Florida International University
FIU Digital Commons

FIU Electronic Theses and Dissertations | University Graduate School

11-12-2010

An Empirical Study of Kirkpatrick's Evaluation Model in the Hospitality Industry

Ya-Hui Elegance Chang
Florida International University, elegance5@hotmail.com

Follow this and additional works at: http://digitalcommons.fiu.edu/etd

Part of the Other Education Commons

Recommended Citation
Chang, Ya-Hui Elegance, "An Empirical Study of Kirkpatrick's Evaluation Model in the Hospitality Industry" (2010). *FIU Electronic Theses and Dissertations.* Paper 325.
http://digitalcommons.fiu.edu/etd/325

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

AN EMPIRICAL STUDY OF KIRKPATRICK'S EVALUATION MODEL IN THE HOSPITALITY INDUSTRY

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF EDUCATION

in

ADULT EDUCATION AND HUMAN RESOURCE DEVELOPMENT

by

Ya-Hui Elegance Chang

2010

To: Dean Delia C. Garcia
College of Education

This dissertation, written by Ya-Hui Elegance Chang, and entitled An Empirical Study of Kirkpatrick's Evaluation Model in the Hospitality Industry, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

_____
Tonette S. Rocco

_____
M. O. Thirunarayanan

_____
Douglas H. Smith, Co-Major Professor

_____
Thomas G. Reio Jr., Co-Major Professor

Date of Defense: November 12, 2010

The dissertation of Ya-Hui Elegance Chang is approved.

_____
Dean Delia C. Garcia
College of Education

_____
Interim Dean Kevin O'Shea
University Graduate School

Florida International University, 2010

© Copyright 2010 by Ya-Hui Elegance Chang

All rights reserved.

DEDICATION

This dissertation is dedicated with the greatest love and pride to my family. To my parents, the completion of this dissertation would not have been possible without their love, sacrifice, vision, and confidence in me. They have given my sisters and me the whole world, literally. And to my two younger sisters as my best friends, who make me laugh, wipe my tears, see me stumble, cheer me on, watch me succeed, and keep me strong. I hope they all will be proud of me as a daughter and as a big sister. For their endurance and endless love through this long journey, I am eternally grateful.

ACKNOWLEDGMENTS

The completion of this dissertation was made possible by the valuable contribution of many individuals, to whom I owe much gratitude. First and foremost, I extend my sincere appreciation for the support and guidance provided by my chairs, Drs. Douglas H. Smith and Thomas G. Reio, Jr. In the few times I had serious doubts of the possibility of the study, Dr. Smith stood firmly with me and never gave up on me. I am privileged to have his wisdom, support, and patience throughout my entire doctoral program. I hope he will be proud of me as his very last student he chaired in his long and distinguished career. Dr. Reio rescued me in midstream of this dissertation process, but his humor, enthusiasm and confidence about my completing the study, helped me through many hurdles. His leadership style is truly one that adult educators and HRD professionals can look up to.

I also greatly appreciate the expertise and suggestions received from Drs. Tonette S. Rocco and M.O. Thirunarayanan, the other two members of my doctoral committee. I admire Dr. Rocco's intelligence, personality, and her humor. Dr. Thiru is a quiet gentleman, but his insightful comments were heard loud and clear. Their breadth and depth of knowledge have helped me to better formulate and express my thoughts in this dissertation.

Two other individuals have played important roles in facilitating the completion of this study. First, I have greatly admired Dr. Paulette Johnson's sharp sense about statistics since the first day I sat in her stat class. Then, working those long hours with SPSS she put me at ease when we also talked about travels, tropical fruits, etc., those things we both love. Finally, I wish to thank Dr. Mike Hampton as my mentor in the

hospitality industry. Not only was his expertise about the hospitality industry, and help in connecting with the hotel and gaining access to the data that made this study possible, but also he and his entire family embraced me as a family member.

All these individuals not only made an impact in my brain, but more importantly they also left a permanent mark in my heart.

ABSTRACT OF THE DISSERTATION

AN EMPIRICAL STUDY OF KIRKPATRICK'S EVALUATION MODEL IN THE HOSPITALITY INDUSTRY

by

Ya-Hui Elegance Chang

Florida International University, 2010

Miami, Florida

Professor Thomas G. Reio, Jr., Co-Major Professor

Professor Douglas H. Smith, Co-Major Professor

This study examined Kirkpatrick's training evaluation model (Kirkpatrick & Kirkpatrick, 2006) by assessing a sales training program conducted at an organization in the hospitality industry. The study assessed the employees' training outcomes of knowledge and skills, job performance, and the impact of the training upon the organization. By assessing these training outcomes and their relationships, the study demonstrated whether Kirkpatrick's theories are supported and the lower evaluation levels can be used to predict organizational impact.

The population for this study was a group of reservations sales agents from a leading luxury hotel chain's reservations center. During the study period from January 2005 to May 2007, there were 335 reservations sales agents employed in this Global Reservations Center (GRC). The number of reservations sales agents who had completed a sales training program/intervention during this period and had data available for at least two months pre and post training composed the sample for this study. The number of agents was 69 ($N = 69$).

Four hypotheses were tested through paired-samples *t* tests, correlation, and hierarchical regression analytic procedures. Results from the analyses supported the hypotheses in this study. The significant improvement in the call score supported hypothesis one that the reservations sales agents who completed the training improved their knowledge of content and required skills in handling calls (Level 2). Hypothesis two was accepted in part as there was significant improvement in call conversion, but there was no significant improvement of time usage. The significant improvement in the sales per call supported hypothesis three that the reservations agents who completed the training contributed to increased organizational impact (Level 4), i.e., made significantly more sales. Last, findings supported hypothesis four that Level 2 and Level 3 variables can be used for predicting Level 4 organizational impact. The findings supported the theory of Kirkpatrick's evaluation model that in order to expect organizational results, a positive change in behavior (job performance) and learning must occur. The examinations of Levels 2 and 3 helped to partially explain and predict Level 4 results.

TABLE OF CONTENTS

CHAPTER PAGE

1. INTRODUCTION 1
Background of the Problem 1
Statement of the Problem 3
Research Question and Hypotheses 4
Research Question 4
Research Hypotheses 5
Significance of the Study and Anticipated Consequences 5
The Need for Examining Kirkpatrick's Evaluation Model 5
The Need Within the Hospitality Industry 6
The Need Within the Body of AE/HRD Research and Theory 7
Definition of Terms 8
Assumption of This Study 11
Delimitations of This Study 11
Summary 12

2. REVIEW OF THE LITERATURE 13
Theoretical Review 13
The Purpose and Importance of HRD Program Evaluation 13
The Challenges in Conducting Evaluations 18
A Review of Selected Evaluation Models 21
Summary of the Theoretical Review of the Literature 34
Empirical Review 36
Surveys of the Use of Kirkpatrick's Model 40
Limited Utilization of Kirkpatrick's Model 43
Studies Examining Barriers to Utilizing the Higher Levels of Kirkpatrick's Model 44
Studies Utilizing All Four Levels of Kirkpatrick's Model 47
Summary of the Empirical Literature 52
Summary of Chapter 2 and Research Question 54
Research Question 54
Research Hypotheses 54

3. METHOD 56
Research Question and Hypotheses 56
Research Question 56
Research Hypotheses 56
Methodological Rationale and Review of Methodological Literature 57
Research Design 57
Effect Size 59
Population and Sample 60

Sample Size 60
The Training Program/Intervention 61
Data Collection 62
Analysis of the Data 66
Limitation of This Study 69
Summary 69

4. ANALYSIS OF DATA 70
Research Question and Hypotheses 70
Research Question 70
Research Hypotheses 70
Population and Sample... 71
Job Title 72
Length of Employment 72
Finings Pertaining to Hypothesis One 73
Finings Pertaining to Hypothesis Two... 74
Finings Pertaining to Hypothesis Three 78
Finings Pertaining to Hypothesis Four 81
Summary 83

5. SUMMARY, CONCLUSIONS, AND IMPLICATIONS 85
Summary of the Study 85
Discussion 86
Research Question 87
Research Hypotheses 87
Implications for Theory, Research, and Practice 91
Implications for Theory 91
Implications for Research 93
Implications for Practice 100
Closing Remarks 101

REFERENCES 102

APPENDIXES 111

VITA 119

LIST OF TABLES

TABLE PAGE

1. Studies Related to Training Evaluation Utilizing Kirkpatrick's Model……..37

2. Summary of the Variables Needed and Statistical Tests Used to Analyze Each of the Four Hypotheses……67

3. Summary of the Population and Sample Sizes ……...71

4. Training Dates and the Number of Participants ...……..72

5. Knowledge and Skills Variable and Statistical Results for Hypothesis One ...……. 74

6. Job Performance Variables and Statistical Results for Hypothesis Two ...……75

7. Organizational Impact Variables and Statistical Results for Hypothesis Three ...….79

8. Costs of Training Intervention ……...80

9. Correlations of Organizational Impact Change from Pre to Post with Changes in Employee Learning and Job Performance Variables for Hypothesis Four ……...82

10. Summary Hierarchical Regression Analysis with Employee Learning and Job Performance, Predicting Sales per Call ……83

CHAPTER 1

INTRODUCTION

This study examined Kirkpatrick's training evaluation model (Kirkpatrick & Kirkpatrick, 2006) by assessing a sales training program conducted at an organization in the hospitality industry. The study assessed the employees' training outcomes of knowledge and skills, job performance, and the impact of the training upon the organization. By assessing these training outcomes and their relationships, the study demonstrated whether Kirkpatrick's theories are correct and the lower evaluation levels can be used to predict organizational impact. This introductory chapter discusses the background of the problem and the basic research question and hypotheses addressed in the study. It then provides an overview of the conceptual framework of the study that will be fully discussed in chapter 2, and the purpose, significance, and anticipated consequences of the study. This chapter concludes with the definitions of key terms, the assumptions, and the limitations of the study.

Background of the Problem

The field of human resource development (HRD) and HRD professionals are responsible for developing effective HRD programs within organizations. According to Werner and DeSimone (2005), there are a number of challenges to HRD, including increasing workforce diversity, competing in a global economy, eliminating the skills gap, meeting the need for lifelong learning, and facilitating organizational learning. The increasing complexity of the workplace demands more on-the-job training and a more educated and trained workforce (Hudson, 2002; Newman & Hodgetts, 1998). With the increasing costs for advanced training, many organizations are trying to become more

aggressive in determining the value of training upon employees' performance, and in turn the value of the employees' performance upon the continuous growth of the organization. This is generally referred to as the return on investment (ROI) of training and development (Abernathy, 2003; Cascio, 2000; Delerno, 2001; Gagné & Medsker, 1996; Hall, 2001; Philips, 2003a; Swanson, 2001). In addition, as learning and skill development increases and becomes more integrated with business strategies, the need to evaluate the learning function is increasing.

While evaluation has long been an integral part of learning, organizational learning executives and HRD professionals continue to struggle with developing an evaluation system that measures the value of the learning function with the same precision as financial and accounting evaluation measures. Various evaluation models have been considered, but one of the earliest models that continues to be one of the most widely utilized, and adapted into other evaluation models, is Donald Kirkpatrick's four-level evaluation model (ASTD, 2009; Kirkpatrick, 1959a; Kirkpatrick, 1959b; Kirkpatrick, 1960a; Kirkpatrick, 1960b; Kirkpatrick & Kirkpatrick, 2006). The Kirkpatrick model evaluates a training program on four levels or areas: (a) the participants' reactions to the program, (b) an assessment of the content, or what the participants learned, (c) the participants' performances on the job, and (d) the impact of the training upon the organization. Most training programs, however, have primarily relied on the first two levels of Kirkpatrick's model (Alliger & Janak, 1989; ASTD, 2009; Kirkpatrick, 1998; Kirkpatrick & Kirkpatrick, 2006), with less emphasis on levels three and four. This is due, in part, to the increased difficulty in assessing job performance and organizational impact. It is also due to the question of a specific training program, and

even training and development in general, being able to evaluate total performance and organizational impact.

Statement of the Problem

As discussed above, this study addresses three issues. First, there is increasing interest by organizational and human resource development (HRD) professionals to pursue higher levels of evaluation to track their training and development investments (American Society for Training and Development, 2007; Phillips, 1999; Phillips, 2003a; Van Buren, 2001). Second, while various training and development evaluation systems and models have been developed, and will be presented in chapter 2, Kirkpatrick's four-level evaluation model continues to be the most widely used by HRD practitioners, and referred to in the HRD literature (Alliger & Janak, 1989; ASTD, 2009; Kaufman & Keller, 1994; Kirkpatrick, 1998; Kirkpatrick & Kirkpatrick, 2006; Phillips, 1998; Russ-Eft & Preskill, 2001; Warr & Bunce, 1995).

The HRD field has primarily utilized and relied on Levels 1 and 2[1] of Kirkpatrick's evaluation model, the participant's reaction of the program, and an assessment of the learning from the program content, with less focus on Levels 3 and 4, performances on the job and organizational impact (Alliger & Janak, 1989; ASTD, 2009; Kirkpatrick, 1998; Kirkpatrick & Kirkpatrick, 2006). The *ASTD 2002 State of the Industry Report* found that only one-third of companies profiled tried to measure learning gained, and that 12 % or less tried to measure job performance and business impact (Bersin, 2003). Similar findings were also evident in an ASTD (2009) recent research.

[1] Note: In this proposal, when discussing specific evaluation levels of any evaluation model, the format will be listed as Level 1, Level 2, etc., rather than level one, level two, etc.

