enough. The program must also be aligned with organizational goals. If it is, funding is more secure, as is evidenced by a leader who said, "The outcome of [training evaluation] is effective because when we were doing some budget cutbacks, one of the things we were asked not to cut is this [training] program." Another leader said,

> There would be a real question as to the value of what we're doing in education [if we did not have training evaluation reports]. I would say education budgets could be cut easily. If you're not able to show value for the dollars you're investing, especially in healthcare, those dollars could go away.

In addition to wanting to demonstrate the value of a program, leaders are also acutely aware of the cost of training as a whole. One leader said, "Education is being viewed as a significant expense. Unless you have champions within the organization, it can create a real challenge." Another leader said,

> Without the [training evaluation] reports, the program would probably have been more vulnerable, especially during tough economical times. Without understand the outcomes of the program, it's hard to justify a couple hundred thousand dollars worth of investment. Its strength is in the outcomes piece.

Another leader reiterated this sentiment by saying,

> I would be challenged to defend the monies that we're spending on people participating in these programs. I'm a real advocate of this [training program] and have launched a regional initiative where all the system hospitals in the region are offering learning programs together. There's a cost to offering all of this. There's justifying the productivity impact for people spending time in learning initiatives versus doing their day to day work. All of this helps me when I have to justify what I'm budgeting to support these programs and learning initiatives and to coordinate this regional learning stuff. I have to coordinate resources and send my staff around the region to be able to teach these courses. Those investments are things I have to justify and these reports are helpful.

It was clear that leaders thought demonstrating value was critical to the long-term existence of training programs and that without training evaluation reports, there would be less reason to keep a training program alive. The leadership perceptions were closely aligned with the HRD professional perceptions.

As mentioned in the analysis of research question one, no HRD professional provided examples of level five (ROI) training evaluation. While five of the eight organizations that participated in Phase 1 of the research indicated they accurately conducted level five training evaluation, none of those organizations spoke about those metrics. When specifically asked about ROI, HRD professionals indicated that ROI studies were not formally done, but rather that level four (results) data were used to demonstrate the value and this seemed sufficient for leaders. Only one of the leaders in Phase 2 mentioned ROI by saying,

> We would look for whether learning had taken place and whether we are able to do something with that learning to move the rock; to provide ROI. You can look at ROI from the construct of demonstrating visible, quantifiable results that are recognized as being a value. Alternatively you could look at the cost to the

business if you didn't do it.

This description of ROI was actually a description of level four because it focused on what the organization gained from the training evaluation, not on comparing the cost to the gain. In fact, none of the leaders interviewed in Phase 2 mentioned concerns or desires to better understand the cost of the training programs in question. There were subjective mentions of the high cost of training, but no specific desire to calculate ROI. There also was not a desire to calculate ROI by the HRD professionals expressed in any way other than a vague desire to prove training programs had an impact. Knowing that more was gained from training than spent on the program did not arise as a critical factor.

Best Practice Data Use

The high level training evaluation practices used by most organizations included the collection of stories about learner application of things learned in class. In fact, the Success Case Method used by numerous organizations in this study was largely built on the concept of learner provided stories. These stories were not only used to demonstrate the value of training programs, they were also used as a source of best practices. One leader explained their organization's process for best practice collection and dissemination as,

> We get [training evaluation] reports at the conclusion of each cohort. Then it's broadly shared throughout the system. You can go through this and see people who have solved problems that we may be dealing with here in our organization but where no one's had the time to work on. It's a good way to share the data and best practices.

The training evaluation report referred to by this leader included learner reports of how the learner applied things learned from class back on the job. While the original intent of these stories was to demonstrate the value of the training program, leaders liked the

tangential effect of best practice identification. Another leader expressed a similar sentiment by saying,

> The reports that have been useful for us from an HR perspective, I pass those on to the staff here so we can evaluate our processes and see if we can benefit from that. I've also seen some best practices sent on to other departments saying, "you may want to look at this."

Organizations that collected stories for level four training evaluation attempted to disseminate best practices uncovered across the organization. These leaders looked for internal and external best practices. One leader explained a successful story of best practice collection as,

> For each project a [leadership program] participant is engaged in, we get an outcome report. One participant's project was reducing door to EKG time in Emergency Department. The clinical outcome is that it reduced time from 25 minutes to 10 minutes average, which is below the average. Six patients with acute myocardial infarctions were captured within 6 or less minutes that would have waited over 45 minutes with the previous process. That's a huge patient safety issue and significant clinical outcome. In this case, the individual studied the process and modified the process to reduce the length of time it takes to get those types of patients. There's example after example of things like this. What's cool about this is that this activity occurred in another hospital in the system, but we can all see that so we can tap the other hospital to see what the change and best practice was and share that information.

This leader used a best practice shared by another organization through a training evaluation in order to improve clinical results at their own hospital.

Although many leaders indicated the importance of collecting and spreading best practice across the organization, most of these leaders also indicated that their organizations were not skilled at doing this. One leader said,

> We need to follow-up if this is a policy that other [hospitals in our system] are going to use if that's going to be a policy. We don't transfer knowledge wisely as an organization.

Another leader explained their organization's difficulty in transferring best practices by saying,

> The piece we don't capture is the replication [of best practice]. I've heard many people say since [they heard of another organization's practice], "Oh, I'm gonna call blank hospital because I'd like to take that piece of work and apply it in my own organization." We don't have a good way of capturing that part of the return.

The difficulty in following up on best practice dissemination was frequent in leader interviews. Another leader said,

> We share all the topics from other hospitals in the system because we want leaders to pick up the phone and call other systems for best practices. Do I know that they actually follow-up? No. But, I do know that our CNO and our senior leadership team are pretty astute about "hey this is a great idea and somebody needs to follow-up on this." I think this could be developed to be an even stronger system of sharing. It's a shame that all of this best practice information is out there and people aren't tapping into it and even sharing it outside of the system.

Leaders conveyed a strong desire to collect and disseminate best practice, although the actual practice of doing so seemed to elude most. Although causes for this were not specifically explored, it is possible that the time and effort required to do this was prohibitive.

Research Question 4: Leadership Perceptions of Training Evaluation

The last research question in this study focused on what leaders thought about the training evaluation reports they received. We've already explored the sporadic prevalence of various training evaluation methods. We've also discovered that while HRD professionals did not always measure all levels of evaluation, they believed there was great value in the highest levels of evaluation. Leaders shared this same opinion that high level training evaluation was important. But believing something is important is only half of the story. Once that important training evaluation is presented to the leader,

what do they think about it? This section explores leadership perceptions derived from Phase 2 leader interviews related to those training evaluations. First, we'll explore the factors that enhanced belief in training evaluation reports and then we'll explore those factors that lessened belief.

Factors that Enhanced Belief in Training Evaluation Data

Leaders were articulate when describing their need in high level training evaluation data. However, they were less articulate when asked to explain what made them believe, or not believe, training evaluation data put in front of them. The previous research questions clearly indicated that leaders believed that outcomes demonstrated value of training. In fact, one leader said, "You know training adds value because you see it in outcomes." But what made the leader believe in the outcomes presented to them? This section explores those elements that enhanced leader belief in outcomes.

Trustworthiness of Data Source

Some of the metrics associated with high level training evaluation included cost savings, quality indicators, revenue, promotion rates, 360 performance review scores, retention, voluntary turnover, etc. When these types of metrics were included in training evaluation reports, leaders did not question the validity of this type of data because the sources of this data were believed to be trustworthy. One leader said,

> One way we know our [leadership training] program is working is that we have statistics that show the number of internal promotion rates. We pull these from our succession planning tool to show that the people are getting something out of the training.

While this leader did not mention the correlation between the promotion metric and training's impact on that metric, it was clear the leader felt training made a difference in this metric. The succession planning tool referenced was a standard tool maintained by

the HR department that had a strong presence and respect across the organization. Leaders believed in data sources that spread across the organization and that had universal authenticity and support. Another leader mentioned reliable data sources by saying,

> What's really great is that we're seeing more promotions from the inside, higher [performance] reviews, and less turnover. When our employees perform well, they want to work here, and when they want to work here, they perform well.

Again, this leader referenced HR provided data they deemed to be from a reliable source. The leader also seemed to inherently believe there was a connection between leadership training and these results, although that connection was not explicitly stated. Another leader felt the same way about training's impact on the organization by saying,

> We know associate satisfaction, turnover rates, promotion rates, and who are the up and comers from those who attended training. I'm convinced from this that the time and money we put into this is extremely valuable.

Leader Provided Data

Another interesting finding was that leaders looked at metrics related to leadership training even when those metrics were not included in the training evaluation reports. For example, one leader said,

> The things I was looking at were 360 [performance review] scores. I was looking for higher scores, but I didn't see a change. That would really have been an indicator that [the leadership training program] worked.

The training evaluation report this leader received did not include 360 performance review data. Even so, this leader felt there should have been a connection there. Perhaps the HRD professional left off 360 performance review data from the training evaluation report because it was unfavorable, as the leader later determined alone.

In other situations, the leaders themselves were tapped to provide evaluative data of training programs. For example, one HRD professional explained their process of getting feedback from leaders as,

> We have a course for frontline supervisors. We get feedback from CEOs and regional leadership teams on the effectiveness of that. The outcome of that is effective because when we were doing some budget cutbacks, one of the things we were asked not to cut is this program.

Several leaders reinforced this practice by indicating that valuable training evaluation data derives from the opinions of the supervisors of learners attending training. Supervisor data sources were deemed to be highly reliable.

Learner Reports to Leaders

The most common element leaders cited regarding their belief of a data source was related to report out techniques. As mentioned in previous sections of this study, a common practice in high level training evaluation in the target organizations was the process of having learners report what they did with the information they learned in class. These learner reports were sometimes collected by instructors to be included in final training evaluation reports. Other times, the learner reports were presented by the learner back to leadership. There was a clear preference from several leaders that the learners should be required to report their results back to their own leadership, not just report back to the training department. One leader said, "If there is only the training and not expectation for the user to be implementing some practices, then I don't believe the training is effective." Another leader said,

> If the [training evaluation] survey had asked their bosses what the student achieved, that would be great. Instead, the survey asked the student to self-identify. That's just not good enough.

This leader felt that learners who were only required to report their findings back to the training department without that data being reviewed by leadership was insufficient proof. When leaders reviewed the claims of their employees, there was an opportunity to refute those claims. One leader said,

> Yes, [leaders believe training evaluation data is credible] because the top leaders are skeptical of this kind of information anyway. We have a transparent organization on things like this data. They will challenge success. If I'm saying across the system that we're at this level, another region may say that they're not that successful.

This concept of reporting organizational outcomes resulting from training program participation in public forums not only validated the data, but also served as a mechanism to transfer best practices. Because other leaders had seen and examined the results of the training program, those results were deemed to be accurate.

Some leaders expressed concerns that while learner reports to leadership were present in some courses, they were not present in as many as they should be. One leader expressed concern that the level of accountability was not high enough for the training evaluation reports by saying,

> I think it would make it more credible if there was more accountability to report out how behaviors have changed. But we don't have that infrastructure. I think we will in the future. There is a project to do this. We do this in [another training] program where people come back during graduation to explain their project. We do see it there, but we don't see it consistently in all of our courses.

It is clear that leaders see premium value in learners describing their application of learning back on the job as long as those reviewing that learning are able to critique and challenge the data.

Longitudinal Studies

HRD professionals were also asked about making the connection between outcomes and the training program's contribution to achieving those outcomes. Most of the HRD professionals did not have specific correlation data to share on this. In fact, one HRD professional said,

> We know that lots of metrics are improving. We know that voluntary turnover is going down. We know that internal promotion rates are going up. We know that performance reviews are improving. We believe that these outcomes are coming from our leadership program, but I can't tell you that for sure.

When asked about control groups, no organization compared the performance of people who attended training with the performance of people who did not attend training. A move in the result of the target metric was deemed sufficient by HRD professionals, even if there was not a control group to which to compare.

While HRD professionals did not pursue longitudinal or control group studies, several leaders indicated that these were critical factors to determining the impact of training on the organization. One leader said,

> I think what you ultimately do is look at the effectiveness of the people who have gone through training over time. What are the skills of managers? What is their overall performance level? Are they better now than they were a year or two ago? We don't have a single number we can use to compare two years ago to now, so we probably could do a better job trying to identify that.

This leader would be more influenced if presented data with a baseline and/or a control group that would have indicated what performance would have looked like without the training program. The pre- and post-measure techniques described earlier in this study were particularly important to leaders of this study. However, most organizations were not conducting these types of training evaluations.

Inherent Belief in Training Value

The previously described factors all contributed to the leaders' ability to believe the training evaluation data presented to them. While not all of these factors were actively pursued by all of the HRD professionals, the leaders consistency mentioned these factors as a way to increase believability. However, during interviews with leaders, the researcher noticed that almost all leaders believed that training was inherently valuable. Leaders tended to describe education as a means to achieving organizational goals. Some comments from leaders indicating their support of training activities included,

> I think [training] is a high value.
>
> It's a good thing that we have these educational offerings.
>
> I believe that training is essential to develop skills and competencies.
>
> I'm definitely an advocate for training.
>
> First of all, I feel there's an inherent value in training – without question.
>
> Education is core aspect of our brand.

Almost every leader interviewed directly stated their support for training. One leader described the role of training as,

> Training is a significant component to ensure the readiness of our workforce to be aligned. And to deliver the results that are required to provide good clinical care in all of our settings. So, yes I think training is a valued asset to meet outcomes.
>
> So, I believe that training is an essential component of ensuring these areas are effective in our organization. I believe that people value [training] if it is provided in a carefully constructed manner so that it meets business needs and meets the challenge of enhancing competence or capability of developing capacity.

This leader clearly believed that training was a critical component to achieving organizational goals. Several other leaders agreed with this perception. One such leader said,

> Education is being viewed as a significant expense. Unless you have champions within the organization, it can create a real challenge. Our CEO is very pro learning and pro development, so that's a real help. But it behooves us to be a good steward of resources and to develop good metrics of success.

A few leaders even said that they felt an obligation to provide development to employees to help those employees perform on the job. Although not all of the leaders stated the value of training quite this urgently, all leaders clearly expressed their belief that training had an important role in the organization. This inherent belief may have influenced their perceptions about training evaluation reports. If a leader inherently believes in the value of training, they are more likely to believe in data that in-turn proves that value.

This inherent belief in the value of training also seemed to be exhibited by leaders who cited lower levels of training evaluation as indicators that training was effective. For example, several leaders indicated that learner satisfaction in a training program was a sign that the training program was effective. One leader suggested that level two (learning) and level three (behavior) training evaluation were indicators of training effectiveness. While these measures may contribute to a full picture understanding of the value of a training program, they do not represent training effectiveness on their own. It is unlikely that leaders who do not inherently believe in the value of training would be as willing to accept low level training evaluations as indicators of overall training program success.

Factors that Lessened Belief in Training Evaluation Data

It goes without saying that HRD professionals who did not use all of the factors listed in the previous section had training evaluation reports that could have been more believable by leaders. Reports that did not include trustworthy data, leader provided data, learner reports to leaders, and/or longitudinal or control groups were less believable than reports that did include all of these elements.

Poor Metrics

When leaders expressed disappointment in training evaluation reports, they often were looking for data not included in the reports. This was an indicator that poor metrics were chosen to represent the effectiveness of the training program. One leader indicated a report was not sufficient by saying,

> It gave me a summary of the program to date in the sense of the number of people who have been through, how many retained, who's promoted, what the gender diversity was, how the content of the program has changed over time. Other than that, it was just kind of a general overview.

While one metric from this statement was mentioned by other leaders as important (promotion rates), this metric appeared to be lost in all of the activity data provided in the training report. This leader went on to tell the researcher that their comments about training program effectiveness were not based on the training evaluation report created by the HRD professional, but rather based on their personal interactions with past learners in the program. This leader was mingling their personal experiences and understanding of the training program with the data provided by the HRD professional to create a complete picture of the effectiveness of the training. However, the leader seemed to put a lot more emphasis on their own experiences rather than those of the training evaluation report. Another leader gave a similar answer by saying,

> I basically said, "I'm not sure how helpful this [training evaluation report] is really." It's too focused on the customer [the learner] – that's the credibility piece. That doesn't help me understand how the program benefits the organization.

This leader indicated that the metrics provided in the training evaluation report were at too low a level. They were level one (reaction) metrics plus information about how the learner planned to apply what they learned back on the job. This leader felt the scope of metrics was insufficient to demonstrate the value of the training program. Learner self-identified data were described in detail in the analysis of research question two. Several leaders indicated that data solely provided by learners indicating how they intended to apply what they learned in class was insufficient without further knowledge of whether the learner actually followed up on their commitments.

Incomplete Data

While some metrics were the wrong things to measure, other metrics may have been the right things to measure, but were not completely or accurately measured. Leaders cited a lack of data in some metrics that indicated a lack of rigor or a lack of information that led to an inability to tell a complete story. One leader said, "the [training evaluation] report didn't have the data for this year, only for last." This leader explained that the training evaluation report was stagnant because data were incomplete. Another example of the right metric without complete data was told by one leader as,

> We know what we're getting from the [leadership training] program in the sense that we can see organizational results. However, I can't dig down into the data to see what it means to me. I can't tell where the promotion rates come from. Is one region better than another? I don't think we know this.

This leader indicated that training evaluation reports showed metrics that were meaningful, but those reports did not tell the whole story. Those reports only told the

high level story, not the individual regional story. This was likely caused by data only being available at the high level. In any event, not having all of the data for a metric made the entire training evaluation report less meaningful, and perhaps even less believable by leaders.

Unmemorable Reports

One of the first questions asked of leaders in Phase 2 was, "How familiar does this evaluation report look to you?" Four of the nine leaders interviewed indicated they were not well acquainted with the training evaluation report or in some cases, had never seen the report. Some leaders indicated they had not read the report for some time, which is not surprising because they would not have had a reason to read the report after it was initially presented to them. Even so, many of the leaders indicated that although they read the report, they didn't think others did the same. One leader said,

> I don't know if other leaders read the training evaluation reports. Because I put so many people through these programs, I stay very connected. I can't tell you that our CEO looks at it as closely as I do, but it's certainly readily available for any leader.

This leader was very familiar with the training evaluation report at their organization because this leader had a vested interest in knowing about the program's success. Other leaders indicated that many people do not refer to the reports. In fact, one leader said, "Just because you send out a message with an attachment, doesn't mean they opened it." Clearly this indicated that many leaders do not read things that were merely sent to them. They had to be actively engaged in reviewing the training evaluation report to encourage their active review of the data.

While the majority of leaders vaguely recalled receiving and reviewing the report in the past, one leader didn't remember ever seeing the report. This was interesting

because the HRD professional indicated this leader was one of the people who asked for the report to be generated in the first place. This may indicate a disconnect between what leaders want when they ask to see a training program's value demonstrated and what HRD professionals provide in return.

CHAPTER 5 – DISCUSSION

This chapter expands upon the findings from Chapter 4 by interpreting those findings in context of the research questions. A discussion of what the findings mean is organized by research question. The discussion is followed by clarifications regarding the limitations of this study and recommendations for further related studies.

Research Question 1: How Training is Evaluated

As mentioned in Chapter 4, both HRD professionals and leaders expressed a desire to link training to organizational outcomes. HRD professionals indicated an increased accountability for demonstrating the value of training, which is consistent with Phillips and Phillips (2007). The desire and practice to correlate training activity to organizational impact is in line with Kirkpatrick's (1998) theory that measuring all four levels of training evaluation builds a case that correlates training results and organizational results. When leaders explained what organizational impact was achieved as a result of training, they tended to describe a journey from learning to action to impact. There was a belief that if the learner understood what was being taught to them and how to apply that to their everyday life they would in-turn experience positive performance improvements. This underlying theme was evident in all of the other findings related to training evaluation methods.