With the increasing need for intensive evaluation of learning and performance, what is needed is more research on Levels 3 and 4.

The increasing interest in more extensive evaluation, particularly in higher levels, has resulted in just conducting Level 3 or 4 assessments (Bersin, 2003; Hackett, 1997; Pine & Tinkley, 1993; Shelton & Alliger, 1993; Strunk, 1999; Swanson & Gradous, 1988). However, Kirkpatrick contends it is risky to conduct evaluation just at certain levels and expect the results will provide the overall conclusions of the training intervention. Positive reaction to the training experience (Level 1) does not guarantee that learning (Level 2) occurred. Similarly, if employees did learn from the training, it does not mean they will change their behaviors and apply what they learned back onto their jobs (Level 3). Therefore, no organizational results/impact (Level 4) can be expected unless a demonstrated change in behavior occurs. Thus, Kirkpatrick contends it is important to conduct the evaluation on all four levels to determine what areas have improved and what still needs further improvement. This study addresses these issues by examining the impact of Kirkpatrick's evaluation model, with particular assessments of the inter-level relationships between the four levels.

Research Question and Hypotheses

This study was guided by the following research question and four research hypotheses.

Research Question

Do the data from a training program implemented at an organization in the hospitality industry support the theories of Kirkpatrick's evaluation model (Kirkpatrick & Kirkpatrick, 2006)?

Research Hypotheses

To answer this research question, four research hypotheses served as the guides for the data to be collected and analyzed.

Hypothesis one (H_1). Employees who completed the training will improve their knowledge of the content and required skills (Level 2).

Hypothesis two (H_2). Employees who completed the training will improve their job performance (Level 3).

Hypothesis three (H_3). Employees who completed the training will contribute to increased organizational impact (Level 4).

Hypothesis four (H_4). Employee learning (Level 2) and job performance (Level 3) will predict organizational impact (Level 4).

The general context of this study is the hospitality industry. Specifically, the data for this study were the evaluations of training provided to reservations sales agents of a large international hotel chain.

Significance of the Study and Anticipated Consequences

The need for conducting this study is significant in the areas of examining Kirkpatrick's evaluation model, the hospitality industry, and the body of adult education (AE) and HRD research and theory.

The Need for Examining Kirkpatrick's Evaluation Model

Despite its having been introduced a half century ago, Kirkpatrick's model has been extensively studied, widely accepted, but also legitimately criticized (Alliger & Janak, 1989; ASTD, 2009; Brinkerhoff, 1987; Bushnell, 1990; Holton, 1996; Kirkpatrick & Kirkpatrick, 2006; Kraiger, Ford & Salas, 1993; Phillips, 2003). In many applied

studies that evaluate Kirkpatrick's four levels, the data used are not uniform, nor standardized (Attia, 1998; Bledsoe, 1999; Lockwood, 2001; Tidler, 1999; Wertz, 2005). There have been very few studies that apply and assess the four levels in a single evaluation/application, where data are collected for a single training program to evaluate all four levels. Therefore, what will be examined in this study is Kirkpatrick's model applied to a single training program where data were collected to evaluate the knowledge and skills, job performance, and organizational impact of employees completing the training. It is expected that from this study, (a) a guide for data collection will be established, and (b) evaluation procedures will be more consistent and standardized. This will be further described in chapter 3, Methods.

The Need within the Hospitality Industry

A core concept of all business development is the need to maintain or improve profit either through increasing revenue and/or lowering expenses. Profits in the hospitality industry are increased by pursuing both of these directives. The job of being a hotel reservations sales agent is crucial, and perhaps more important than ever before in the history of the industry. Traditionally, reservations agents were viewed and trained as order takers, simply handling the customers' requests for room rates and availability. As competition has increased, reservations agents have become "order makers," taking all the steps possible to get the customer to make a reservation (Farrell, 2005; Hospitality Services Alliance International, 2007). With increased competition and softening demand being experienced in the travel industry today, it is more important than ever to keep building loyal and returning guests. This study examined if revenue is generated from the

investment in the training, and how training can be used to create new revenues and provide services to the customers at the same time.

If HRD program evaluation of performance and organizational impact has been marginal, it has been even less marginal in the hospitality industry. It is believed that this study is one the few, perhaps the first, to gather and analyze data for all four levels of Kirkpatrick's evaluation model. Hence, this will be a unique study for the hospitality industry. This study contributed to the knowledge of training outcomes as measured by the learned knowledge and skills, job performance, and organizational impact. In addition, modifications of program design to target weaker performance areas in future training may result from this study. Finally, the results provided recommendations for the future utilization of different measurement variables for interpreting levels of performance in the hospitality industry.

The Need within the Body of AE/HRD Research and Theory

The current emphasis on accountability reveals a critical need to enhance knowledge and skills in the area of adult learning. To provide adult learners with an effective learning experience, more evaluation is needed on the impact of learning. Similarly, HRD professionals who develop programs to serve the growing employee population must address the issues influencing the effectiveness of HRD programs. A comprehensive study of learning acquisition in a training program over a period of time that addresses the employee's program perception, knowledge and skills learned, on-the-job performance, and organizational impact is needed. While there are numerous research studies on program evaluation and the adaptation of different evaluation models, limited information exists in the literature on the use of higher levels of evaluation and what

needs to be examined. The findings of this study contributed to AE/HRD research, theory of program evaluation, and how selected measurement variables translate into different levels of performance outcomes. In addition, the results can serve as basis for future program evaluation strategic planning and implications in different industries and academic environments.

Definition of Terms

The context of this study is a reservations call center for a leading international hotel. Therefore, the following terms utilized throughout this document are defined in order for readers to understand the evaluation measures of the study:

Average Daily Rate (ADR)

The average of all rates charged for all occupied guest rooms during one day of business. The method of computing the ADR is to add the total of all guest room revenues and divide that by the number of rooms sold (Feiertag & Hogan, 2001). The average annual figures used are reports provided by Smith Travel Research (Bowers, 2007; Freitag, 2006a, 2006b, 2006c; Lomanno, 2005), an international research company that collects and reports comprehensive performance data for the hospitality industry, and is considered the industry standard and index.

Average Processing Time per Call

The total processing time divided by the total number of calls received. The time was recorded and reported in seconds.

Average Talk Time per Call

The total talk time divided by the total number of calls received. The time was recorded and reported in seconds.

Book

To sell hotel space, either to an individual needing a room or to a group needing a block of rooms (Feiertag & Hogan, 2001).

Call Conversion Ratio

The total number of reservations booked divided by the total number of received calls (Hospitality Services of Alliances International, 2007).

Call Quality Assessment Score

The assessment of the reservations sales agents' knowledge and skills in handling calls. Utilizing the Hotel's scoring criteria (see Appendix B), the call center supervisors randomly review each reservations sales agent's recorded calls and conversations each month. The score is calculated on a 100-point scale.

Central Reservations Office (CRO) or Call Center

A central reservations office, or call center, typically deals directly with the public, advertises a central (usually toll-free) telephone number, provides participating properties with necessary communications equipment, and bills properties for handling reservations. They may also be called Central Reservation Services, especially if they represent independent operators or more than one brand as part of an affiliate reservation network for many hotels (Feiertag & Hogan, 2001).

Cost of Training

The total training costs are calculated by the sum of all the costs related to the training intervention. According to the Director of the Hotel's human resource department, the costs include training materials for each agent, the agents' wages, and the learning coach's (facilitator's) fee. Since the training sessions were conducted at the call

center, it was agreed by the Director that the costs of utilizing the facility were minimal. Therefore, these costs were excluded for calculating the total costs of training.

Cost of Training/Sales Ratio

A ratio calculated by dividing the costs of the training by the sales (the number of room nights times the ADR).

Organizational Impact

Level 4 of Kirkpatrick's evaluation model. At this level, organizations attempt to measure actual organizational change due to their training efforts, and determine a monetary value on those changes. Training programs targeted to increase sales, reduce accidents, lower turnover, decrease costs, or increase production can often be evaluated in terms of organization wide results (Kirkpatrick & Kirkpatrick, 2005, 2006).

Reservations Sales Agent

An employee who accepts, verifies and confirms lodging reservations (often by telephone), frequently using a computerized reservation system (Feiertag & Hogan, 2001).

Sales

The number of room nights times the average daily room rate (ADR).

Sales/Call Ratio

A ratio calculated by dividing the sales by the number of calls received.

Total Processing Time

The sum of time a reservations sales agent used to enter information received from a call, whether a reservation was made or not made, for the entire month. The time is recoded and reported in seconds.

Total Talk Time

The sum of time a reservation sales agent spends on the telephone conversations for the entire month. The time is recoded and reported in seconds.

Total Time Saved

The sum of the time every agent spent after the training program minus the sum of the time every agent spent prior to the training.

Total Wages Saved

The total time saved times the agents' average hourly wage.

Assumption of This Study

Despite all the reservations sales agents participating and completing the training sessions at different times, it was the same learning coach that facilitated all the sessions. Thus, to conduct this study, one important assumption is made: It is assumed that all the reservations sales agents received the same training from the same learning coach (facilitator).

Delimitations of This Study

This was a study to examine the utilization of Kirkpatrick's model, using collected data that enables its use to examine a training delivered within the parameters of all four levels of Kirkpatrick's model, and to also determine the inter-level relationship of the four levels. Because the study population came from only one hotel chain, the research results may not be generalized to other hotels with different operations, target customer segments, or geographical regions.

Summary

The purpose of this study was to examine Kirkpatrick's training evaluation model (Kirkpatrick & Kirkpatrick, 2006) by assessing a sales training program conducted at an organization in the hospitality industry. The conceptual framework, purpose, significance, and expected consequences of the study were introduced in this chapter. The next chapter will review the theoretical frameworks and empirical research in adult learning theories and training program evaluation as they apply to this study.

CHAPTER 2

REVIEW OF THE LITERATURE

This study examined the training evaluation model of Kirkpatrick (Kirkpatrick & Kirkpatrick, 2006) when applied to assess a sales training program conducted at an organization in the hospitality industry. The study assessed the employees' training outcomes on knowledge and skills, job performance, and the impact of the training upon the organization. This chapter reviews the related literature that addresses the theoretical frameworks and empirical research relevant to this study.

Theoretical Review

This theoretical review discusses the theories of the purpose and importance of HRD program evaluation, the challenges in conducting evaluations, and a review of selected evaluation models, with particular emphasis on Kirkpatrick's four-level evaluation model.

The Purpose and Importance of HRD Program Evaluation

Merrill Anderson, Chief Executive Officer at MetrixGlobal, wrote in the forward of Kirkpatrick's latest edition of *Evaluating Training Programs: The Four Levels* (Kirkpatrick & Kirkpatrick, 2006), "Every year new challenges emerge in the field of training and development – for example, competency development, outsourcing, e-learning, and knowledge management, to name a few. In spite of the variety and complexity of these challenges, there is a common theme: business leaders want to see value for their investment (p. ix)." This emphasizes the need for more information on the impact of adult learning programs and the emphasis on accountability (Barrow-Britton, 1997; DeVeau, 1995; Hart, 1992; Tung, 1998). Accordingly, adult educators and HRD

professionals who are responsible for programs to serve this growing population must recognize the effectiveness of the program at different performance levels (DeSimone & Harris, 2002; Werner & DeSimone, 2005).

HR efforts are not complete until the outcomes have been measured. However, among many most prominent evaluation theorists and/or researchers, their views of what evaluation is and how it should be carried out are differ widely.

Gilley, Eggland, and Gilley (2002) indicate "evaluation is a process, not an event, that involves all key decision-makers, stakeholders, and influencers, and should be influenced by a clear understanding of the organization's performance and business needs, as well as its strategic goals and objectives" (p. 381). According to Caffarella (1988), training program evaluation is "the process used to determine the effectiveness of the training activities and the results of those activities (p. 190)." Brinkerhoff (1981) defined training program evaluation as "systematic inquiry into training contexts, needs, plans, operations, and effects (p. 66)." HRD evaluation can also be defined as "the systematic collection of descriptive and judgmental information necessary to make effective training decisions related to the selection, adoption, value, and modification of various instructional activities (Goldstein, 1986, p. 237)." Evaluating the HRD effort means collecting and using information to make effective decisions about the choice, implementation, and follow-up of all development, education, and training efforts of an organization (Werner & DeSimone, 2005; Phillips, 1999; Phillips, 2003).

With such diverse definitions of what evaluation is, Kirkpatrick and Kirkpatrick (Kirkpatrick & Kirkpatrick, 2006) concluded that there are three general objectives or reasons to evaluate training: "(1) to justify the existence and budget of the training

department by showing how it contributes to the organization's objectives and goals, (2) to decide whether to continue or discontinue training programs, and (3) to gain information on how to improve future training programs (p. 17)."

Tanke (1999) further indicates that the short-term need for organizations to conduct evaluation of their training programs is to ensure that they provide employees sufficient knowledge and skills to performance their job, or change their behaviors or attitudes in order to improve productivity and/or efficiency. In addition to increasing productivity, higher job satisfaction, and improving work environment, the evaluation results can provide guidelines toward the organizational goals to ensure long-term success (Tanke, 1999).

Depending on the constitution or culture of the organization, educational and workplace evaluations usually have very different goals and purposes. According to Strunk (1999), "educational evaluations most often use a combination of summative and formative evaluation to render judgment about the value of the program being evaluated (p. 13)." The focuses are between the purpose, goals, objectives, roles, and uses of evaluation in academic settings. However, on the other hand, in today's competitive environment, for-profit organizations are more concerned with performance and the impact of training in the work place (Phillips, 1999; Strunk, 1999; Swanson, 2001; Van Buren, 2001).