Prevalence of Training Evaluation

The researcher found in Table 9 that organizations in this study are more likely to conduct high levels of training evaluation than other organizations. The higher prevalence of organizations in this study measuring levels four and five evaluations could be a result of the focus on this topic by the abundance of professional organization literature on this topic (Bennett & Griswold, 1984; Bingham & Galagan, 2007; Moore, 2009; Naughton, 2008; Phillips, 1996; Radhakrishnan, 2008). One strong case for conducting evaluation at these higher levels was the recommendation from Phillips and Phillips (2007) that all organizations should measure at least some of their programs at all levels. Moreover, Clarke (2004) found that larger organizations tend to conduct formal learning assessment more often than small organizations. Since the organizations in this study were all large organizations, it is logical that this study found a higher prevalence for level four and level five evaluation.

Even though the organizations in this study were more likely to measure high levels of training evaluation than past studies have indicated, there were still a number of organizations that did not measure these levels of training evaluation at all. This was not surprising based on the findings of Yadapadithaya (2001) who found that 14% of private organizations and 19% of public organizations felt that measuring training effectiveness was a major difficulty. Interview data from this study indicated that many HRD professionals were not measuring high levels of evaluation because of the difficulty of doing so coupled with the time and effort it takes to do this. There also was a problem with access to the organizational impact data required to calculate high level training evaluation. These reasons were consistent with the literature on this topic being that

HRD professionals do not have the time or skill to conduct high level training evaluations (Brewer, 2007; Parry, 1996; Phillips, 2003). Even so, HRD professionals in these organizations felt that high levels of evaluation were important. However, if HRD professionals truly felt that this activity held significant value with executives, one would think they would make time to conduct level four and level five training evaluation. Perhaps HRD professionals did not feel an inherent value to high level training evaluation.

The fact that all four organizations that participated in Phase 2 provided level four training evaluation reports, not level five reports, was curious. There seemed to be an air of discomfort related to ROI. Perhaps this was because return on investment is less understood by organizations. However, it seemed like many HRD professionals and leaders felt that knowing the impact (level four) was enough. As long as a training program did not have an extravagant cost, there did not seem to be a need to conduct cost-benefit analysis. Some organizations did not have a grasp on the cost of individual training programs, which makes return on investment calculations difficult.

Methods Used to Evaluate Training

Chapter 4 discussed multiple ways that high level training evaluation was conducted. Methods for evaluating level four were consistent with Kirkpatrick (1998) and Phillips and Phillips (2007). The overall themes of these findings were that high level training evaluation required an understanding of past and present performance in relation to the performance that was desired. Linking learning to organizational priorities was a common recommendation in the literature that was reinforced with this study (Brinkerhoff & Apking, 2001; Gilley & Maycunich, 2000; Holton, 1996; Kirkpatrick,

1998; Phillips, 1999). The strong desire to do this was shared by HRD professionals and leaders. However, there seemed to be difficulty in clearly articulating what strong performance looked like. This could stem from the training evaluations reviewed being for leadership training programs that tended to have indistinct objectives anyway. If the training evaluations were for a less subjective topic, such as a clinical skill, it would likely have been easier to articulate exactly what desired performance looked like.

Best practices from the HRD literature were also appreciated by leaders interviewed in this study. Those best practices included:

- Comparing high and low performers to better identify premium performance
- Observing learners applying skills directly on the job
- Taking pre- and post-measures to determine the amount of improvement using baseline data

I believe that HRD professionals should ensure their high level training evaluation practices include these concepts to build credibility with leaders. HRD professionals need to have a good understanding of what desired performance looks like so they can design training to achieve this and know what to look for during evaluative periods. If the HRD professional does not understand this or there is no baseline data to use for comparisons, leaders are unlikely to see the data as credible.

Accuracy of Methods Used for Training Evaluation

While all organizations who measured levels four and five did so using accurate methods as defined by Plant and Ryan (1994), all organizations also augmented the strong evidence methods with weak evidence methods. I believe this was because when an organization measured level four (results) and level five (ROI), they were grasping at straws. They were trying to collect any data they could to build the case for training's

value. This activity caused them to couple inaccurate methods of training evaluation with accurate ones. Mixing different measures together in training evaluation reports also made the reports quite large. This study found reports averaging 30 pages in length. Leaders would be unlikely to read reports of this size. The reports not only indicated how the training impacted the organization, they also provided activity measures, such as how many learners attended and what learners thought of the training. Using supplemental methods laid the ground work for including supplemental data that may or may not have been interesting to leaders. However, mixing this data also provided a comprehensive view of a training program, which may have been something some leaders would want to see.

Measuring the Effectiveness of Leadership Development

The fact that all organizations in Phase 2 chose to focus on level four and level five training evaluations for leadership development says something about the use of high level training evaluation. Several comments were made in the study about the difficulty of demonstrating the value of leadership development specifically. I believe that HRD professionals were most concerned about demonstrating the value of leadership development because it was a high cost and high profile learning initiative that had a subjective value statement. Leaders also experienced leadership development directly, so it may be more top of mind for them, so they may ask HRD professionals more questions about this topic. The visibility of leadership development may have contributed to the focus on this topic specifically related to training evaluation.

Another key factor to why all participants in Phase 2 of this study used leadership development as the topic of choice could be linked to leadership development being a

topic that was woven into the cultural fabric of these organizations. The study by Alvarez, Salas, and Garofano (2004) indicated that training programs that were more closely aligned with organizational culture were more likely to be seen positively. The HRD professionals may have chosen to use leadership development for this study because they inherently knew leaders felt leadership development training was valuable because that topic was important to the organization.

Research Question 2: Why Measures are Chosen

Leaders expressed an interest in wanting to see tangible results regarding the effectiveness of learning interventions. Although leaders were not always clear on what that meant, they seemed to know it when they saw it. Leaders who saw data in training evaluation reports they liked could share what those metrics were off the top of their heads. Leaders who did not see data they liked in training evaluation reports could not seem to put their finger on exactly what they did not like, nor could they articulate exactly what they would like to see in the future. This section explores leader perspectives on metrics used by HRD professionals.

Objective Metrics

Leaders often spoke in terms of quantifiable metrics during the Phase 2 interviews. These quantitative metrics included those listed in Chapter 4 (i.e., 360 reviews, finances, satisfaction scores). Leaders were comfortable discussing these metrics as a cornerstone of their individual success. Leaders seemed relatively open to the idea that these outcomes could be correlated to the effectiveness of learning. Leaders wanted to see these types of metrics used to demonstrate the value of a particular learning program. However, not all leaders fully understood what that meant or how it would

work. It was clear that in order for the leader to believe that training had an impact on results, the training evaluation had to compare pre- and post-performance of the employee. I believe that HRD professionals need to heed this advice by (a) ensuring training evaluation plans compare performance before and after a training program and (b) explaining to leaders how training is being evaluated so everyone is on the same page regarding how results will be used.

Subjective Metrics

The power of stories was clear in this study. Leaders relied more on their hearts than their heads. This could be related to the fact that these are leaders in non-profit organizations, but I don't believe that is the case because Catholic healthcare works similarly to other industries. I think HRD professionals need to master the art of storytelling to weave the story of how a training program benefits the organization. The story cannot just be a series of numbers. The story must also speak to the heart of the leader and the organization.

Although the literature on training evaluation recommends using both quantitative and qualitative data, the literature is heavily based on quantitative data. I think HRD professionals would benefit by focusing equally on both of these areas instead of only trying to quantify their value. Leaders tend to remember stories more than they remember raw data. I think the story format is especially critical for training evaluation reports because HRD professionals are attempting to make a correlation between training programs and organizational results. This correlation requires a leap of faith that must be supported by well woven story. The HRD professionals in this study often used Success Case Methods (Brinkerhoff, 1983) to tell the story of training effectiveness. The Success

Case Method appeared to be appropriately applied in many situations and tended to be well received by leaders, even if the training evaluation data included therein was not comprehensively presented.

These stories often include information provided directly by the learners. There was clearly an important role for self-identified data in high level training evaluation. Training evaluation methods proposed by Kirkpatrick (1998) do not specifically recommend using learner self-identified data to measure level four. Moreover, Plant and Ryan (1994) indicate that self-identified data should not be used for any level of training evaluation except level one. In practice, HRD professionals appeared to be heavily using self-identified data, probably because of the ease of data collection. While I think it is important to use self-identified data, I also think it is important for HRD professionals to be cognizant of overusing this type of data, which in turn may make results less credible.

The data from this study also indicated that leaders felt learner satisfaction was critical to program success. Most of the literature downplays the importance of level one (reaction) data (Holton, 1996; Phillips and Phillips, 2007). Kirkpatrick (1998) believed there was a linear path that leads from learner satisfaction to results. Most studies on training evaluation focus on levels three (behavior), four (results), or five (ROI) (Alvarez, Salas, and Garofano, 2004; Clemenz and Weaver, 2003; Kontoghiorghes, 2001; Rosti and Shipper, 1998). Even though the HRD literature puts the majority of emphasis on high level training evaluation, leaders tend to consider that an important indicator of success. Because the training programs examined in this study were all leadership development programs, it is possible that leaders put more credence in learner reaction

because the learners in these programs are fellow leaders. Perhaps more tactical or skill based training would carry less importance for level one (reaction) data.

Less Meaningful Metrics

The lack of enthusiasm from leaders regarding training activity and efficiency metrics was not surprising. Phillips and Phillips (2007) indicated these metrics should be downplayed. However, making sure that the benefits of training outweigh the cost is a critical element of training evaluation that requires the collection of training activity and efficiency measures (Kirkpatrick, 1959; Phillips, 1999). Therefore, it was understandable that HRD professionals collected this data, but the researcher recommends downplaying this information in final reports sent to leadership because they did not have an appetite for this type of information.

Research Question 3: Who Receives Reports and Why

The data from this study indicated that HRD professionals felt pressure to demonstrate their value to leadership and they responded to that pressure by providing training evaluation reports. This section analyzes information from this study about who receives training evaluation reports and why.

Training Evaluation Data Requestors

Although leadership examined in this study did not specifically request training evaluation reports in most cases, leadership appreciated the effort to demonstrate value. Leaders felt an inherent need to know more about training effectiveness, whether or not they knew how or what to ask of HRD professionals. The lack of leader request for high level training evaluation data is consistent with the findings of Jones (2008) who found that only 52% of hospitals in Illinois measure level four training evaluation due, in part,

to the fact that leaders do not ask for this information. This finding also supports the Corporate University Xchange's (2008) finding that 27% of leaders inherently believe in the value of training and therefore do not ask for measures to prove the value. Based on this, I believe that many HRD professionals are not directly asked for training evaluation to support the existence of the training function; however, I do not believe that the lack of formal request should discourage HRD professionals from gather this data anyway.

Another finding in this study was that leaders seemed to be equally requesting training activity data as well as training result data, which is contrary to Becker, Huselid, and Ulrich (2001) who put a much stronger value on outcome metrics. This is not to say that leaders did not request outcome data. Rather, it was more likely that leaders did not know what to ask for to determine training effectiveness, so they asked for whatever was easiest, which was activity based data. Common training evaluation metrics cited in several studies indicate training activity metrics as common and at least marginally important (Becker, Huselid, & Ulrich; Bernthal, 2005; Rosenburg, 2001). I believe that HRD professionals need to think of training evaluation holistically to include both activity metrics (what we do) and outcome metrics (what we achieve) to tell the full picture. This belief is in alignment with Becker, Huselid, & Ulrich.

Training Evaluation Data Use

The first finding related to the use of training evaluation data was that HRD professionals did not think that leaders were using training evaluation reports for as much as what the leaders themselves indicated. I believe this perception from the HRD professionals stems from the fact many leaders are not asking for high level training evaluation reports. If there is no request for the data, one could assume that there is also

no planned use for the data. However, I think that HRD professionals need to be aware that leaders are thinking about the improvement of training, the value of training, and ways to expand best practices from training across the organization. If HRD leaders understand this, they can better market and obtain value from training evaluation efforts.

As mentioned above, a significant finding of this study was that both HRD professionals and leaders felt that training evaluations should be used for formative purposes. These purposes include changing content, changing facilitation, increasing frequency of a program, or even discontinuing a program. HRD professionals tended to explain the positive and the negative aspects of their programs in training evaluation reports in the light of what was being done about the feedback. Feedback was not merely shared on a program; it was also related to an improvement action for the future. Kirkpatrick (1998) originally suggested formative evaluation as a critical intent of training evaluation, but he focused the audience of this information to the HRD professionals, not the leaders. I believe HRD professionals need to not only conduct formative evaluations, but also share the results and actions with their leaders.

Beyond formative use of training evaluation data, the most common reason to evaluate training was to demonstrate the value training provided back to the organization. HRD professionals and leaders were aware (at least subjectively) about how expensive training is to provide and that there needs to be a demonstration of value. This is aligned with research on the topic that indicates leaders want to see program value articulated (Bersin & Associates, 2009 February; Phillips & Phillips, 2001). It was not surprising that leaders felt HRD professionals needed to demonstrate the value for the programs they provided because this desire is both logical and supported by the literature.

The findings also supported literature indicating that high level training evaluation should only be conducted for the most costly and important programs (Becker, Huselid, & Ulrich, 2001; Blanchard, Thacker, & Way, 2000; Dixon, 1996). Several HRD professionals and leaders indicated that they did not feel a need to calculate level four or level five for many programs. This simply reinforces that HRD professionals should pick the programs with the most opportunity to influence perceptions about the training department for high level training evaluation activities.

One interesting finding related to the use of high level training evaluation was a link between organizational culture and goals with training program success. Dixon (1996) indicated that training programs were much more likely to be perceived as successful if they clearly and adequately supported organizational goals. Leader interviews always had the spin of what the organization was trying to achieve first, only then followed by how training supported that goal. I think HRD professionals need to constantly think about their role in supporting the organizational goals and then frame all training evaluation activity toward achieving this goal.

The final, and potentially largest opportunity area, for using training evaluation reports was the collection and dissemination of best practices. The Brinkerhoff (1983) Success Case Method is based on learner stories, which are in themselves best practices; however, Brinkerhoff does not specifically suggest a knowledge management approach when conducting a Success Case Method. The knowledge management and best practice field was not specifically referenced in the preparation of this study; however, Wenger (1998) indicates that best practices are collected and disseminated organically across organizations. The organic nature of best practices most likely describes why other

training evaluation literature did not mention the by-product of best practices. I believe that high level training evaluation and the collection and dissemination of best practices go hand in hand and should therefore be purposively pursued during evaluation planning.

Research Question 4: Leadership Perceptions of Training Evaluation

All of the findings of this study reinforced the claim in the literature that leadership thinks high level training evaluation is important (Bennett & Griswold, 1984; Bingham & Galagan, 2007; Moore, 2009; Naughton, 2008; Phillips, 1996; Radhakrishnan, 2008). Even so, there were mixed findings related to leadership perceptions of the data presented to leaders. There were clearly factors that increased the believability of the data, but there were also factors that did not. I believe that HRD professionals should ensure that their practices for high level training evaluation include the factors that increase believability because the practice of conducting this type of evaluation is incredibly time consuming and expensive. If the training evaluation is not believed in the end, what is the point of doing it in the first place? This section provides a further examination of how HRD professionals should apply these findings to their practice.

Factors that Enhanced Belief in Training Evaluation Data

Several factors arose during data analysis that indicated what made leaders more likely to believe data. The findings of this study seem to reinforce those summarized in Table 4 from the Corporate University Xchange (2008). That study indicated that while some leaders were less likely to believe training evaluation data provided to them, others were quite likely to believe the data when it is well collected and written. There will always be some leaders who do not believe in training evaluation reports, but it is

imperative for HRD professionals to focus on how they can help willing leaders believe in training evaluation reports.

A key area of believability was based on the three dimensions of believability from Prat and Madnick (2007), which included (a) trustworthiness of source, (b) reasonableness of data, and (c) temporality of data. This study found that leaders were especially influenced by the trustworthiness of the source. Information that came from organizationally recognized systems or groups was deemed to be much more reliable than other sources of questionable or unfounded data. Reasonableness of data was also cited as important in the sense that leaders wanted to validate the findings by understanding them firsthand. If a leader did not find the data to be reasonable during this process, they in-turn did not believe in the data to the extent they otherwise may have. These perceptions of leaders were also closely related to the temporality of data, which required leaders to consider the outcomes to be within the parameters of what the training program was designed to do in the first place. I was surprised that none of the leaders indicated an overinflated result on a training evaluation report. I had expected that leaders would have felt the impact of training was less than what the HRD professionals felt it was. The lack of concerns related to overinflated results may be due to the leaders' inherent belief in training value back to the organization. If a leader inherently believes in the value of training, they are more likely to believe in data that proves that value.

The training evaluations that leaders indicated they believed were very much in line with Phillips (2007, April) who suggested that the reports should be of the right quality and quantity and should come from a credible source. Some of the top influencers of leader believability in this study mirrored those suggested by Phillips

(2007, December) including executive provided data and auditable data. However, several other recommendations from this same reference were not mentioned by any of the participating organizations including isolating training, conducting error rate adjustments, and only using partial results. I believe that none of the organizations who participated in Phase 2 of this study were conducting advanced training evaluations so these techniques were not employed. There certainly was not any aversion to these techniques expressed by HRD professionals or leaders, but I'm not convinced that application of these techniques would have increased leader believability of the training evaluation reports.

One factor that multiple leaders indicated would increase believability was the use of longitudinal or control group studies. Leaders wanted to be able to determine what would happen if a training program didn't exist. The only way to do that is to try the program on some people and compare their performance to the performance of people who did not participate in the training program (Kirkpatrick, 1998; Phillips, 1999; Rothwell & Kazanas, 1998). Although the control group concept is common in the training evaluation literature, none of the organizations studied in this research used that technique. I believe the use of this technique should be increased because so many leaders indicated that they believed training evaluation results were lacking in that area. Leaders wanted to know how performance differed as a result of training, which is something they had to make a leap of faith about without having a control group.

The final, and perhaps most influential, factor that increased leader belief in training evaluation data was related to the leaders inherently believing in the value of training. This was the most interesting finding to the researcher because it provided

insight into who was selected to participate in Phase 2. Participants for Phase 2 were selected by HRD professionals from Phase 1. These HRD professionals had a vested interest in the training programs about which they were providing training evaluation data. It was possible that HRD professionals purposely provided leaders for Phase 2 who were already advocates for training and therefore more likely to have positive things to say about training evaluations. In the end, these leaders said positive and negative things about the training evaluation reports they received, so I believe they were all honest in their responses.

This inherent belief likely influenced leadership willingness to believe that low level training evaluations (reaction, learning, and behavior) were strong indicators of training program effectiveness. Much of the literature indicates that low level training evaluation is not sufficient to prove training program value (Becker, Huselid, & Ulrich, 2001; Figis, 2001; Parry, 1996; Phillips, 1999; Russ-Eft & Preskill, 2001; Tesoro & Tootson, 2000). However, I believe this inherent belief in training's value to the organization caused these leaders to be more open to low level training evaluation results. HRD professionals should be sure to consider their audience's inclination toward the value of training when preparing reports for them.