As mentioned in the chapter 1, according to the American Society for Training and Development (ASTD, 2007), the success of organizations depends on the skills and capabilities of their employees. However, there is a growing gap between employee skills and today's job requirements. Organizations still struggle to find the right people with the

right skills. Most organizations recognize the problem by increasing their investment in training and development (Swanson, 2001). Until 2005, spending on training and development had been flat for several years (Rivera & Paradise, 2006). According to the *ASTD 2008 State of the Industry Report* (Paradise, 2008), organizations are now recognizing that, to sustain a competitive position, employee learning and skill development are more important to the business than ever before. ASTD estimated that U.S. organizations spent $134.38 billion on employee learning and development in 2007, with nearly two-thirds ($83.62 billion) spent on the internal learning function and the remainder ($50.77 billion) being spent on external consultant services (Paradise, 2008).

Despite the worst economic conditions in several decades, ASTD's latest 2009 report estimated that U.S. organizations still spent $134.07 billion on employee learning and development in 2008 (Paradise & Patel, 2009). The average annual expenditure per employee in the ASTD's sample organizations increased to $1,103 per employee in 2007, an increase of 6 percent from 2006 (Paradise, 2008). The finding in 2008 was slightly down 3.8 percent from the 2007 level to $1,068 (Paradise & Patel, 2009). Average expenditure per employee in the sample of "BEST" organizations, defined by ASTD as organizations that demonstrated enterprise-wise success as a result of employee learning and development, was $1,531 in 2006. The average number of hours of formal learning per employee in the sample organizations increased to 35.06 hours per employee in 2006. In the Best organizations, the average number of learning hours per employee rose from 36 in 2004 to 44.34 in 2006 (Paradise, 2007; Rivera & Paradise, 2006).

As Kirkpatrick and Kirkpatrick (2005) indicate, "the economy has been tight since late 1990s, and 9/11 only made things worse. Competition remains fierce.

Executives are looking everywhere for opportunities to generate income and cut costs. With that goes the need to increase training effectiveness and efficiency (p.11)." This is particular significant in travel and hospitality industry in general. Both business and leisure travels had been stalled due to the economic downturn, security concern, etc. Many companies have had to restructure their organizations and retrain their workforce in order to survive. Retaining highly skilled and productive workforce is critical to an organization's overall success, and even more crucial to the hospitality industry with high turnover rate (Newman & Hodgetts, 1998).

According to Delerno (2001), the cost for on-site (classroom) training for a hotel reservations department, the focus of this study, can range from $6,000 to over $10,000 and includes such costs as the instructor's travel expenses and the cost of the training course itself. This on-site training estimate assumes that only the instructor will incur travel expenses. If the course is taught in a cluster format, with reservations sales agents coming from other hotels, all of the participants will incur travel expenses with the exception of the participants from the host hotel. Lost productivity and revenue can actually be higher if classroom days include not only travel time, but also total time away from the office. Because on-site training is a live, one-time event, and the turnover rate of the front-office/reservations department is considerably high, more expenses are incurred when new hires are to be trained (Newman & Hodgetts, 1998). Many hotel companies failed to recognize the significance of reservations sales agents' contribution and their association with the companies' bottom-lines. For most reservations sales training programs delivered in the industry today, the fundamental concept is emphasizing the reservations agents' performances and their contribution to the bottom-lines (HSA,

2007). Honeycutt, Ford, and Rao (1995) indicated that a common problem faced by many companies is an inadequately trained sales force and the area in greatest need of additional research is the determination of sales training effectiveness.

The roles of HRD have changed. With increasing competition, investments in the HR endeavors, and the emphasis on the accountability, today one of the primary global trends in training is to show the organizational results/impact of the training investments (Phillips, 1999; Phillips, 2003a; Van Buren, 2001; Van Buren & Erskine, 2002). HRD functions have moved from producing competent workforce to achieving organizational impact (Benabou, 1996; Bomberger, 2003; Bushnell, 1990; Jackson, 1989; Shelton & Alliger, 1993). The issues surrounding organizational results/impact as a way to measure the contribution of HRD endeavors have received increasing attention (Brinkerhoff & Gill, 1994; Werner & DeSimone, 2005). Due to the increased needs and trends in the industry, ASTD established its latest Evaluation and ROI Community in 2002. In August 2002, ASTD affiliated with the ROI Network, an association of more than 500 practitioners of training evaluation with a specific interest in ROI evaluation. The purpose is to promote the significance of accountability. The network facilitates information sharing about effective measurement and evaluation research practices, particularly in the human and organizational performance improvement field, and how to disseminate these findings so as to foster organizational learning.

The Challenges in Conducting Evaluations

As stressed by Attia (1998), evaluation is a very essential and important phase of training. However, it is also the most neglected. Many organizations and HRD practitioners understand the importance of the training program evaluation, but various

challenges usually restrict its full implementation (Galvin, 1983; Phillips, 2000; Phillips, 2003b; Strunk, 1999).

As argued by many HRD professionals and organizations (Driscoll, 2001; Speizer, 2005), measuring organizational impact is very difficult, especially when establishing a relationship between training and an increase in profits. Many organizations fail to conduct an evaluation of the training investment within the framework of its contribution to profits (Setaro, 2001). The *ASTD 2002 State of the Industry Report* found that only one-third of companies profiled try to measure learning gained, and that only 12 percent or less try to measure job performance and business impact (Bersin, 2003). A 2002 Bersin and Associates study of more than 30 training organizations found that the leading reason companies failed to measure training more rigorously is not the lack of interest or importance, but rather they lack the experience, tools, and infrastructure to do so.

According to Larsen (1985), the reasons training evaluation is not taking place include limited time, resources and access to personnel for follow-up, HR personnel lacking the expertise to conduct effective evaluations, current methods are not useful and practical, training results are not measurable during the evaluation periods, and the lack of commitment from top management. Russ-Eft and Preskill (2001, p. 17) indicate more reasons why many organizations fail to conduct evaluation:

- Organization members misunderstand evaluation's purpose and role.
- Organization members fear the impact of evaluation findings.
- There is a real or perceived lack of evaluation skills.
- Evaluation is considered an add-on activity.
- Organization members don't believe the results will be used; data are collected and not analyzed or used.

- Organization members view evaluation as a time-consuming and laborious task.
- The perceived costs of evaluation outweigh the perceived benefits of evaluation.
- Organizational leaders think they already know what does and does not work.
- Previous experiences with evaluation have been either a disaster or disappointing.
- No one has asked for it (p. 17).

Phillips (1991) argued that these are myths and false assumptions about the training evaluation. Other false assumptions include some training cannot be quantitatively measured, there are too many variables affecting the behavior change other than training, and evaluating training programs is very expensive (Swanson, 2001).

Because of these myths and assumptions, program evaluation has long been focused on the employees' reactions to the program and learning and knowledge gained in the training, i.e., Kirkpatrick's Levels 1 and 2 (Alligar & Janak, 1989). Organizations are now aggressively searching ways to examine Level 3, the overall performance following the training, and Level 4, converting the performance to measurable organizational results, and in turn, determining the contribution of HRD to the organization. All these measures are intertwined and highly dependent upon each other (Kirkpatrick, 1998; Kirkpatrick & Kirkpatrick, 2006; Phillips, 1996a, 1996b, 1996c, 2003). Even the Kirkpatrick model, which has been known for almost 50 years, is often misunderstood and implemented with varying degrees of fidelity. For the few times results were examined, it was most often with technical training because of it being easy to measure (Hackett, 1997).

Sales training is more important in some industries than others. As stated by Attia (1998), a common problem faced by many companies is an inadequately trained sales

force. Because of the high need of collecting sales data and the difficulties encountered in measuring the effects of sales training, 57% of the surveyed sales training executives said that the area in greatest need of additional research is the determination of sales training effectiveness (Attia, 1998). Thus, a critical need is to determine if performance following the training, and assessing the performance contributes to measurable organizational results, and, in turn, the contribution of HRD to the organizational results (Honeycutt, Ford, & Rao, 1995).

A Review of Selected Evaluation Models

According to Posavac and Carey (1997), the overall purpose for program evaluation activities is contributing to the provision of quality services to people in need. Many program evaluation experts have developed various guidelines and models for determining the value of training interventions. Organizations and HRD professionals have a wide selection of evaluation guidelines and models to measure their training initiatives and calculate the value. Seven evaluation models will now be presented: Kirkpatrick's (Kirkpatrick & Kirkpatrick, 2006) four-level evaluation, Bushnell's (1990) IPO model, Stufflebeam's (1983) CIPP model, Warr, Bird and Rucham's (1970) CIRO model, Brinkerhoff's (1987) six stage model, Kauffman and Keller's (1994) five level model, and Holton's (1996) three level evaluation model.

Kirkpatrick's Four-level Evaluation Model. One of the most well-known and widely used models is articulated by Donald Kirkpatrick. Introduced in 1959, it has stood the test of critical review, gaining support over time to be one of the most widely accepted and influential models (Phillips, 2003). Kirkpatrick formed a logical framework

to examine results and impact from both individual and organizational performance perspectives (Setaro, 2001).

According to Kirkpatrick and Kirkpatrick (Kirkpatrick & Kirkpatrick, 2005), when the four levels of evaluation were first introduced in the late 1950s and early 1960s, HRD professionals were struggling with the concept of evaluation, as there was no common language and easy way to communicate what evaluation meant and how to accomplish it.

The conceptualization evolved from Kirkpatrick's doctoral dissertation in 1952. From November 1959 to February 1960, Kirkpatrick published a series of four articles, *Techniques for Evaluating Training Programs*, in the *Journal for the American Society of Training Directors* (Kirkpatrick & Kirkpatrick, 2006). Originally, Kirkpatrick used four steps to describe his theories. Soon in the industry and in the literature, HRD professionals referred to the four steps as four levels. They also began to accept his four levels, and it became recognized as the Kirkpatrick Model (Kirkpatrick & Kirkpatrick, 2005, 2006).

The model is the most well-known and utilized model for evaluating training programs. Not surprisingly, it has also been criticized over the past five decades (Alliger & Janak, 1989; Brinkerhoff, 1987; Bushnell, 1990; Hilbert, Preskill, & Russ-Eft, 1997; Holton, 1996; Kraiger, Ford, & Salas, 1993; Spitzer & Conway, 2002; Swanson, 2001). Despite these criticisms, and the development of other comprehensive evaluation models, Kirkpatrick's model is still being widely utilized due to its simplicity and practicality (Kirkpatrick & Kirkpatrick, 2006; Twitchell, 1997).

Kirkpatrick contends that training can be evaluated using four criteria or levels of evaluation: reaction, learning, job performance, and organizational impact (Kirkpatrick & Kirkpatrick, 2006). Galvin (1983) identified the four levels as reaction, learning, behavior, and results (RLBR). From individual to organizational performance, the four levels represent a sequence or continuum of complexity. Moving from one level to the next, the evaluation process becomes more difficult and time-consuming, but it also provides increasingly more valuable information.

At Level 1, the focus is on the learner's perceptions about the program and its effectiveness. The measurement instruments usually request comments about the training content, materials, instructors, facilities, delivery methodology, etc. This is important because positive reactions to a training program may encourage employees to attend future programs. In contrast, negative comments about the program may discourage learners from attending and/or completing the program. The negative comments can be used to modify the program and to ensure organizational support for the training program. Because favorable reactions to training do not, by itself, guarantee that learning (Level 2), performance (Level 3) has occurred, Kirkpatrick stressed that many organizations and HRD professionals are overlooking the importance of Level 1 evaluation (Kirkpatrick, 1959a; Kirkpatrick & Kirkpatrick, 2005, 2006).

Kirkpatrick's Level 2 is content evaluation, the examination of what employees learned in the training program. Kirkpatrick defined learning "as the extent to which participants change attitudes, improve knowledge, and/or increase skill as a result of attending the program (Kirkpatrick & Kirkpatrick, 2006, p. 22)." Although research does not support that acquired knowledge and skills equates to the behavioral changes or on

the job performance (Strunk, 1999), it is also evident in the literature that Level 2 evaluations is still one of the most popular forms to evaluate the effectiveness of training programs (Bersin, 2003). By implication, HRD professionals need to prove that the employees acquired knowledge and skills from the training demonstrates the worth of the program. As Kirkpatrick and Kirkpatrick stressed (Kirkpatrick & Kirkpatrick, 2006), "Evaluating learning is important. Without learning, no change in behavior will occur (p. 50)."

Level 3 measures employees' job performance by determining the extent to which employees apply their newly acquired knowledge and skills on the jobs (Kirkpatrick, 1960a). This level of evaluation is critical, as it addresses the issue of learning transfer. If employees do not apply what they learned to their job, the training effort cannot have an impact on the organizational results (Level 4). No final results can be expected unless a positive change in behavior (performance) occurs. According to Kirkpatrick and Kirkpatrick (Kirkpatrick & Kirkpatrick, 2006), evaluation of the behavior is more complicated, difficult, and time-consuming than the reaction and learning evaluations (Levels 1 and 2). Consequently, as Kirkpatrick stated: "I believe that level 3 is the forgotten level. Lots of time, energy, and expense are put into levels 1 and 2 by training professionals because these are the levels that they have the most control over. Executives are interested in level 4, and that is as it should be. That leaves level 3 out there on its own with no one really owning it (Kirkpatrick & Kirkpatrick, 2006, p. 83)."