Factors that Lessened Belief in Training Evaluation Data

In addition to HRD professionals not incorporating the factors mentioned above, there were several other factors that specifically decreased the belief in training evaluation data. The first of these was a lack of meaningful metrics. Blanchard, Thacker, and Way (2000) indicated that metrics should be directly aligned to the training program's intent in order to be meaningful. However, several metrics used by HRD

professionals in this report were more closely aligned with metrics of interest to HRD professionals rather than leaders. The literature tells us that courses closely aligned with organizational strategy are more likely to impact organizational results (Alvarez, Salas, & Garofano, 2004). Even so, some of the metrics in use in some of the training evaluation reports did not appear to be correlated with organizational results. It seemed that a more conscientious selection of metrics would have improved leadership believability of reports.

The second factor that lessened leader belief in training evaluation was incomplete data. Leaders tended to describe incomplete pictures when indicating incomplete data was a problem. They explained that although some of the information they needed was there, some things were still missing. This situation could have likely been addressed if the HRD professional would have triangulated multiple data sources, which also would have isolated training effects and built credibility (Fusch, 2001; Phillips, 1997).

The third factor that lessened leader belief in training evaluation was unmemorable reports. Something that was missing from the reports was a well crafted story. Figgis (2001) found that a well-written story increased leader believability in training evaluation reports. Figgis also found that although leaders avoided reading the reports, once they did read the report, they had positive comments about what they read. Reports that were more closely aligned to organizational priority were more likely to be read, which is consistent with the literature (Brinkerhoff & Apking, 2001; Gilley & Maycunich, 2000; Holton, 1996; Kirkpatrick, 1998; Phillips, 1999).

Limitations

The qualitative case study design for this study was chosen because there was limited literature on leadership perceptions of training evaluation reports. In order to better understand this core question, a flexible research tradition was chosen. This research tradition, while enabling the researcher to fully explore the topic at hand, limited the ability to generalize the findings from this study due to a small and narrow sample size. The sample for this study was sub-section of a single industry. This narrow sample was chosen because the study sought to interview executive leaders at organizations. These types of leaders were unlikely to be willing to participate in research on this topic directly, so the researcher chose to narrow the sample to organizations in which he had connections. While this decision was imperative to getting access to the target population, it also raised the question of whether leaders in different industries would have different perceptions related to this topic.

A related issue with the target population is that the organizations studied were faith-based non-profit healthcare systems. A faith-based organization has a unique culture and tends to attract employees who have a complimentary belief system to that of the organization. One might assume that leaders in a faith-based non-profit healthcare system may have a greater tendency to be more open to subjective ideas and stories. However, in my experience of working within both for-profit and non-profit organizations, leaders in faith-based organizations exhibit many of the same behaviors as leaders in other types of organizations. Leaders in non-profit healthcare are just as concerned about increasing profit, decreasing cost, and increasing quality as their counterpart leaders in for-profit healthcare or even in other industries. I believe the

results of this study can be generalized beyond Catholic healthcare because non-profit faith-based healthcare functions in similar ways to other organizations. Further evidence of the generalizability of this study can be derived in the consistency of the findings with literature on the topic. Because the findings of this study were similar to the claims of other experts in the field, it is safe to assume the results of this study could be applied to those in other organizations.

A limitation of this study was that within-case sampling techniques were used to identify leaders for Phase 2. The HRD leaders in Phase 1 identified leaders for Phase 2 interviews. Although this sampling technique is not random, it was required for this study because (a) the researcher needed the HRD professionals' introductions to get access to the leaders and (b) leaders in Phase 2 needed to have received the training evaluation report in the past and the only way to ensure this was through the HRD professionals. It is possible that leaders in this study may not have been representative of the average leader within an organization because the HRD professionals wanted to provide leaders who were inclined to support training.

A final limitation of this study was that the researcher did not validate the accuracy of training evaluations. This study assumed that level four and level five training evaluations conducted using appropriate methods were accurate representations of training's impact on the organization. The researcher acknowledged that these training evaluations may or may not have been well prepared, but the validation of their quality was outside of the scope of this study. It is possible that leadership perceptions of training evaluation reports were tainted by the quality of the reports leaders saw. The researcher felt that leadership perceptions were valuable and valid regardless of the

quality of the reports they were given, which is why the quality of the reports was not a variable in this study.

Recommendations for Further Research

This study focused on the practices of large Catholic healthcare organizations that conduct level four and level five training evaluation. Eight of the ten largest Catholic healthcare providers in the U.S. responded to the Phase 1 survey and seven of those organizations indicated they measured level four and/or level five training evaluation. This study excluded the one organization that did not measure levels four or five from Phase 2. It might be interesting for a future study to do the opposite of this and only focus on organizations that do not measure either level four or level five to better understand why. The mirror side of this study would be an interesting accompaniment to the HRD industry's collective knowledge on why and how high level training evaluation is conducted.

As mentioned earlier, all of the leaders included in Phase 2 of this research believed there was an inherent value to training. A future study may be able to compare opinions of those leaders who inherently believe in the value of training with those who do not. It would be interesting to see what factors differ for leaders who do not inherently believe in the value of training.

Conclusions and Implications

As an HRD professional, I struggle with how much time and effort should be spent on doing HRD work versus measuring the effectiveness of HRD work. The results of this study were important to me because I wanted to strike a more perfect balance between doing and evaluating. HRD professionals like myself, should alter their high

level training evaluation practices by ensuring that metrics included in training evaluation reports match the metrics leaders expect to see. Thorough and accurate collection of data related to those metrics should take precedence over expansion to include more metrics in reports.

HRD professionals should focus on depth of understanding, not breadth of measurement possibilities. We need to be able to explain what performance looks like for someone who has experienced our training programs compared to someone who has not. We cannot rely solely on organizational metrics to tell our story. We have to weave in our contribution as HRD professionals to more clearly indicate how we help the organization succeed. We need to be better at using control groups, longitudinal studies, and baseline data to compare past performance to improved future performance.

We cannot simply rely on the opinions of those leaders closest to us. We have to open our eyes to the larger community of leaders in our organizations, some who believe in our work and some who do not. We need to engage those leaders to not only tell our story, but also to improve our training offerings. HRD professionals are focused on improving the performance of employees. Now, we need to focus on telling our story in a way that leaders believe our claims.

REFERENCES

Alvarez, K., Salas, E., & Garofano, C. M. (2004). An integrated model of training evaluation and effectiveness. *Human Resource Development Review, 3*(4), 385-416.

American Hospital Association (2007). *2007 Annual Survey*. Retrieved June 15, 2009 from http://www.aha.org/aha/resource-center/Statistics-and-Studies/fast-facts.html and http://www.chausa.org/NR/rdonlyres/68B7C0E5-F9AA-4106-B182-7DF0FC30A1CA/0/FACTSHEET.pdf

Arnst, C. (2008, March 25). Health care: Not so recession-proof. *Business Week.* Retrieved April 28, 2009 from http://www.businessweek.com/technology/content/mar2008/tc20080324_828167.htm

Becker, B. E., Huselid, M. A., & Ulrich, D. (2001). *The HR scorecard: Linking people strategy, and performance*. Boston: Harvard Business School Press.

Bennett, B., & Griswold, D.F. (1984). Proving our worth: The training value model. *Training and Development, 38*(10), 81-83.

Bernthal, P. (2005). Measurement gets strategic. *Training and Development, 59*(5), 53-59.

Bersin and Associates (2006, November). *High-impact learning measurement* (v 1.0). Berkeley, CA: Bersin, J.

Bersin and Associates (2009, February). *The corporate learning factbook 2009: Healthcare edition* (v1.0). Berkeley, CA: O'Leonard, K.

Bersin and Associates (2009, April). *The state of learning and talent measurement* (v1.0). Berkeley, CA: Bersin, J.

Billings, D. M. (2000). A framework for assessing outcomes and practices in web-based courses in nursing. *Journal of Nursing Education, 39*(2), 60-67.

Bingham, T., & Galagan, P. (2007, January). Satyam creates value through learning. *Training and Development,* 30-34.

Blanchard, P. N., Thacker, J. W., & Way, S. A. (2000). Training evaluation: Perspectives and evidence from Canada. *International Journal of Training and Development, 4*(4), 295-304.

Bober, C. F., & Bartlett, K. R. (2004). The utilization of training program evaluation in corporate universities. *Human Resource Development Quarterly, 15*(4), 363-383.

Brewer, T. K. (2007). *Use of Phillip's five level training evaluation and return on investment framework in the U.S. non-profit sector*. Unpublished doctoral dissertation, University of North Texas. Denton, TX. (UMI No. 3288246)

Brinkerhoff, R. O. (1983). The success case: A low-cost, high-yield evaluation. *Training and Development Journal, 37*(8), 58-60.

Brinkerhoff, R. O., & Apking, A.M. (2001). *High impact learning*. Cambridge, MA: Perseus.

Carlson, J., & Galloro, V. (2009, June 8). Into the red. *Modern Healthcare, 39*(23), 26-30.

Clarke, N. (2004). HRD and the challenges of assessing learning in the workplace. *International Journal of Training and Development, 8*(2), 140-156.

Clemenz, C. E., & Weaver, P. A. (2003). Dimensions of perceived training quality: A comparison of measurements. *Journal of Quality Assurance in Hospitality & Tourism, 4*(1/2), 47-70.

Coghlan, D. & Brannick, T. (2001). *Doing action research in your own organization.* London, UK: Sage.

Corporate University Xchange (2008). *8th annual benchmark measurement study chapter.* Retrieved May 1, 2009, from http://www.corpu.com/research/document/365/8th-annual-benchmark-measurement-study-chapter/

Cousins, J. B., & Leithwood, K. A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research, 56*(3), 331-364.

Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among the five traditions*. Thousand Oaks, CA: Sage.

DiPietro, R. B. (2004). Return on investment in managerial training: Does the method matter? *Journal of Foodservice Business Research, 7*(4), 79-96.