Level 4 is the most important and also the most challenging level to assess (Werner & DeSimone, 2005; Kirkpatrick, 1960b; Kirkpatrick, 1998; Phillips, 1996a). Typically at Kirkpatrick's Level 4, organizations search the business results for their

training efforts. At this level, organizations attempt to measure actual organizational changes due to training and determine a monetary value on those changes. Programs that target to increase sales, reduce accidents, lower turnover, decrease costs, or increase production can often be evaluated in terms of results (Kirkpatrick & Kirkpatrick, 2005, 2006). It should be emphasized that many HRD professionals recognized Phillips' return on investment (ROI) theory and considered it as Phillips' ROI model (Phillips, 1999; 2003). However, in essence, the Phillip's ROI framework is built upon Kirkpatrick's model only by its expansion of the fourth level, and identifies a fifth level that tries to further determine the organizational benefits of training by converting training results to monetary values and comparing them with the cost of training to obtain the true return on the training investment, or ROI. This is evident in Lockwood's (2001) research where she addressed ROI as part of the Kirkpatrick's model.

Critique of Kirkpatrick's Evaluation Model. Since Kirkpatrick introduced his four-level evaluation model in 1959, there have been many discussions about the model and the four levels. In Alliger and Janak's (1989) 30-year review of articles evaluating the model three major problematic assumptions were identified. The first assumption is that the levels are arranged in ascending order and the model is hierarchical in nature. Therefore, the higher levels are more valuable and important than the lower ones. With this notion, many HRD professionals purport to skip the lower levels of evaluations and focus on the higher levels of evaluations. This is questionable, as will be shown in the empirical review later in this section, few reported studies have addressed Levels 3 and 4. Also, Kirkpatrick (1959a; Kirkpatrick & Kirkpatrick, 2005, 2006) contends that it is a serious mistake to bypass Levels 1 and 2 evaluations and only conduct Level 3 and 4

evaluations. This will easily lead to the wrong conclusions about the effect of each level and the training program's overall result.

The second assumption is that the four levels of evaluation are causally linked. Based on this assumption, many researchers and HRD professionals presume that positive reactions are prerequisite for learning to occur. Once learning has occurred, desired behaviors will change and ultimately lead to organizational results (Alliger & Janak, 1989; Hilber, Preskill & Ress-Eft, 1997; Kirkpatrick & Kirkpatrick, 2005, 2006).

The second assumption leads to the third assumption, that the four levels are positively intercorrelated. If these two assumptions were true, it would be sufficient just to evaluate whether employees have positive reactions (Level 1) to the training program, from which it could be assumed they learned from the training, they ultimately would improve their job performance, and positively contribute to the organizational results. Addressing these assumptions, Kirkpatrick (1959a; Kirkpatrick & Kirkpatrick, 2005, 2006) emphasized that there is no guarantee that a favorable reaction to the training program assures learning, positive behavioral change, and favorable organizational results. This is why it is important to evaluate both reaction (Level 1) and learning (Level 2) in case no change in behavior (Level 3) occurs. "Then it can be determined whether the fact that there was no change was the result of an ineffective training program or of the wrong job climate and lack of rewards (Kirkpatrick & Kirkpatrick, 2006; p. 24)." By evaluating both Levels 1 and 2, it also makes employees more accountable for their own learning and performance.

Having examined Kirkpatrick's model and examined its criticisms, following are six other evaluation models. As these are presented, it would be beneficial to keep in

mind Kirkpatrick's four levels of evaluation, noting how many of these models are variations of the four levels.

Systems approach or input, process, output (IPO) model. Bushnell (1990) contends that Kirkpatrick's model focuses only on what happens after the training but not the entire training process. Therefore, Bushnell's evaluation model is more similar to many systematic instructional design models. The IPO Model is the acronym of the initials of the three stages of the model – input, process, and output. The input stage contains all the elements that may impact the effectiveness of the training, such as trainer competency, training materials, facilities, and equipments. In the process stage, the trainer plans, designs, develops, and delivers the program. The output stage, or short-term benefits, actually consists of Kirkpatrick's first three levels – participant reaction, knowledge gained, and improved job performance. Bushnell includes Kirkpatrick's fourth level, identifying it as long-term benefits to the organization's bottom-line, which include profitability, customer satisfaction, productivity, etc.

The IPO Model combines elements of Kirkpatrick's four-level Model and Brinkerhoff's six-stage Model (1987), discussed below. It is the model IBM (International Business Machines) utilizes with their corporate training programs. As Bushnell (1990) indicates, the organizations that use this model can easily determine whether the training programs meet their goals, what kinds of changes are needed for program improvement, and whether trainees actually acquired the needed knowledge and skills (Bushnell, 1990; Galvin, 1983; Phillips, 2000). Bomberger (2003) claims that the IPO model provides both formative and summative information, and it also goes beyond the Kirkpatrick model, attempting to show the worth of training in financial terms.

Context, input, process, and product (CIPP) model. As chair of The Phi Delta Kappa National Study Committee, Stufflebeam (1983) developed a model to improve curriculum evaluation throughout the field of education. His model is commonly known as the CIPP model, an acronym for the four types of decision-making factors – context, input, process, and product. Context refers to the decisions to determine objectives and goals. Input refers to structuring and designing the program. Process focuses on the implementation of the program, and Product refers to the outcome of the programs. Stufflebeam's evaluation model is similar to many instructional design models based on the ADDIE (analysis, design, development, implementation, and evaluation) framework (Dick & Carey, 1996).

Context, input, reaction, and outcome (CIRO) model. Warr, Bird, and Rackham (1970) presented another four-level framework, which consists context, input, reaction, and outcome (CIRO). Context evaluation involves obtaining information about the current situation to determine training needs and objectives. This is similar to context evaluation in the CIPP model. Input evaluation involves obtaining information about possible training resources, and is also similar to input phase of the CIPP model. Reaction evaluation involves obtaining information about the participant's reactions to improving the training process, and is similar to Kirkpatrick's Level 1, reaction evaluation. Outcome evaluation involves obtaining information about the results or outcomes of the program. This outcome phase has three different levels: immediate, intermediate, and ultimate outcomes, and are similar to Kirkpatrick's levels of learning, behavior, and results (Phillips, 2003).

It is easy to identify the great similarity between CIRO model, CIPP model, and Kirkpatrick's model. According to Galvin (1983), "systems theory is useful to analyze and synthesize these approaches to evaluation. The CIRO model is a simplified approximation of how a synthesis of the RLBR and CIPP model might appear." However, Warr, Bird, and Rackham (1970) stated that ultimate outcome evaluation does not always need to be used. This is a different emphasis than the current trend in the field of HRD, which is wanting to focus on the results level evaluation.

Brinkerhoff's six-stage evaluation model. As a proponent of systematic evaluation, Brinkerhoff (1987; Brinkerhoff & Gill, 1994) advocated circular evaluation by measuring all the instructional design elements. The Six-Stage Evaluation Model starts with needs assessment and identifies the goals of training. Stage two evaluates the program design, and stage three evaluates program implementation, which is similar to Kirkpatrick's Level 1 evaluation. Stage four evaluates the learning, and is identical to Kirkpatrick's Level 2. Stage five evaluates behavior, and is similar to Kirkpatrick's Level 3 evaluation. Stage six evaluates how much learning transferred to the results, as does Kirkpatrick's Level 4.

Similar to Bushnell's criticism, Brinkerhoff (1987) criticized Kirkpatrick's model contending that it lacks the examinations of the instructional design functions of needs analysis, instructional planning and development, implementation, etc. However, as identified by Bomberger (2003) and Phillips (2003), Brinkerhoff's model is also similar to Kirkpatrick model, although he adds an additional stage 1 to address and evaluate the instructional design functions, which are collectively called goal setting or needs analysis.

Kaufman and Keller's five levels of evaluation. Like most of the models mentioned to this point, Kirkpatrick's model is reflected in Kaufman and Keller's (1994) five levels of evaluation. Level one was expanded to include enabling and reaction. Level 2 is acquisition, Level 3, application, and Level 4 organizational outputs. Level 5 is the evaluation beyond the organization, and examines the extent to which programs enhance society and the environment surrounding the organization (Kauffman, Keller, & Watkins, 1996; Phillips, 2003a; Werner & DeSimone, 2005). While it was their intent to improve on Kirkpatrick's model, Kauffman and Keller's (1994) five levels evaluation are still aligned with Kirkpatrick's four levels, with just the addition of the expansion of Level 1 and the addition of Level 5 that examines consumer satisfaction and societal impact (Bledsoe, 1999).

Holton's three levels HRD evaluation and research model. Of Kirkpatrick's critics, Holton (1996) is the most critical, claiming that Kirkpatrick failed to specify the causal relationships between the four levels. He suggests that rather than a model, Kirkpatrick's work represents a taxonomy or classification. It lacks the research necessary to further the theory of evaluation. As Holton claimed (1996 & 2005), theories or models generally have a complete set of objects, relationships, influencing factors, hypothesis, predictions, and limits of generalization.

Holton's model identifies three outcomes of training – learning, individual performance, and organizational results, all of which are still similar to Kirkpatrick's Levels 2, 3, and 4. The missing element is the first level, reaction (Holton, 1996; 2005). Holton stressed that reactions should not be considered a primary outcome of training. He believed that favorable reactions and learning are not necessarily related (Holton, 1996;

Holton & Naquin, 2004). His model shows reaction as influencing the learning outcome; thus, its influence is not entirely disregarded. As Bomberger (2003) stated, "Holton's model shows the expected outcome from training and the influences that promote or inhibit them. It is a good addition to the roster of training evaluation models since it identifies several variables known to affect effectiveness of a training program. However, it has not been used nearly as widely as the Kirkpatrick model (p.22)." This reflects in one of Holton's recent studies. After almost a decade later criticizing Kirkpatrick's model, Holton (2005) indicated that "unfortunately, a full test of Holton's model has not been possible because tools to measure the constructs in the model did not exist (p. 37)."

The similarities and differences of the selected evaluation models. While well received and popular, the Kirkpatrick model is often challenged by other training evaluation scholars, researchers, and practitioners. Some researchers further developed their own models. Bushnell (1990) contends that Kirkpatrick's model focuses only on what happens after the training but not the entire training process. Similar to Bushnell's claim, Brinkerhoff (1987) identifies needs assessment, planning, and implementation as these training processes. Kraiger, Ford and Salas (1993) contend that Kirkpatrick's model fails to specify what kinds of changes can be expected from the HRD program, and what assessment techniques should be used to measure learning at each level. Similar remarks were made by Hilbert, Preskill, and Russ-Eft (1996). They indicated that Kirkpatrick's model lacks of diagnostic capability and the inability to account for factors that may affect outcomes of each level of evaluation. Alliger and Janak (1989) claimed that Kirkpatrick's model is misleading, with users easily accepting the notion that by achieving the outcomes in the higher levels assumes the achievement of outcomes at the

lower levels. Spitzer and Conway (2002) criticized the framework indicating it is conceptual and lacks the tools to increase business results. Phillips (1998) believes that Kirkpatrick did not adequately elaborate the fourth level and adds a fifth level that evaluates the cost benefit, or the return on the investment (ROI) in training. Of all these criticisms, Holton (1996) is the most critical, claiming that Kirkpatrick's four-level framework is incomplete as a model, and he fails to specify the causal relationships between the four levels. Spitzer and Conway (2002) also suggested that the process does not recognize the disconnection between behavior (Level 3) and impact (Level 4).

Researchers have categorized the frameworks and models based on their respective foci of evaluation. Some argue that Kirkpatrick's model is conceptual, defining it as a framework or taxonomy (Alliger & Janak, 1989; Holton, 1996; Spitzer & Conway; 2002). Others contend that Kirkpatrick's model only focuses on the outcomes, evaluating what happens after the training intervention. As Galvin (1983) pointed out, the RLBR model, as he referred to Kirkpatrick's model, is outcome and objective-oriented and focuses on determining the effectiveness of a program. In other words, it is a summative evaluation model, which only takes place after the training program has been conducted in order to assess the merit and worth of the training program, and provide a summary report of the training outcomes for consideration of its continuation and/or its improvement.

On the other hand, because instructional system design (ISD) theorists Mager (1984), and Dick and Carey (1996) incorporate the evaluation process into every aspect of their training model, some contend the ISD model is an evaluation model. This is evident in the IPO, CIPP, CIRO, and Brinkerhoff's models, which are, at least in part,

evolved from the ISD fundamentals (Tidler, 1999). These models are considered formative evaluation (Dick & Carey, 1996; Galvin, 1983; Phillips, 1991), and are procedure and process-oriented, focusing on providing information to make decisions about the entire scope of the curriculum development process. However, as argued by Kirkpatrick, based on the evaluation results, decisions to continue or alter the training program can be made accordingly. The summative evaluation results can turn into formative evaluation for future program improvements and/or modifications (Kirkpatrick & Kirkpatrick, 2006).

Despite all the criticisms of Kirkpatrick's model, and how researchers try to differentiate their models from Kirkpatrick's, most of the evaluation models found in the literature are generally based upon the original four levels (Bomberger, 2003; DeSimone & Harris, 2002; Werner & DeSimone, 2005; Goldwasser, 2001). The seven training evaluation frameworks or models presented in this section represent a similar framework – the use of levels or categories by which to report training data. The following figure (Figure 1) illustrates the categorizations of evaluation and the relationships between Kirkpatrick's model and the training program of this study.

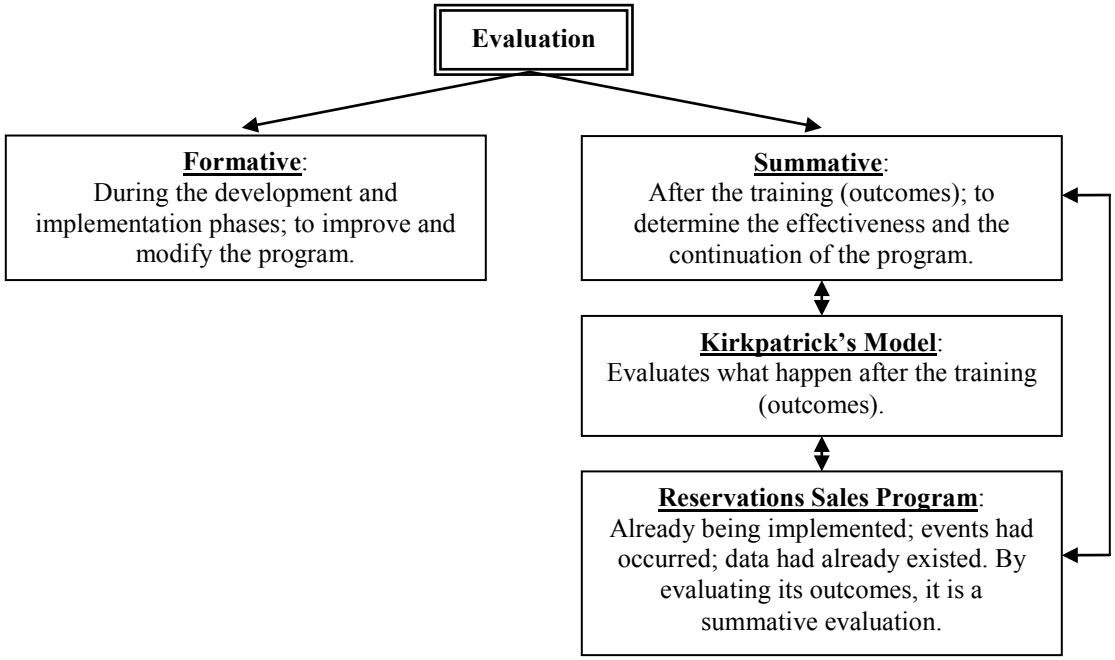

Figure 1. The Categorization of Evaluation & the Relationships between Kirkpatrick's Model & the Training Program in This Study

Summary of the Theoretical Review of the Literature

Fifty years after Kirkpatrick conceptualized his four-step approach to evaluation, the discussions of its utilizations and assumptions still dominate the literature. It has since been called a model, a system, a framework, taxonomy, a methodology, typography, and a vocabulary. The four steps have been called stages, criteria types, categories of measures, and, most commonly, levels of evaluation (Hillbert, Preskill, & Russ-Eft, 1996). Many evaluation theorists and researchers have critiqued and criticized Kirkpatrick's model, especially the lack of research to support its utilization at the higher levels (job performance and organizational impact).

In response to Holton's harshest criticism, Kirkpatrick stated, "I don't care whether it's a model or taxonomy as long as training professionals find it useful in evaluating training programs (Kirkpatrick, 1996, p. 55)." Kirkpatrick further stressed, "the model remains essentially the same. The concepts, principles, and techniques are as applicable today as they were when the model was first introduced. I am still getting requests from universities and professional and private organizations to present the four levels in keynote addresses at their conferences (Kirkpatrick & Kirkpatrick, 2005; p. 4)."

Kirkpatrick's model is still well received and being adapted in different disciplines. Organizations like AT&T, Motorola, Intel, Cisco, The Gap, First Union National Bank, Kemper Insurance, Duke Energy, the City of Los Angeles, St. Luke's Hospital, and the University of Wisconsin's Management Institute have been utilizing Kirkpatrick's model for evaluating their HRD endeavors (Kirkpatrick & Kirkpatrick, 2005, 2006). The book, *Evaluating Training Programs: The Four Levels*, has been a best seller and has been translated into Spanish, Polish, and Turkish (Kirkpatrick & Kirkpatrick, 2006).

At the ASTD national conference in 2004, Kirkpatrick was given the Lifetime Achievement Award in Workplace Learning and Performance for his body of publications and research that has had a significant impact on the field of workplace learning and performance. The power of Kirkpatrick's model is its simplicity and its ability to help people easily understand the concepts of training evaluation. Although the Kirkpatrick model is recognized as the being influential, it still has not been widely implemented in its entirety after a half of century (Alliger & Janak, 1989; Kraiger, Fords & Salas, 1993; Phillips, 2003). Training program evaluations are still stalled at the

reaction (Level 1) and learning (Level 2) levels (ASTD, 2009; Bassi, Benson, & Cheney, 1996; Bersin, 2003; Bomberger, 2003; Bromley & Kitson, 1994; Paradise & Patel, 2009; Plant & Ryan, 1992).

As Kirkpatrick and Kirkpatrick (Kirkpatrick & Kirkpatrick, 2006), Brinkerhoff (1989), and other researchers expect HRD practitioners to continue to evaluate their programs, there is a need for more empirical research to test more innovative methods and approaches that can be utilized to measure all four levels, especially the behavior (Level 3) and results/impact (Level 4).

Empirical Review

In this section reviews the empirical literature research of studies germane to this study, the utilization of Kirkpatrick's four-level evaluation model, are reviewed. Table 1 presents the 14 studies that are reviewed, listed in alphabetical order of the researchers, followed with the purpose of the study, the population of each study, data collection procedures, and salient results.

Table 1

Studies Related to Training Evaluation Utilizing Kirkpatrick's Model

Author	Purpose and Methodology	Subjects	Data Collection Techniques	Selected Salient Results
Attia, A. M. (1998)	To enhance the understanding of current sales training evaluation practices, to provide an example companies can utilize to evaluate sales training effectiveness, and to propose and test a model for evaluating sales training programs' effectiveness; experimental design, but not random assignment	59 trainees & 42 non-trainees, for a total of 101 sales supervisors of one large multinational company operating in Egypt	Surveys (Level 1); pre- posttests (Level 2); self-evaluation and supervisory-evaluation (Levels 3 & 4 – sales, sales/quota); staff/ management analysis (trainer's evaluation of trainees and utility analysis)	No differences were found between anonymous and non-anonymous responses; behavior change (Level 3) was significant from pretest to posttest but insignificant between experimental and control groups; the trainer's evaluation of trainees and the utility analysis are two complementary techniques that were found to be useful when conducted in conjunction with the Kirkpatrick's model; ROI = $17:1 over 5 years
Bledsoe, M. D. (1999)	To examine the Kirkpatrick model at each of the four levels as they related to corporate computer training courses	69 employees of a Midwest financial organization	Satisfaction survey (Level 1); pre & posttests (Level 2); 2 weeks after a self-report survey (Level 3); 2 weeks after surveyed supervisors (Level 4)	Moderate relationship between Levels 1 & 3; weak relationship between Levels 1 & 4; weak relationship between Levels 3 & 4
Bomberger, D. W. (2003)	To determine how three human service organizations in a large bureaucracy determine what training is needed by their staff, and how evaluation criteria are selected for evaluating training programs within those organizations; qualitative case study	3 large Pennsylvania State Department	Interviews, review of documentations, observation	None of the programs evaluated beyond Level 1; all satisfied with their current training methods, no plans of changes

(table continues)

Table 1 (continued)

Author	Purpose and Methodology	Subjects	Data Collection Techniques	Selected Salient Results
Galvin, J. C. (1983)	To determine if any relationship exist between the model of evaluation preferred by training specialists and their attitude toward valuation of management education; survey research	300 of the ASTD members (80% response rate)	Mail questionnaire	More training specialists preferred the CIPP model over the RLBR model; those who preferred the RLBR model had a more favorable attitude toward evaluation of management education
Kim, I. Y. (2006)	To implement an evaluation of a church instructor training program	269 of the 405 church members in Seoul, Korea	Initial & follow-up surveys	Most participants satisfied with the program; motivation, perceived changes (self-efficacy), increased; knowledge/skill gained; knowledge/skills acquisition was inversely related to a plan implementing the training
Lanigan, M. L. (1997)	To determine how well the Theories of Reasoned Action and Planned Behavior explain and predict training outcomes assessed by measures of the three levels of Kirkpatrick's model	214 new students of the Indiana University's MBA program	Surveys right after the training; 3 weeks later an email follow-up behavioral survey	Theory of Planned Behavior is the most appropriate theory to support the Kirkpatrick model; high correlation between attitude and self-efficacy; self-efficacy is the strongest predictor of behavior
Larsen, N. B. (1985)	To assess how useful and practical the success case method is for evaluation; success case method meta-evaluation	A Fortune 500 company – 9 success case trainees was selected from the population of 37 health care administrators	Daily reaction sheets; true/false pre-post tests; interviews	Estimated total costs of SCM is about 2% of the total budget of the training
Lockwood, S. L. (2001)	To diagnose, design, implement, and evaluate an orientation program; action research	103 staff of San Diego District Attorney's Office	Focus groups; survey (level 1); test (level 2); interviews (level 3); review of budget (level 4)	70% of the trainees scored 90% or higher at level 2; 200%ROI

(table continues)

Table 1 (continued)

Author	Purpose and Methodology	Subjects	Data Collection Techniques	Selected Salient Results
Phillips, J. H. (2000)	To investigate evaluation practices and processes used by companies to measure training results; qualitative design	8 training directors and instructional designers/trainers of five large companies	2 phases of interviews	Training directors emphasize on level 3; instructional designers/ trainers focus on levels 1 &2; all 5 organizations use Kirkpatrick's model but mainly focus on levels 1 & 2, rare on level 4
Strunk, K. S. (1999)	To determine the status of and barriers to financial impact evaluations in employer-sponsored training programs	ASTD group: random 1000 members (153 returned, 15.3%); ROI Network group: all 110 members (33 returned, 30%)	A national survey	98% evaluated Level 1; 88% evaluated Level 2; 76% evaluated Level 3; over 50% evaluated Level 4; time constraints, complexity of analysis, lack of support for the process, cost are main barriers to impact evaluation
Tidler, K. L. (1999)	To determine if CME (continuing medical education) training could be evaluated using Kirkpatrick's four levels of evaluation; historical research	84 healthcare professionals (only 21 of these are physicians) at one southwestern healthcare institution	Questionnaires (Levels 1 & 2); pharmacy and billing systems (Level 3); charges for treatment (Level 4)	The high correlation between Level 1 and Level 2; no changes in Levels 3 and 4
Wertz, C. (2005)	To determine if the current CLAD training teachers are receiving is making a difference in their classrooms, and if so, what kind of difference; primarily a qualitative study	17 K-12 teachers in 3 northern California counties; 25 teachers & 12 administrators	Pre & posttests; interviews	Positive changes from Levels 1 to 4; mix perceptions between teachers and administrators about supports

(table continues)

Table 1 (continued)

Author	Purpose and Methodology	Subjects	Data Collection Techniques	Selected Salient Results
Yaw, D. C. (2005)	To examined the effectiveness of e-learning in the industrial setting at Level 3 based upon the Kirkpatrick model and compared e-learning to traditional classroom learning; experimental design	200 production employees	Posttest for Level 1; pre & posttests for Level 2; supervisor focus groups & incident reports pre & post training for Level 3	No significant difference between the post-test scores of e-learners & classroom learners; a significant difference between the pre-test and post-test of the e-learners; no significant differences between the pre-test & post-test of the classroom learners; no significant difference between the two groups at Level 3

Surveys of the Use of Kirkpatrick's Model

According to ASTD, the areas that separate leading-edge from average organizations are a high performance workforce, the number of employees trained, training expenditures, outsourcing, course topics, delivery methods, and review and evaluation (Bassi & Van Buren, 1999). Organizations that are willing to make a greater financial investment are shown to train a larger percentage of employees, have higher rate of spending per employee, and have a greater use of innovative training practices. The *ASTD 2002 State of the Industry Report* found that only one-third of companies profiled try to measure learning (Level 2) gained, and that 12% or less try to measure job performance (Level 3) and business impact (Level 4; Bersin, 2003). This finding yet again proves the lack of implementation of Kirkpatrick's model in all four levels. These findings are supported by several empirical researches.

Strunk (1999) surveyed 1,000 randomly selected ASTD members and all 110 ROI (Return of Investment) Network members attempting to determine the status of and barriers to financial impact evaluations in employer-sponsored training programs. Her study revealed that 98% of the organizations evaluate at Level 1, 88% evaluate at Level 2, 76% evaluate at Level 3, and over 50% evaluate at Level 4. The significant difference was ROI Network members were more likely to use both Level 2 and Level 4 evaluations.

Similar to Strunk's study, P.P. Phillips (2003) sought to gain the understanding of training evaluation practices in public sector organizations. Her samples consisted of public sector members of the ASTD representing public sector organizations (excluding consultants, training suppliers, and professors), and human resources (HR) directors/managers/staff with responsibility for training and training directors/managers/staff who are members of the International Public Management Association for Human Resources (IPMA-HR). From 523 responded survey questionnaires, public sector organizations evaluate training predominantly at Level 1 (72.18%) and Level 2 (31.65%). The typical methods for conducting these types of evaluation are the end-of-course questionnaire (Level 1) and facilitator/instructor assessment (Level 2). There is use to some extent of Level 3 (20.42%), Level 4 (12.21%), and, using J.J. Philips (2003) model, Level 5 ROI (5.25%) evaluation. Large organizations (federal agencies) tend to evaluate at all levels, while small organizations (county, municipal, city/local) evaluate at Levels 1 and 2. In general, there was less use of the five levels of evaluation in public sector organizations as compared to private sector organizations (P.P. Phillips, 2003).