Dixon, N. M. (1996, May). New routes to evaluation. *Training and Development*, 82-85.

Expertus & Training Industry Inc. (2008). *Measuring learning as budgets tighten.* Retrieved April 14, 2009, from http://www.expertus.com/sites/expertus.com/files/public-files/Survey-Results_Learning-Measurement_Expertus-2008.pdf

Figgis, J. (2001). What convinces enterprises to value training and learning, and what does not? In Smith, A. (Ed.), *Return on Investment in Training: Research Readings* (103-121). Kensington Park, Australia: National Centre for Vocational Education Research.

Florida, R. (2005). *The flight of the creative class*. New York: HarperCollins.

Fusch, G. E. (2001). What happens when the ROI model does not fit? *Performance Improvement Quarterly, 14*(4), 60-76.

Gilley, J. W., & Maycunich, A. (2000). *Organizational learning, performance, and change: An introduction to strategic human resource development*. Cambridge, MA: Perseus.

Grammatikopoulos, V., Papacharisis, V., Koustelios, A., Tsigilis, N., & Theodorakis, Y. (2004). Evaluation of the training program for Greek Olympic education. *The International Journal of Educational Management, 18*(1), 66-73.

Holton, E. F., III (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly, 7*(1), 5-21.

Jones, P. A. (2008). *Training evaluation in large Illinois hospitals*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, IL. (UMI No. 3223624)

Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. *Journal of American Society for Training and Development, 13*(11), 3-9.

Kirkpatrick, D. L. (1998). *Evaluating training programs: The four levels* (2nd ed.). San Francisco: Berrett-Koehler.

Kontoghiorghes, C. (2001). Factors affecting training effectiveness in the context of the introduction of new technology – A U.S. case study. *International Journal of Training and Development, 5*(4), 248-260.

Krathwohl, D. R. (1998). *Methods of educational and social science research* (2nd ed.). Long Grove, IL: Waveland Press.

Longenecker, C. O., & Fink, L.S. (2005). Management training: Benefits and lost opportunities (part II). *Industrial and Commercial Training, 37*(2), 73-79.

Mattox, J. R., II & Jinkerson, D. L. (2005). Using survival analysis to demonstrate the effects of training on employee retention. *Evaluation and Program Planning, 28*, 423-430.

McKillip, J. (2001). Case studies in job analysis and training evaluation. *International Journal of Training and Development, 5*(4), 283-289.

Meister, J. C. (1998). *Corporate universities: Lessons in building a world-class work force.* Boston: McGraw Hill.

Miles, M. B., & Huberman, A.M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Mitchell, R. J. (2001). Evaluation as an organizational development tool. *International Journal of Training and Development, 5*(4), 275-282.

Moore, C. (2009, February). Selling the C-suite on results. *Chief Learning Officer, 8*, 18-23.

Morgan, G. A., Leech, N.L., Gloeckner, G. W., & Barrett, K. C. (2007). *SPSS for introductory statistics* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Naughton, J. (2008, September). IOL: Determining the impact of learning. *Chief Learning Officer, 7*, 32-27.

Parry, S. B. (1996, May). Measuring training's ROI. *Training and Development*, 72-77.

Phillips, J. J. (1996, April). How much is the training worth? *Training and Development, 20*(4), 20-25.

Phillips, J. J. (1997). *Return on investment: In training and performance improvement programs.* Houston, TX: Gulf.

Phillips, J. J. (1999). *Accountability in human resource management*. Woburn, MA: Butterworth-Heinemann.

Phillips, J. J. (2007, April). Measuring ROI: Fact, fad, or fantasy? *Training and Development*, 42-46.

Phillips, J. J. (2007, December). Measuring the ROI of a coaching intervention, Part 2. *Performance Improvement*, *46*(10), 42-46.

Phillips, J. J. & Phillips, P. P. (2007). Show me the money: The use of ROI in performance improvement, Part 1. *Performance Improvement, 46*(9), 8-22.

Phillips, P. P. (2003). Training evaluation in the public sector. *Dissertation Abstracts International, 64* (09), 3162A. (UMI No. 313677)

Plant, R. A., & Ryan, R. J. (1994). Who is evaluating training? *Journal of European Industrial Training, 18*(5), 27-30.

Prat, N., & Madnick, S. E. (2007). *Measuring data believability: A provenance approach.* MIT Sloan School of Management Working Paper, 4672-07. Retrieved May 29, 2009 from http://ssrn.com/abstract=1075723

Radhakrishnan, M. (2008, October). Learning measurements: It's time to align. *Chief Learning Officer, 7*, 36-39.

Richards, L. (2005). *Handling qualitative data: A practical guide*. Thousand Oaks, CA: Sage.

Roffe, I. (2002). E-learning: Engagement, enhancement and execution. *Quality Assurance in Education, 10*(1), 40-50.

Rosenberg, M. (2001). *E-Learning: Strategies for delivering knowledge in the digital age*. New York: McGraw-Hill.

Rosti, R. T., Jr., & Shipper, F. (1998). A study of the impact of training in management development program based on 360 feedback. *Journal of Management Psychology, 13*(1), 77-89.

Rothwell, W. J., & Kazanas, H. C. (1998) *Mastering the instructional design process: A systematic approach* (2nd ed.). San Francisco: Jossey-Bass.

Russ-Eft, D., & Preskill, H. (2001). *Evaluation in organizations: A systematic approach to enhancing learning, performance, and change*. Cambridge, MA: Perseus.

Senge, P. M. (1990). *The fifth discipline.* New York: Doubleday/Currency.

Tesoro, F., & Tootson, J. (2000). *Implementing global performance measurement systems.* San Francisco: Jossey-Bass Pfeiffer.

Thurmond, V. A., Wambach, K., Connors, H. R., & Frey, B. B. (2002). Evaluation of student satisfaction: Determining the impact of a Web-based environment by controlling for student characteristics. *American Journal of Distance Education, 16*(3), 169–189.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity.* New York: Cambridge University Press.

Wexler, S., Hart, J., Karrer, T., Martin, M., Oehlert, M., Parker, S., Schlender, B., Thalheimer, W. (2008, September 23). *E-Learning 2.0*. Retrieved April 21, 2009 from http://www.elearningguild.com/research/archives/index.cfm?action=viewonly&id=134

Yadapadithaya, P. S. (2001). Evaluating corporate training and development: An Indian experience. *International Journal of Training and Development, 5*(4) 261-274.

APPENDIX A – HUMAN SUBJECTS APPROVAL

Colorado State University
Knowledge to Go Places

Research Integrity & Compliance Review Office
Office of the Vice President for Research
321 General Services Building - Campus Delivery 2011
Fort Collins, CO
TEL: (970) 491-1553
FAX: (970) 491-2293

NOTICE OF APPROVAL FOR HUMAN RESEARCH

DATE:	January 11,2010
TO:	Gilley, Jerry, Education Waite, Alina, 1588 School of Education, Davies, Timothy, Education, Preston, Kevin, Education
FROM:	Barker, Janell, CSU IRB 1
PROTOCOL TITLE:	Leadership's Perceptions of Results and Return on Investment Training Evaluations
FUNDING SOURCE:	NONE
PROTOCOL NUMBER:	09-1047H
APPROVAL PERIOD:	Approval Date: January 11,2010 Expiration Date: January 07,2011

The CSU Institutional Review Board (IRB) for the protection of human subjects has reviewed the protocol entitled: Leadership's Perceptions of Results and Return on Investment Training Evaluations. The project has been approved for the procedures and subjects described in the protocol. This protocol must be reviewed for renewal on a yearly basis for as long as the research remains active. Should the protocol not be renewed before expiration, all activities must cease until the protocol has been re-reviewed.

If approval did not accompany a proposal when it was submitted to a sponsor, it is the PI's responsibility to provide the sponsor with the approval notice.

This approval is issued under Colorado State University's Federal Wide Assurance 00000647 with the Office for Human Research Protections (OHRP). If you have any questions regarding your obligations under CSU's Assurance, please do not hesitate to contact us.

Please direct any questions about the IRB's actions on this project to:

Janell Barker, Senior IRB Coordinator - (970) 491-1655 Janell.Barker@Research.Colostate.edu
Evelyn Swiss, IRB Coordinator - (970) 491-1381 Evelyn.Swiss@Research.Colostate.edu

Barker, Janell

Includes:
Approval is for 10 survey participants for Phase 1 and 30 interview participants for Phase 2 using the approved electronic cover letters to obtain consent.

Page: 1

Colorado State University
Knowledge to Go Places

Research Integrity & Compliance Review Office
Office of the Vice President for Research
321 General Services Building - Campus Delivery 2011
Fort Collins, CO
TEL: (970) 491-1553
FAX: (970) 491-2293

Approval Period:	January 11,2010 through January 07,2011
Review Type:	EXPEDITED
IRB Number:	00000202

APPENDIX B – SURVEY COVER LETTER

Dear Participant,

My name is Kevin Preston and I am a researcher from Colorado State University in the School of Education's Organizational Performance and Change (OPC) department. We are conducting a research study on Catholic healthcare leadership's perception of training evaluation reports. The Principal Investigator is Jerry Gilley (Program Chair of OPC Program) and the Co-Principal Investigator is Kevin Preston (doctoral student in the OPC Program).

We would like you to complete a 15 minute on-line survey to share your organization's training evaluation practices with the researchers. Participants who conduct certain types of training evaluation will be contacted within three weeks to better understand your organization's training evaluation practices. You will be asked for additional information about how you and your organization use the training evaluation data. You will also be asked to provide contact information for one to three leaders within your organization that receive training evaluation data for their participation in phase two of the study. Your participation in this research is voluntary. If you decide to participate in the study, you may withdraw your consent and stop participation at any time without penalty.