With the intension to explore organizational practices and processes at the impact level, J. H. Phillips (2000) interviewed eight training directors and instructional designers/trainers for five large organizations. These five organizations have at least 5,000 employees and a training department consisting of at least five instructional designers/trainers. They represent different businesses – property/casualty insurance, banking, automotive manufacturing, health care, and furniture manufacturing. The findings indicated that training directors emphasized Level 3 (job performance), but instructional designers/trainers focused on Levels 1 (reaction) and 2 (learning). All five organizations used Kirkpatrick's model, but mainly focused on Levels 1 and 2, but rarely on Levels 3 or 4. The results further indicated that the main methods of conducting a fourth level evaluation were a discussion with the manager or a survey (Phillips, 2000).

Similar to Phillips' (2000) study, Bomberger (2003) examined three of the larger departments in the Pennsylvania State governments' training functions as case studies. The purpose was to determine how the training staffs decide what training is needed and how evaluation criteria were selected for evaluating training programs within those organizations. Bomberger first interviewed the staff, and then reviewed evaluation documents of each organization. Bomberger then participated in one training activity conducted by each department. All the departments were satisfied with their current evaluation methods and none of them were planning to change their evaluation process. However, they all voiced needs for improvement and admitted that they were not evaluating to determine if the training was effective. None of the organizations evaluated their training programs beyond Level 1 of the Kirkpatrick Model. The organizations seemed to give little thought as to models for evaluation and methods that accompany

these evaluation models. All of the organizations used a participant satisfaction feedback form to obtain feedback from the training, but none went beyond the reaction level (Level 1). Two of the organizations asked participants to indicate what they perceived they have learned or what they thought they would do to use the newly acquired information. However, no one measures what the participants actually learned (Level 2).

Limited Utilization of Kirkpatrick's Model

In a limited experimental study, Yaw (2005) designed, developed, and delivered a safety-training program to the 200 production employees at the ZF Boge Elastmetall-Paris (France). The same curriculum was presented to both e-learning and classroom groups. The pre-test was administered identically to both groups two weeks prior to the training to determine the trainee's previous knowledge about the safety training. Upon completion of the training program, Levels 1 and 2 evaluations were administered to each group of learners. Level 3 evaluation was administered one month following the training in order to assess if there was a behavior change in the workplace. The Level 3 evaluation consisted of supervisor focus groups and a comparison of the number of safety incidents for the one month post-training to one month pre-training. There was a significant difference in the pre-test assessment of e-learners and classroom learners. However, there was no significant difference between the post-test scores of e-learners and classroom learners. For the e-learners, there was a significant difference between the pre-test and post-test scores indicating that learning did occur. For classroom learners, there were no significant differences between the pre-test and post-test scores. For Level 3, there was no significant difference between the two groups.

From an academic setting, Lanigan (1997) studied 214 new students enrolled in the Indiana University's MAB program who attended the email/computer training program. The main objective of Lanigan's study was to select an appropriate theory to support the Kirkpatrick model by uncovering the particular variables that predict behavior. The finding suggested that the Theory of Planned Behavior is the most appropriate theory to support the Kirkpatrick model, and perceived control enhances the prediction of actual behavior. Additionally, it also confirmed that the Kirkpatrick model is hierarchical in nature and the levels are sequential. Level 1 is the lowest level on the hierarchy. While Level 1 data can predict Level 3 outcomes, the prediction is enhanced by Level 2 data. Moreover, the prediction of behavior is further enhanced by adding the one item control measure to the Level 2 data. As a result, Lanigan suggested that the Kirkpatrick model should be modified so that the perceived control variable is added within the hierarchy as a new Level 3. The new model would include five levels as Level 1 reactionnaire, Level 2 change in learning, Level 3 perceived control factors, Level 4 behavior on-the-job, and Level 5 return on investment.

Studies Examining Barriers to Utilizing the Higher Levels of Kirkpatrick's Model

A 2002 study by Bersin and Associates of more than 30 training organizations found that the leading reason companies failed to measure training more intensely is not the lack of interest or importance, but rather they lack the experience, tools, and infrastructure to do so. These findings are supported by several more rigorous empirical studies.

Strunk's (1999) survey, cited previously, wanted to determine the status of and barriers to financial impact evaluations in employer-sponsored training programs. Her

study revealed that time constraints, complexity of analysis, lack of support for the process, cost (too expensive), of little value (not necessary), and not familiar with the higher level processes are the main barriers to organization impact evaluation (Level 4). Misunderstanding of what constitutes financial impact evaluations continues to be a concern. These issues were also echoed by P. P. Phillips (2003), also cited previously, who stressed in her research that even with the increased emphasis on the higher levels, training evaluation is still predominantly conducted at Levels 1 and 2. This is primarily due to the costs, lack of training, and the fact that higher levels of evaluation are not required. Barriers to training evaluation within public sector organizations are similar to those barriers that prevent evaluations in other organizations (Phillips, 2003). Similar barriers were also found in the studies conducted in healthcare, and business (Hill, 1999; Twitchell, 1997).

Many comments by respondents in P. P. Phillips' (2003) study indicated that small staffs and limited resources prohibited the pursue of training evaluation. This was confirmed by the correlation between the use of the five levels of evaluation and the percentage of training staff involved in evaluation. All five levels have a relationship at the .01 level of significance with the percentage of training staff involved in evaluation. Within public sector organizations there was a relationship between manager experience and use of training evaluation. Significant relationships exist between job title and Levels1 and 4 evaluation, job function and Levels 1 and 2 evaluation, number of years in training and Level 4 evaluation, and academic preparation and Level 5 evaluation. Significantly higher levels of evaluation were conducted at all five levels when an evaluation policy is in place than when it is not (Phillips, 2003).

Galvin (1983) studied the relationship between preference for the RLBR or CIPP model of evaluation, and attitude toward evaluation of management education in corporations among training specialists. By selecting 300 samples from ASTD's member directory, mail questionnaires were sent out to collect the data and 80% were returned. The results indicated that more training specialists preferred the CIPP model to the RLBR model. Those who prefer the RLBR model had a more favorable attitude toward evaluation of management education. In addition, the study indicated that misconceptions of evaluation are common and often acted as the barrier to the initiation and implementation of evaluation.

In Bomberger's (2003) case study, previously cited, he also found some barriers to conducting at higher levels of evaluation. The misconceptions from the staffs indicated that they perceived that evaluations beyond Level 1 are difficult, will require more expertise to conduct more comprehensive research projects when using control groups. If new models were used, they may need to pursue further education and training or at least review and update their education and training. These types of activities would require time, effort, and financial resources that they seemed unwilling to commit. These misconceptions were also found in Phillips' study (2000). This exposed the staff's lack of knowledge and expertise for conducting evaluations. It also reflects why many organizations remain satisfied with how they evaluate their training programs as long as they receive positive reactions to the training programs and they remain financially stable.

Studies Utilizing All Four Levels of Kirkpatrick's Model

A few limited studies evaluated organizations that utilized all four levels of Kirkpatrick's model. The organizations ranged from government agencies, businesses, and, healthcare, academic, and religious organizations. The findings were diverse.

Lockwood's (2001) study was an action research project to diagnose, design, implement, and evaluate an orientation program for the 103 San Diego District Attorney's Office (DA) employees. Two instruments, the *DA Reaction Survey* and the *DA Orientation Knowledge Test* were created to assess Kirkpatrick's (1998) Levels 1 (reaction) and 2 (learning). Both were given to the employees immediately upon completion of a live orientation presentation. The presentation received the highest rating, and approximately 70% of the trainees scored ninety percent or higher on the 19-question *Orientation Knowledge Test*. A follow-up survey was developed and administered to the managers and supervisors of the new employees four weeks after the employees attended the orientation program. In general, the results from this survey indicated that the managers and supervisors strongly agreed that the trainees needed less attention, were more focused on satisfying both internal and external customers, and had increased communication with peers. To evaluate Kirkpatrick's Level 4 (results), the orientation training was linked to the DA's operating budget. According to Lockwood's (2001) forecast, there was a cost saving for the managers and supervisors, co-workers, and the new employees, resulting in a potential benefit to the organization. Lockwood estimated the gross benefits from the orientation training were projected to total from approximately $60,000 to $100,000 per year, depending on the number of new hires. These figures are based on both increased individual performance and reduced reliance

on co-workers, and managers and supervisors. It was projected that in the first year of the orientation program the organization could potentially realize a savings of approximately $45,000 to $77,000 dollars, an almost 200% ROI, i.e., the ROI was almost double, and would increase further with each year of implementation. Lockwood (2001) stressed that the savings was based on subjective evaluation developed from the three criteria areas – savings in time for managers/supervisors, co-workers, and trainee.

A study by Bledsoe (1999) was designed to evaluate the relationship of the four levels of the Kirkpatrick Model as they related to corporate computer training courses. The subjects for this study were employees of various departments of a medium-sized financial organization (1,200+ employees) in the Midwest. Participants voluntarily registered for a 4-hour Advanced Microsoft Outlook training class. The total number of employees that participated in all four evaluation levels for this study was 69. The objective of Bledsoe's (1999) study was to provide the first fully implemented study to investigate correlations among all four levels of the Kirkpatrick model as they related to corporate computer training courses. Six relationships were examined. Only one of those (Reaction and Behavior) was found to be significant and at the moderate level. This study also concluded that evaluations conducted at Level 1 does not provide evidence of the overall success of a training program.

Attia (1998) studied a total of 101 sales supervisors of one large multinational company operating in Egypt. The study was designed to test a model for evaluating sales training programs' effectiveness. Due to management's role in deciding who would attend the training during the study period, the assignment of sale supervisors to experimental (59 trainees) and control (42 non-trainees) groups was non-random. While

all four levels were utilized, Attia's study primarily focused on the Levels 3 and 4, which utilized an experimental-control group design with pre-and-post measurements. The findings indicated that no differences were found between anonymous and non-anonymous responses for Level 1. There was a significant improvement in behavior (Level 3) from pretest to posttest, but the behavior improvement was insignificant between the trainees and non-trainees for both the self-evaluation and supervisory-evaluation. The trainer's evaluation of trainees and the utility analysis were two complementary techniques found to be useful when conducted in conjunction with the Kirkpatrick model. The utility analysis suggested a 17:1 ROI, i.e., that each dollar invested in conducting sales training generated $17 in revenue over a five-year period. According to Attia (1998), this ROI justified the large amount of money invested in sales training programs.

Larsen (1985) conducted a study to assess how useful and practical the success case method was for evaluating an administrator training program (ATP) in a public sector business, and a training program of a Fortune 500 company that had internally developed and implemented a new training program for health care administrators. The success case method focuses on assessing the performance changes and results of training, and helps explain how successful trainees make use of the training by collecting descriptive data about the uses and benefits of the training. As Brinkerhoff (1983) stressed, observations, work samples, and sales records, the typical methods for gathering data about results and impact of training, are expensive and time consuming. The success case method can gather significant formative data at little cost.

Larsen designed a success case method interview instrument based on the overall company concerns. Those concerns addressed by in the instrument were determined by an evaluation consultant and the training director. Training managers were asked to select the success case trainees who: "(a) learned the content of the training better than most, (b) were more positive and contributed more than others during discussion periods, (c) were more likely to apply the skills and knowledge taught in training, and (d) believed in the utility and worth of their training experiences" (Larsen, 1985, p. 44). Of the 37 administrators completing the training, nine were selected for the success case telephone interviews conducted by the training managers two to three weeks after training. Larsen (1985) emphasized the difficulty to adapt training costs and benefits into tangible numbers, but acknowledged there must be an attempt to quantify some of these costs. As Larsen estimated, the total costs for carrying out the success case method was approximately $1,400. This represents about 2% of the $70,000 budget to develop, implement, and evaluate the administrator training program.

Tidler (1999) used Kirkpatrick's model in assessing the effectiveness of training in the treatment of pediatric asthma. Her study was a historical research since the five training sessions she examined had already taken place, and all the data were stored in three databases for three years. The sample was 84 healthcare professionals (21 were physicians) at a healthcare institution in the Southwest.

The training was an instructor-led classroom-based format. Due to the nature of the participants' self-response to the questionnaires of Levels 1 and 2, there was a high correlation between Levels 1 and 2 indicating the participants' satisfaction, and, their willingness to learn, what Kirkpatrick claims. However, the study did not support any

behavioral (performance) changes (Level 3), nor increased revenue charges (Level 4). Tidler claimed there was a lack of evidence demonstrating whether continuing medical education activities can be evaluated using all four levels of Kirkpatrick's model. Levels 3 and 4 assessments in the healthcare industry are rare.

Wertz (2005) studied the effectiveness of the CLAD (cross-cultural, language, and academic development) training for a group of K-12 teachers currently teaching in Shasta, Trinity, and Tehama Counties in northern California. The data were collected from three areas: pre-and post-surveys from 17 teachers who were currently taking a CLAD class, interviews with 25 teachers who had taken the CLAD training in the last year, and interviews with 12 administrators. The pre- and post-evaluations were reported by the 17 teachers by completing a Likert-scale survey, with the findings showing significantly positive responses and changes at Level 1 (reaction), Level 2 (learning), and Level 3 (behavior). A few of the teachers reported success by their students that were considered Level 4 achievements. One mixed finding was that on the surveys, 53% of the participants stated that either their administrators were not aware they were taking the training, or they were not aware of changes in the classroom. This differed from the perception of administrators, who saw themselves as being supportive.