Although this survey is not anonymous in order to enable follow-up with certain participants, your responses to this survey and your participation in follow-up

conversations about your training evaluation practices will be kept confidential in final reports. Individual organizations will not be linked to specific comments or practices in the final report. Only the researchers will have access to data tagged with organizational sources which will be stored in a secured location and this tagged data will be destroyed 3 years after study completion. While there are no direct benefits to you, we hope to gain more knowledge on how leaders perceive training evaluation reports from learning professionals.

There are no risks associated with this research. It is not possible to identify all potential risks in research procedures, but the researcher(s) have taken reasonable safeguards to minimize any known and potential, but unknown, risks.

If you have any questions, please contact Kevin Preston at 303.383.2774 or Kevin.Preston@ColoState.edu. If you have any questions about your rights as a volunteer in this research, contact Janell Barker, Human Research Administrator, at 970-491-1655.

To acknowledge your understanding of this consent and continue to the survey, click "I accept" below.

I Accept I Don't Accept

APPENDIX C – SURVEY

TRAINING EVALUATION SURVEY

This survey is Phase 1 in a 2 Phase study. Phase 1 seeks to identify the types of training evaluation conducted and how those training evaluation results are used. Phase 2 seeks to understand leadership's perceptions of training evaluation results at the organizational impact and/or return on investment levels.

For Phase 1, a single educator from each facility is asked to complete this survey about your facility's training evaluation practices. The data from this survey will be used to determine how facilities measure the effectiveness of training and who uses the training evaluation reports. Name and contact information for this survey are only collected to follow-up with you in case your facility is selected to participate in Phase 2. Individuals and organizations will not be associated with responses in the final report. Participation in this survey does not impact your performance.

The leadership at facilities conducting organizational impact or return on investment training evaluations may be contacted for follow-up interviews about their perceptions of this data. If your organizational leadership is identified as a follow-up interview candidate, we will first contact you to clarify your practices.

This survey should take you approximately 10-20 minutes to complete. Please complete this survey no later than _____.

I appreciate your willingness to participate in this study. If you have questions about this process, please contact Kevin Preston at 303.363.2774.

(0) CONTACT INFORMATION

This information will only be used to contact you for follow-up if your facility is selected to participate in phase 2 of the study. This information will not be associated with the data analysis.

YOUR NAME:
YOUR PHONE:
FACILITY NAME:

Training Evaluation Survey

(1) Level 1: Reaction Training Evaluations

These are questions about how you measure learner satisfaction (training evaluation level 1 - Reaction) for your courses.

For what percentage of your training do you evaluate learner satisfaction?
Select one

- ❑ 100%
- ❑ 75-99%
- ❑ 50-74%
- ❑ 25-49%
- ❑ 1-24%
- ❑ None – we don't evaluate this

Measuring reaction (i.e. learner satisfaction) training evaluation is...
Select one

Critical — Important — Waste of Time

How do you measure learner satisfaction?
Select all that apply

- ❑ Syllabus review with course sponsors
- ❑ Learner satisfaction survey
- ❑ Observation of the learner on the job
- ❑ Learner identifies financial benefit
- ❑ Learner's manager identifies financial benefit
- ❑ Examination or test
- ❑ Accreditation of course by external bodies
- ❑ Cost/Benefit analysis
- ❑ Organizational Metrics (revenue, safety, etc.)
- ❑ None – we don't do this
- ❑ Other (please specify)

For each of the following purposes, who uses reaction training evaluation results?
Select all that apply for each purpose

Purposes	Educators use to:	Course Sponsor use to:	Education Leaders use to:	Business Unit Managers use to:	Senior Executives Use to:	External People (Public, Vendors) use to:
Modify the course	❑	❑	❑	❑	❑	❑
Improve the instructor's facilitation	❑	❑	❑	❑	❑	❑
Identify if additional courseware is needed	❑	❑	❑	❑	❑	❑
Determine learner job placement	❑	❑	❑	❑	❑	❑
Market the program	❑	❑	❑	❑	❑	❑
Identify barriers to transferring learning to the job	❑	❑	❑	❑	❑	❑
Determine if vendor relationships should continue	❑	❑	❑	❑	❑	❑
Determine if the course should continue	❑	❑	❑	❑	❑	❑
Justify the existence of the course	❑	❑	❑	❑	❑	❑
Justify the existence of the education department	❑	❑	❑	❑	❑	❑
Other Uses – Please specify:						

Training Evaluation Survey

(2) Level 2: Learning/Testing Training Evaluations

These are questions about how you measure learner learning (training evaluation level 2 - Learning) for your courses.

For what percentage of your training do you test learner learning?

Select one

- ❑ 100%
- ❑ 75-99%
- ❑ 50-74%
- ❑ 25-49%
- ❑ 1-24%
- ❑ None – we don't evaluate this

Measuring/testing learning is...

Select one

Critical — Important — Waste of Time

How do you test learning?

Select all that apply

- ❑ Syllabus review with course sponsors
- ❑ Learner satisfaction survey
- ❑ Observation of the learner on the job
- ❑ Learner identifies financial benefit
- ❑ Learner's manager identifies financial benefit
- ❑ Examination or test
- ❑ Accreditation of course by external bodies
- ❑ Cost/Benefit analysis
- ❑ Organizational Metrics (revenue, safety, etc.)
- ❑ None – we don't do this
- ❑ Other (please specify)

For each of the following purposes, who uses learning training evaluation results?

Select all that apply for each purpose

Purposes	Educators use to:	Course Sponsor use to:	Education Leaders use to:	Business Unit Managers use to:	Senior Executives Use to:	External People (Public, Vendors) use to:
Modify the course	❑	❑	❑	❑	❑	❑
Improve the instructor's facilitation	❑	❑	❑	❑	❑	❑
Identify if additional courseware is needed	❑	❑	❑	❑	❑	❑
Determine learner job placement	❑	❑	❑	❑	❑	❑
Market the program	❑	❑	❑	❑	❑	❑
Identify barriers to transferring learning to the job	❑	❑	❑	❑	❑	❑
Determine if vendor relationships should continue	❑	❑	❑	❑	❑	❑
Determine if the course should continue	❑	❑	❑	❑	❑	❑
Justify the existence of the course	❑	❑	❑	❑	❑	❑
Justify the existence of the education department	❑	❑	❑	❑	❑	❑
Other Uses – Please specify:						

Training Evaluation Survey

(3) Level 3: Behavior Training Evaluations

These are questions about how you measure how learners apply training to the job (training evaluation level 3 - Behavior).

For what percentage of your training do you measure learner application of skills on the job?

Select one

- ❑ 100%
- ❑ 75-99%
- ❑ 50-74%
- ❑ 25-49%
- ❑ 1-24%
- ❑ None – we don't evaluate this

Measuring behavior/application training evaluation is...

Select one

Critical — Important — Waste of Time

How do you measure behavior/application of training to the job?

Select all that apply

- ❑ Syllabus review with course sponsors
- ❑ Learner satisfaction survey
- ❑ Observation of the learner on the job
- ❑ Learner identifies financial benefit
- ❑ Learner's manager identifies financial benefit
- ❑ Examination or test
- ❑ Accreditation of course by external bodies
- ❑ Cost/Benefit analysis
- ❑ Organizational Metrics (revenue, safety, etc.)
- ❑ None – we don't do this
- ❑ Other (please specify)

For each of the following purposes, who uses behavior training evaluation results?

Select all that apply for each purpose

Purposes	Educators use to:	Course Sponsor use to:	Education Leaders use to:	Business Unit Managers use to:	Senior Executives Use to:	External People (Public, Vendors) use to:
Modify the course	❑	❑	❑	❑	❑	❑
Improve the instructor's facilitation	❑	❑	❑	❑	❑	❑
Identify if additional courseware is needed	❑	❑	❑	❑	❑	❑
Determine learner job placement	❑	❑	❑	❑	❑	❑
Market the program	❑	❑	❑	❑	❑	❑
Identify barriers to transferring learning to the job	❑	❑	❑	❑	❑	❑
Determine if vendor relationships should continue	❑	❑	❑	❑	❑	❑
Determine if the course should continue	❑	❑	❑	❑	❑	❑
Justify the existence of the course	❑	❑	❑	❑	❑	❑
Justify the existence of the education department	❑	❑	❑	❑	❑	❑
Other Uses – Please specify:						

Training Evaluation Survey

(4) Level 4: Organizational Impact Training Evaluations

These are questions about how you measure how the training impacted the organization including financial impact, clinical results, etc. (training evaluation level 4 - Results).

For what percentage of your training do you measure organizational impact/results of training?

Select one

- ❑ 100%
- ❑ 75-99%
- ❑ 50-74%
- ❑ 25-49%
- ❑ 1-24%
- ❑ None – we don't evaluate this

Measuring organizational impact/results training evaluation is...

Select one

Critical — Important — Waste of Time

How do you measure organizational impact/results?

Select all that apply

- ❑ Syllabus review with course sponsors
- ❑ Learner satisfaction survey
- ❑ Observation of the learner on the job
- ❑ Learner identifies financial benefit
- ❑ Learner's manager identifies financial benefit
- ❑ Examination or test
- ❑ Accreditation of course by external bodies
- ❑ Cost/Benefit analysis
- ❑ Organizational Metrics (revenue, safety, etc.)
- ❑ None – we don't do this
- ❑ Other (please specify)

For each of the following purposes, who uses results training evaluation results?

Select all that apply for each purpose

Purposes	Educators use to:	Course Sponsor use to:	Education Leaders use to:	Business Unit Managers use to:	Senior Executives Use to:	External People (Public, Vendors) use to:
Modify the course	❑	❑	❑	❑	❑	❑
Improve the instructor's facilitation	❑	❑	❑	❑	❑	❑
Identify if additional courseware is needed	❑	❑	❑	❑	❑	❑
Determine learner job placement	❑	❑	❑	❑	❑	❑
Market the program	❑	❑	❑	❑	❑	❑
Identify barriers to transferring learning to the job	❑	❑	❑	❑	❑	❑
Determine if vendor relationships should continue	❑	❑	❑	❑	❑	❑
Determine if the course should continue	❑	❑	❑	❑	❑	❑
Justify the existence of the course	❑	❑	❑	❑	❑	❑
Justify the existence of the education department	❑	❑	❑	❑	❑	❑
Other Uses – Please specify:						

Training Evaluation Survey

(5) Level 5: Return on Investment (ROI) Training Evaluations

These are questions about how you measure if the organizational impact or results outweigh the cost to provide the training in the first place (training evaluation level 5 – Return on Investment).

For what percentage of your training do you measure the Return on Investment of training?

Select one

- ❑ 100%
- ❑ 75-99%
- ❑ 50-74%
- ❑ 25-49%
- ❑ 1-24%
- ❑ None – we don't evaluate this

Measuring Return on Investment training evaluation is...

Select one

Critical — Important — Waste of Time

How do you measure Return on Investment?

Select all that apply

- ❑ Syllabus review with course sponsors
- ❑ Learner satisfaction survey
- ❑ Observation of the learner on the job
- ❑ Learner identifies financial benefit
- ❑ Learner's manager identifies financial benefit
- ❑ Examination or test
- ❑ Accreditation of course by external bodies
- ❑ Cost/Benefit analysis
- ❑ Organizational Metrics (revenue, safety, etc.)
- ❑ None – we don't do this
- ❑ Other (please specify)

For each of the following purposes, who uses ROI training evaluation results?

Select all that apply for each purpose

Purposes	Educators use to:	Course Sponsor use to:	Education Leaders use to:	Business Unit Managers use to:	Senior Executives Use to:	External People (Public, Vendors) use to:
Modify the course	❑	❑	❑	❑	❑	❑
Improve the instructor's facilitation	❑	❑	❑	❑	❑	❑
Identify if additional courseware is needed	❑	❑	❑	❑	❑	❑
Determine learner job placement	❑	❑	❑	❑	❑	❑
Market the program	❑	❑	❑	❑	❑	❑
Identify barriers to transferring learning to the job	❑	❑	❑	❑	❑	❑
Determine if vendor relationships should continue	❑	❑	❑	❑	❑	❑
Determine if the course should continue	❑	❑	❑	❑	❑	❑
Justify the existence of the course	❑	❑	❑	❑	❑	❑
Justify the existence of the education department	❑	❑	❑	❑	❑	❑
Other Uses – Please specify:						

APPENDIX D – SURVEY RESPONSE REMINDER A

When will this message be sent?

1 week after the survey is sent when there are 2 weeks left to return it.

Who will this message be sent to?

The HRD professionals of the ten largest Catholic healthcare systems

How will this message be sent?

Via e-mail with the original message forwarded/copied in the text

What will the message say?

Just a friendly reminder to complete the Leadership Perceptions of Results and Return on Investment Training Evaluations survey no later than Friday January 27, 2010. The survey should not take you any longer than 15 minutes to complete.

Thank you in advance for your participation. Please contact Kevin Preston with questions or comments at 303.383.2774 or Kevin.Preston@ColoState.edu

APPENDIX E – SURVEY RESPONSE REMINDER B

When will this message be sent?

2 weeks after the survey is sent when there is 1 week left to return it.

Who will this message be sent to?

The HRD professionals of the ten largest Catholic healthcare systems based on the Catholic Consortium's OEL group who have not completed the survey as of yet

How will this message be sent?

Via e-mail with the original message forwarded/copied in the text

What will the message say?

Just a friendly reminder to complete the *Leadership Perceptions of Results and Return on Investment Training Evaluations* survey no later than Friday January 27, 2010. The survey should not take you any longer than 15 minutes to complete. Surveys have been received by 8 out of the 10 other target organizations. Your input would make the study even more valuable. Thank you in advance for your participation. Please contact Kevin Preston with questions or comments at 303.383.2774 or Kevin.Preston@ColoState.edu

APPENDIX F – HRD PROFESSIONAL INTERVIEW FRAMEWORK

Intent	Question
Clarify incorrect responses between evaluation techniques and the level of evaluation that is being attempted	▪ You indicated that you use 'Examinations or tests' to measure training results. Tell me more about that. ▪ What is used? ▪ How is it used? ▪ Why do you use this? ▪ What kind of training is evaluated?
Learn how level four or level five reports are created and what they look like	▪ Describe the reports you use to communicate level four (results) and/or level five (ROI) data to leadership. What's included? What do they look like? ▪ When were these reports given to leadership? ▪ How do you gather data for these reports?
Determine why level four and level five data are collected	▪ Why did you start collecting levels four and five data? ▪ Did you initiate collecting levels four and five data or was this a practice when you arrived? ▪ Did leadership ask for this?
Clarify the purpose of the data use	▪ In your survey you indicated that [role] used this data for [purpose]. Tell me about this. ▪ What actions have been taken as a result of your reports (e.g., budget changes, staffing changes, strategic plan changes)?
Determine how credible HRD professionals believe their levels four and five evaluation reports to be	▪ What does leadership do with these reports? ▪ What value do you think leadership gains from your reports? ▪ What makes the data you provide accurate and believable? ▪ How do you make the data you provide a realistic reflection of the value of training? ▪ Why do you think leadership thinks the data you provide is accurate?

APPENDIX G – LEADERSHIP INTERVIEW CONSENT

The following text was included in an e-mail sent to leadership interviewees for Phase 2 of the study. Participants of this phase of research acknowledged their consent to participate by responding to the e-mail.

Dear Participant,

My name is Kevin Preston and I am a researcher from Colorado State University in the School of Education's Organizational Performance and Change (OPC) department. We are conducting a research study on Catholic healthcare leadership perceptions of training evaluation reports. The Principal Investigator is Jerry Gilley (Program Chair of OPC Program) and the Co-Principal Investigator is Kevin Preston (doctoral student in the OPC Program).

You were identified by _____ [name of educator at healthcare system] as someone who receives training evaluation reports and would be willing to briefly talk with me about your perceptions of those reports. I would like to talk with you for 20 minutes about your perception of the training evaluation reports you receive. This conversation will be recorded for research accuracy and transcription. Your participation in this research is voluntary. If you decide to participate in the study, you may withdraw your consent and stop participation at any time without penalty.

Your comments during this interview will be kept confidential in final reports. Individual organizations will not be linked to specific comments or practices in the final report. Only the researchers will have access to data tagged with organizational sources, which will be stored on a secured server and this tagged data will be destroyed 3 years after study completion. While there are no direct benefits to you, we hope to gain more knowledge on how leaders perceive training evaluation reports from learning professionals.

There are no risks associated with this research. It is not possible to identify all potential risks in research procedures, but the researcher(s) have taken reasonable safeguards to minimize any known and potential, but unknown, risks.

If you have any questions, please contact Kevin Preston at 303.383.2774 or Kevin.Preston@ColoState.edu. If you have any questions about your rights as a volunteer in this research, contact Janell Barker, Human Research Administrator, at 970-491-1655.

To acknowledge your understanding of this consent please reply to this e-mail indicating your agreement. At that time, I will contact you or your administrative assistant to schedule our 20 minute interview. I thank you in advance for your willingness to help with this doctoral dissertation research.

APPENDIX H – LEADERSHIP INTERVIEW FRAMEWORK

Intent	Question
Clarify what leadership expects in terms of levels four and five evaluation data.	▪ Describe your perception of the value of training within your organization? ▪ How do you know if training is adding the value it was originally intended to add? ▪ What sort of reports do you prefer from HRD professionals to prove the value of training? ▪ Which of these reports are most useful in explaining the value of training? ▪ Your local HRD professional provides _____ type of training evaluation report today. Did you ask for this report? If so, why? ▪ What if you never had this program? What would you gain? What would you lose?
Identify what is done with training evaluation reports. *Ask these questions in context of a specific evaluation provided to the leader*	▪ How familiar does this evaluation report look to you? ▪ When did you last see it? ▪ What made you want to/not want to read it? ▪ Why did you review this information? *Reference uses from survey to cross-validate HRD professional and leader responses* ▪ What actions did you take as a result of reading this information (budget, staffing, communications, direction)
Determine credibility for the training evaluation data.	▪ How did you judge the reports, or similar reports, related to: o Value o Credibility o Accuracy o Relevancy to business o Usability for decision making o Alignment with expectations ▪ What would be different if you didn't receive this report or similar reports?

APPENDIX I – CODING FRAMEWORK

RQ1 - How is training evaluated

- Evaluation Process
  - Correlate Training to Impact
    - External Factors
    - Triangulating Several Data Sources
  - Impact Map & Success Case
  - Method
    - Compare High and Low Performers
    - Measuring non-traditional programs
    - Observations
    - Pre- Post-Measures
      - Baseline Data
      - Control Group
  - Too Hard to Calculate

RQ2 - Why measures are chosen

- Purpose for measurement
  - Application of Learning to Job
  - Linkage to Org Priority
  - Training is Expensive
  - What if you didn't have reports
- Types of Metrics
  - Accurate Impact Metrics
  - Anecdotal
  - Inaccurate Impact Metrics
  - Performance Management
  - ROI
  - Self Identified
    - Learner Satisfaction
  - Training Activity
  - Training Efficiencies

RQ3 - Who Receives Reports and Why

- Senior Leaders Receive
- Use of Evaluation Reports
    - Formative
    - Internal to the Training Dept
    - Receive Funding
    - Transfer Learning to Other Orgs
- Who's Asking

RQ4 - Leadership Perceptions

- Belief in the process
    - Auditable Data
    - Conservative data collection and analysis
    - Error rate adjustments
    - Executive provided data
    - Infer correlations L1-L5
    - Joint credit
    - Long-term Results Longevity
    - Partial result use
    - Reasonableness of data
    - Temporality of data (Estimate = Actual)
    - Trustworthiness of source
- Didn't help
- Don't remember receiving
- Incomplete Data
- Training is Inherently Valuable