Kim (2006) used the Kirkpatrick model in the evaluation of a community church's instructor training program entitled CAL (Called to Awaken the Laity) in Seoul, Korea. The sample was 405 CAL program participants, with 383 of the participants completing an initial survey, and 269 completing a follow-up survey. Like most self-reported survey studies, most participants indicated they were satisfied with the program and perceived that they had learned, and improved their knowledge and skills due to the

training received. One interesting finding from the study was that the knowledge and skill acquisition was inversely related to a plan for implementation of the training at the churches the participants attended, i.e., those who reported a higher degree of knowledge acquisition were less likely to have a plan. This finding was inconsistent with other studies, where a greater knowledge gain was directly related to a higher rate of behavior or performance.

Summary of the Empirical Literature

Research over the past five decades still leaves many unanswered questions about the effectiveness of training interventions in general, and evaluation models in particular (Bomberger, 2003). Although numerous studies have focused on evaluation of training programs, there is no universally accepted model for evaluating training. With great praise and acceptance of Kirkpatrick's model, research findings support the critical need for further studies of Kirkpatrick's model in its entirety (Alliger & Janak, 1989; Bassi, Benson & Cheney, 1996; Bomberger, 2003; Bromley & Kitson, 1994; Plant & Ryan, 1992).

In a 1989 study, 30 years after Kirkpatrick introduced the four-level evaluation model, Alligar and Janak (1989) examined how many articles on training and evaluation used the Kirkpatrick model. They reported that, in practice, most training was evaluated at the reaction level only. Even though organizations have increasingly recognized the importance of evaluating their HR programs at all four levels, there are still large gaps between acknowledging the importance to conduct Kirkpatrick's evaluation in all four levels and putting them into the practices (Phillips, 2003).

For some organizations that have diligently collected all sorts of performance data, the data were not used or examined for a variety of reasons. Most organizations and HRD professionals evaluate programs at Levels 1 and 2, mirroring Kirkpatrick's contention that everyone seems to talk about the importance of evaluating of training programs, but few do anything about it. Most evaluate reactions (Level 1) but seldom go any farther (Kirkpatrick & Kirkpatrick, 2006). Kirkpatrick suggests that evaluations should begin with Level 1 and proceed up the levels as time and resources permit. Bomberger (2003) recommended that additional research is needed to investigate why measurement of job performance (Level 3) and training results (Level 4) are not evaluated more frequently. Although some argue that it is difficult to establish a direct link between training and the resulting behavioral change and organizational impact, Lockwood (2001) claims that the attempts to demonstrating a relationship are often sufficient. More empirical studies of Kirkpatrick's model in all four levels to determine if the four levels are intercorrelated, as some authors claim, are needed (Attia, 1998; Phillips, 2000).

In addition, based on the conclusions drawn from the literature review, few studies of Kirkpatrick's model have been conducted that examine the evaluation of customer service and sales training programs, the focus of this study (Attia, 1998). To date, it has been very difficult to evaluate sales training effectiveness without sound comprehensive research that incorporates all the four levels. Since Kirkpatrick claimed that his concepts, principles, and techniques can be applied to technical, sales, safety, and even academic courses (Kirkpatrick & Kirkpatrick, 2006), this study was designed to demonstrate

whether it is feasible for HRD professionals to use evaluation tools based on Kirkpatrick's model within the context of sales training.

Summary of Chapter 2 and Research Question

This literature review examined the theoretical frameworks and empirical studies upon which this study was based. The significance of Kirkpatrick's four-level evaluation model, its advantages and weaknesses, and the similarities and differences with other evaluation models were presented. The empirical literature presented 14 specific studies of various surveys and training programs. The conclusion from these studies supports the premise that, while not significant in all studies, the utilization of Kirkpatrick's evaluation model should be further studied to explore how to more effectively connect training, learning, performance, and organizational impact. It is within this premise that this study was conducted.

Based on Kirkpatrick's theory (Kirkpatrick & Kirkpatrick, 2006) and the literature review, this study was guided by the following research question and four research hypotheses.

Research Question

Do the data from a training program implemented at an organization in the hospitality industry support the theories of Kirkpatrick's evaluation model (Kirkpatrick & Kirkpatrick, 2006)?

Research Hypotheses

To answer this basic research question, four research hypotheses served as the guides for the data to be collected and analyzed.

Hypothesis one (H_1). Employees who completed the training will improve their knowledge of the content and required skills (Level 2).

Hypothesis two (H_2). Employees who completed the training will improve their job performance (Level 3).

Hypothesis three (H_3). Employees who completed the training will contribute to increased organizational impact (Level 4).

Hypothesis four (H_4). Employee learning (Level 2) and job performance (Level 3) will predict organizational impact (Level 4).

The general context of this study is the hospitality industry. Specifically, the data for this study were the evaluations of training provided to reservations sales agents of a large international hotel chain.

The next chapter discusses the methodological rationale and the review of methodological literature. In addition, the population and sample, the training program and intervention, the data collected, and analysis of the data are introduced. The schedule of the tasks and activities to complete this study are then presented, after which the chapter is summarized.

CHAPTER 3

METHOD

This study examined Kirkpatrick's training evaluation model (Kirkpatrick & Kirkpatrick, 2006) by assessing a sales training program conducted at an organization in the hospitality industry. The research question and four hypotheses, as stated in the chapters 1 and 2, served as the foundation and purpose of this study and are further addressed in this section. This chapter also discusses the methodological rationale and review of methodological literature of the study, the population and sample, the training program/intervention, the data collected, and the analysis of the data.

Research Question and Hypotheses

This study was guided by the following research question and four research hypotheses.

Research Question

Do the data from a training program implemented at an organization in the hospitality industry support the theories of Kirkpatrick's evaluation model (Kirkpatrick & Kirkpatrick, 2006)?

Research Hypotheses

To answer this basic research question, four research hypotheses served as the guides for the data to be collected and analyzed.

Hypothesis one (H_1). Employees who completed the training will improve their knowledge of the content and required skills (Level 2).

Hypothesis two (H_2). Employees who completed the training will improve their job performance (Level 3).

Hypothesis three (H_3). Employees who completed the training will contribute to increased organizational impact (Level 4).

Hypothesis four (H_4). Employee learning (Level 2) and job performance (Level 3) will predict organizational impact (Level 4).

Methodological Rationale and Review of Methodological Literature

Research Design

To examine and confirm Kirkpatrick's theories and demonstrate the assumptions from the literature review, the primary methodology being employed in conducting this research is a one group ex post facto analysis (Gay & Airasian, 2002; Newman, Newman, Brown & McNeely, 2006). In this study, the attempt is to examine a training intervention based on comparative pre- and post-intervention performance outcome data. This is not an experimental study, as the training intervention, described later in this section, cannot be manipulated. In addition, there is no randomization, manipulation of the intervention, or the use of a control group that characterizes experimental research. These factors are a consistent situation when evaluating workgroups in organizations and are considered weaknesses of an ex post facto design (Gay & Airasian, 2002; Merriam & Simpson, 1995; Newman & Newman, 1994). There are, however, extensive and multiple data to assess the various variables; that is, knowledge and skills, job performance, and organizational impact. The relationships between variables may be demonstrated. However, cause and effect relationships cannot be inferred.

Because of the inability to appropriately identify causal relationships, many researchers "tend to regard ex post facto as inferior research that should not be conducted" (Newman & Newman, 1994, p. 115). However, Newman and Newman

(1994) indicated that this is not necessarily true if the research question deals with relationships. By utilizing ex post facto design with tests of hypotheses, it is appropriate to increase the internal validity and explore relationships between variables. According to Newman and Newman (1994), ex post facto design also has the potential for the most amount of external validity when compared to other designs, such as experimental, quasi-experimental, and true-experimental research.

Furthermore, to determine the relationships between the variables and to use relationships in making predictions, there is the need to examine the relationships of Kirkpatrick's four levels of evaluation, and to examine the degree to which the variables are related (Gay & Airasian, 2002). According to Gay (1996), "if two variables are highly related, scores on one variable can be used to predict scores on the other variable" (p. 305). Therefore, "the variable upon which the prediction is made is referred to as the predictor, and the variable predicted is referred to as the criterion" (Gay, 1996, p. 305). Although a relationship study examines how each predictor variable is correlated with the criterion variable, a combination of variables usually results in a more accurate prediction than any one variable (Creswell, 2005; Gay 1996). A prediction study is appropriate for this research because it tests theoretical hypotheses concerning variables believed to be predictors of a criterion. In other words, by employing a prediction design, the study sought to anticipate performance outcomes by using specific variables as predictors. The variables and predictors in this study are the first three levels of performance outcomes, reaction, learning, and job performance, as identified by Kirkpatrick. The organizational impact of Level 4 assessment is, therefore, the criterion. The test is to confirm the theoretical relationships predicted in Kirkpatrick's model.

As concluded from the literature reviews of Attia's (1998) and Phillips' (2000) studies, Flynn's recommendation (1998), and Kirkpatrick's principles (Kirkpatrick & Kirkpatrick, 2006), discussed in chapter 2, Kirkpatrick's evaluation model was used to assess a sales training program. There are extensive data available to examine the multiple levels of dependent variables. In addition, because the general and ultimate goal of most sales training programs is to increase the sales for all organizations, it was critical to follow Kirkpatrick's model by assessing the employees' knowledge and skill gains and job performance improvement, and how they relate to the sales and the organizational impact. Finally, Kirkpatrick and Kirkpatrick (Kirkpatrick & Kirkpatrick, 2005, 2006) recommend that HRD professionals and researchers utilize control groups whenever possible. However, it should not prohibit the attempts of evaluation if control groups cannot be used. In addition, based on Tidler's (1999) study, a one-group ex post facto analysis is appropriate to examine the existing data in this study.

Effect Size

In a recent review of the published training and development literature from 1960 to 2000 (Arthur, Bennett, Edens, & Bell, 2003), the researchers conducted a meta-analysis of 162 training evaluation studies to examine the relationship between specified training design and evaluation features and the effectiveness of training in organizations. By utilizing Kirkpatrick's model as the framework and his four levels as the evaluation criteria, the researchers found that the average or mean effect sizes (*d*s) for training interventions were fairly large, regardless of the topics and methods used. The results are 0.60 for reaction criteria, 0.63 for learning criteria, 0.62 for behavior criteria, and 0.62 for results criteria. These results indicated a medium to large effect size for organizational

training (Arthur et al., 2003), which was considered the guideline for this study as addressed in the following sample size section.

Population and Sample

The population for this study was reservations sales agents from a leading luxury hotel (referred hereafter as hotel) chain's reservations center. During the study period, there were 335 reservations sales agents employed in this central reservations center. The sample of this study was the reservations agents who completed a sales training program/intervention and for whom complete pre-and-post training data were available. The number of available agents was 69. The agents included in the study were compared to those not included on the available variables of: type of agent, length of employment and length of time before training, to ascertain whether or not differences existed.

Sample Size

According to Gay and Airasian (2002), the number of subjects significantly affects the power of a study. The power means the statistical ability to reject a false null hypothesis. In addition, "if the sample is too small, the results of the study may not be generalizable to the population" (Gay & Airasian, 2002, p. 111). However, in many situations, researchers have difficulties accessing large numbers of potential research participants. Because there were 69 participants in this study, the statistical power requirements have been met. This sample size will yield 99% power in paired-sample *t* tests for a medium-large effect size, $d = .6$ (Arthur et al., 2003).

Depending on the type of research involved, some experts consider the minimum sample size of 30 as a guideline for correlational, causal-comparative, and true experimental research (Gay, 1996; Gay & Airasian, 2002). For regression types of

analytic work, a good rule of thumb is 15 participants per variable (Tabachnick & Fidell, 2001). However, because of the factors discussed previously, the appropriate method for this study is ex post facto analysis.

The Training Program/Intervention

The training intervention is a two and one-half day classroom-based comprehensive course for reservation sales agents. As one of the few hotel chains within the luxury hotel segment, providing high standard customer service has long been the main objective for the hotel's reservations center. However, due to the changes of business climate and the continuous increasing competition, the hotel recognizes the significance of the reservations sales agents who deliver the first impression to their customers and have direct impact on their bottom lines. Consequently, a new training program/intervention has been delivered to the reservations sales agents since 2005 to provide the skills and ultimate performance to meet their new business objectives. The training program (see Appendix A) is comprised of three major components: (1) recognizing and possessing the right attitudes for succeeding as a reservations sales agent, (2) knowledge and skills needed for completing the sales, and (3) practicing skills through rehearsals and role plays.

Considering the nature of hotel operation and the demands for providing services 24 hours a day, there were only about 10 agents being scheduled for the each training session. The training was conducted by the same learning coach (facilitator) despite the time the sessions were scheduled. The evaluation of this program has primarily been the participant's end-of-training evaluation and tests of the content learned, Levels 1 and 2 of Kirkpatrick's model (Kirkpatrick & Kirkpatrick, 2006). However, the hotel has collected

comprehensive performance data for both individual and organizational levels, but these data have not been analyzed to evaluate job performance (Level 3), and organizational impact (Level 4) prior to this study.

Data Collection

The data were retrieved from the Hotel's human resource department and central reservations center's database. The specific time period to be studied was the two-and-a-half-year period of January 2005 to May 2007. As stated previously, the emphasis on this particular period is due to the hotel's shift of business practices in 2005 from focusing primarily on customer service to a focus on promoting sales while still maintaining a high standard of customer service. A new training program was launched to implement the new standards and practices to meet their new objectives.

For all the variables examined in this study, the pre training data consisted of the data two months before training, and the post training data were two months after training. Two months was chosen because much variability was observed in a single month's data. The data for three consecutive months pre and post training were unavailable for some agents, which would exclude more reservations sales agents being the study sample. In addition, this also minimized seasonality variability.

The specific data used in this study were data analyzed to test the four hypotheses guiding this study. For the first hypothesis, the examination of knowledge and skills, the data analyzed were the reservations sales agents' call quality assessment scores per month. Immediate two months before and post training call quality assessment scores were used for this hypothesis. This assessment measured the agents' knowledge and

skills in handling calls. Utilizing the hotel's scoring criteria (see Appendix B), the call center supervisors randomly reviewed and scored a selection of each reservations sales agent's recorded calls and conversations each month. The score was calculated on a 100-point scale.

To analyze Hypothesis 2, job performance, the data were the call conversion ratio and time usage as the measurements of productivity immediate two months before and post training (Attia, 1998; Cascio, 2000; Kirkpatrick & Kirkpatrick, 2006; Lockwood, 2001). The call conversion ratio is the total number of reservations booked divided by the total number of received calls. Because all incoming calls were routed randomly to the agents, every agent had an equal opportunity to convert each inquiry call into a confirmed reservation. Call conversion is the industry wide measure for reservations sales agents (HSA International, 2007; Ismail, 2002). In recent years, when competition between hotel chains has increased, reservations sales agents' roles have changed from order takers to order makers (Farrell, 2005; HSA International, 2007). Hotels now expect the reservations agents to convert inquiry calls into confirmed reservations. As a result, the call conversion ratio is not only a job performance measurement, but also a business survival indicator. To successfully convert an incoming call into a confirmed reservation, reservations agents have to apply their knowledge about hotel properties, the services, the destinations, etc., and also their listening, interpersonal, and relationship skills. This affirms how vital the call conversion is being considered as a key job performance measurement for reservations sales agents, and a key indicator for the call center's success. The higher the conversion ratio means the more confirmed reservations and a more productive reservations center.

In addition, time usage, the second measure of Hypothesis 2, is also a key job performance measurement (Attia, 1998; Cascio, 2000; Kirkpatrick & Kirkpatrick, 2006; Lockwood, 2001). Cascio (2000) and Lockwood (2001) indicated that the length of time employees spend doing specific tasks should be measured to identify the results/outputs. Data for this study was collected on the time each reservations agent spent on each telephone conversation, and also the time to process the information. From these data two measurements were examined. The first was the average talk time per call, a key measurement of how each agent handles his/her calls. The second was average processing time per call, the time needed to enter information received from a call, whether a reservation was made or not made. The time was recorded and reported in seconds.

To assess the organizational impact, Hypothesis 3, five measurements were conducted. First, the total time saved per call for each agent was calculated as the difference between the average talk and processing time per call before the training and the average talk and processing per call after the training. In Lockwood's study (2001), time saving was considered one of the important measurements of organizational impact. Therefore, it is acceptable to utilize time savings as one of the measurements for organizational impact of this research.

Second, the total employee wages saved per 1,000 calls were calculated by multiplying the total time saved per call times the agents' average hourly wage. The time savings, according to Cascio (2000) and Lockwood (2001), translate to monetary savings. The average hourly wage was calculated from data provided by the hotel's human resource department.

Third, the measurement of total sales was calculated by multiplying the number of room nights by the average daily room rate (ADR). Fourth, the measurement of sales/call ratio was calculated by dividing the total sales by the number of calls received. Because the ADRs change on a daily basis and vary for different properties and regions, the average annual figures used are reports provided by Smith Travel Research (Bowers, 2007; Freitag, 2006a, 2006b, 2006c; Lomanno, 2005), an international research company that collects and reports comprehensive performance data for the hospitality industry, and is considered the industry standard and index.

The fifth measurement of Hypothesis 3 was a ratio calculated by dividing the costs of the training by the sales calculated in the second measurement (the number of room nights times the ADR). The training costs were calculated by the sum of all the costs related to the training intervention. According to the Director of the Hotel's human resource department, the costs include training materials for each agent, the agents' wages, and the learning coach's (facilitator's) fee. Because the training sessions were conducted at the call center, it was agreed by the Director that the costs of utilizing the facility were minimal. Therefore, these costs were excluded for calculating the total costs of training. According to Attia (1998), Kirkpatrick, and Kirkpatrick (Kirkpatrick & Kirkpatrick, 2006), total sales volume, sales/calls ratio expense to sales ratio, etc., are crucial criteria for measuring sales training program results.

For Hypothesis 4, the examination of the inter-level relationships between learning (Level 2), job performance (Level 3), and organizational impact (Level 4) were examined. The data from Level 2 were call quality assessment scores; from Level 3 were call conversion ratios and time usage (total talk time, average talk time per call, total

processing time, average processing time per call); from Level 4 were total time saved, total wages saved, and projected sales.

Analysis of the Data

All data were computed by the Statistical Package for the Social Sciences (SPSS) program and examined for statistically significance. Table 2 presents the data collected and how they were analyzed for each hypothesis. The analysis of the data involved selected descriptive and inferential statistics. The descriptive statistics introduced the mean, the average performance of a group on a variable, and the standard deviation. For inferential statistics, the paired-samples *t* tests were utilized to determine the difference between the means of two sets of data (pre and post training). An F test was used to determine if the R^2 was significantly different than 0 at an alpha of .05 for correlations and multiple regressions.

This rigorous and systematic approach uses a statistical power analysis by identifying appropriate sample size, the level of statistical significance (alpha), the amount of power desired in a study, and the effect size involved in statistical inference (Cohen, 1992; Creswell, 2005). A significance (or alpha) level is a probability level that reflects the maximum risk to take that any observed differences are due to chance (Creswell, 2005). It is usually set at .05 (Cohen, 1992; Creswell, 2005; Newman & Newman, 1994). One-tailed test of significance was utilized, as the research indicates a probable direction. According to Creswell (2005), a one-tailed test has more power, which means more likely the hypothesis will be rejected.

Table 2

Summary of the Variables Needed and Statistical Tests Used to Analyze Each of the Four Hypotheses

Data Treatment	H_1	H_2	H_3	H_4
Data Collected	Call quality assessment scores	Call conversion ratio; total talk time; average talk time per call; total processing time; average processing time per call.	Total time saved; total wages saved; sales generated; sales/call ratio; costs of training/sales ratio.	Call quality assessment scores, call conversion ratios, time usage (total talk time, average talk time per call, total processing time, average processing time per call), total time saved, total wages saved, and total sales
Analysis of the Data	Call quality assessment score (paired-samples *t* test).	Total # of reservations booked/ total # of received calls; sum of the talk time; talk time/calls; sum of processing time; processing time/calls (paired-samples *t* test).	Sum of time saved; sum of time saved x the average hourly wage; # of room nights x estimated ADR; sales/calls (paired-samples *t* test); costs of training/sales.	F test

To analyze the first hypothesis, the examination of the knowledge and skills, pre and post training was measured by an objective call quality assessment of the employees completing the training. Paired-samples *t* tests were used to analyze the data for the call quality assessment scores.

To analyze Hypothesis 2, job performance, the call conversion ratios from pre and post training were measured. Paired-samples *t* tests were used to analyze the data for the call conversion ratio values. In addition, the paired-samples *t* tests were performed on the talk time and processing time.

To analyze Hypothesis 3, organizational impact, five calculations were conducted for analyzing this hypothesis. First, the total time saved per call was calculated. Second, the employee wages saved per 1,000 calls were calculated by utilizing the total time saved per call times the average employee hourly wage provided by the hotel. Third, the total sales were calculated by utilizing the number of room nights reserved times the average daily room rate (ADR). Fourth, the sales/call ratio was calculated by dividing the sales by the calls received. The total sales and the sales/call ratio were compared from the pre and post training by paired-samples *t* tests. Fifth, the training cost/sales ratio was calculated by dividing the costs of the training by the total sales previously calculated (the number of room nights reserved times the ADR). The training costs were calculated by the sum of all the costs related to the training intervention, as previously described.

To analyze Hypothesis 4, the inter-level relationships among learning as measured by change in call quality assessment scores from pre to post training (Level 2), job performance as measured by change in call conversion ratio, change in average talk time per call, change in average processing time per call from pre to post (Level 3), and organizational impact as measured by the increase in sales per call (Level 4) were examined. The training/intervention data were utilized for a hierarchical regression test to see if the gains in Levels 2 and 3 can be used to predict gains in Level 4. Multiple regression, as defined by Creswell (2005), is a statistical procedure for examining the combined relationship of multiple independent variables (the Levels 2 and 3 outcomes) with a single dependent variable (Level 4 outcome). “In regression, the variation in the dependent variable is explained by the variance of each independent variable (the relative importance of each predictor), as well as the combined effect of all independent variables

(the proportion of criterion variance explained by all predictors), designed by R^2" (Creswell, 2005, p. 336). An F test was used to determine if the R^2 was significantly different than 0 at an alpha of .05. The F test is chosen as it is very robust and is the most frequently used test of significance (Creswell, 2005; McNeil, Newman & Kelly, 1996). Results were considered significant is $p < .05$.

Limitation of This Study

Due to the nature under which this study was conducted, there are two limitations that have been mentioned previously in this chapter. First, the study was an ex post facto study, with the data collected over a two-and-a-half year period, from January 2005 to May 2007. There is an inability to randomly assign and manipulate the independent variables since they had already occurred and were not under the control of the researcher. Also, a control group of non-trainees could not be formed. Second, the data collected, the collection process, and the measurements utilizing the data were already established.

Summary

This methodology chapter discussed the methodological rationale and review of methodological literature of the study, the population and sample, the training program/intervention, the data collected, and analysis of the data. Next, the detailed results of data analysis are presented in chapter 4.

CHAPTER 4

ANALYSIS OF DATA

This study examined Kirkpatrick's training evaluation model (Kirkpatrick & Kirkpatrick, 2006) by assessing a sales training program conducted at an organization in the hospitality industry. The research question and four hypotheses, as stated in the previous three chapters, served as the foundation and purpose of this study. They also served as the guides for the findings addressed in this chapter.

Research Question and Hypotheses

This study was guided by the following research question and four research hypotheses.

Research Question

Do the data from a training program implemented at an organization in the hospitality industry support the theories of Kirkpatrick's evaluation model (Kirkpatrick & Kirkpatrick, 2006)?

Research Hypotheses

To answer this basic research question, four research hypotheses served as the guides for the data to be collected and analyzed.

Hypothesis one (H_1). Employees who completed the training will improve their knowledge of the content and required skills (Level 2).

Hypothesis two (H_2). Employees who completed the training will improve their job performance (Level 3).

Hypothesis three (H_3). Employees who completed the training will contribute to increased organizational impact (Level 4).

Hypothesis four (H_4). Employee learning (Level 2) and job performance (Level 3) will predict organizational impact (Level 4).

Population and Sample

The population for this study was a group of reservations sales agents from a leading luxury hotel chain's reservations center. During the study period from January 2005 to May 2007, there were 335 reservations sales agents employed in this global reservations center (GRC). The number of reservations sales agents who had completed a sales training program/intervention during this period was 270. There were 65 newly hired reservations agents who had not completed the training and, therefore, were not considered for the study. Of the 270 agents who completed the training, 69 of them had data available for at least two months before and after the training program, so these reservations sales agents composed the sample for this study (Table 3).

Table 3

Summary of the Population and Sample Sizes

Criterion	Number of Reservations Sales Agents
Total number of reservations sales agents	335
Number of agents who completed the training	270
Number of agents who completed the training and had two months of pre and post training data available	69

Table 4 outlines the dates of the sales training sessions during the study period and the number of reservations sales agents that attended each of the sessions.

Table 4

Training Dates and the Number of Participants

Sales Training Dates	Number of Reservations Sales Agents
May 26, 2005	8
September 9, 2005	6
October 13, 2005	9
November 5, 2005	6
May 3, 2006	9
September 7, 2006	8
October 5, 2006	8
December 7, 2006	7
February 14, 2007	8
TOTAL	69

Job Titles

Among the 69 agents, 40 of their job titles are Senior Reservations Sales Agents and 23 of them were GRC (Global Reservations Center) Reservations Sales Agents. The remaining 6 included two Customer Service Leaders, one Concierge, one Global Sales Coordinator, one Tour Coordinator, and one Tour Distribution Specialist. There was a significant difference (p = .04) of agent types between the study and the remaining groups. The study group contained more GRC Reservations Sales Agents (33.3%) instead of senior agents (58%) while the remaining group contains 21% GRC Reservations Sales Agents and 56.9% of Senior Reservations Sales Agents.

Length of Employment

The length of employment for the 69 agents ranged from 9 to 123 months (M = 31.4, SD = 24.2). The mean length of employment before receiving training ranged from 4 to 104 (M = 18.1, SD = 21.2). The median length of employment before receiving training was 13 months with 68.1% of the agents receiving their training within the first thirteen months of employment. Of the remaining agents who completed the training and

have employment records available, the length of employment for those 198 agents ranged from 10 to145 months (M = 74.3, SD = 30.5). Length of employment was significantly shorter for the 69 agents in the study compared to the remaining 198 agents, t (265) = 10.58, $p < .001$.

In addition, it should be noted that the length of employment was not significantly correlated with any of the study variables. This indicates that the length of employment was not associated with job performance.

Findings Pertaining to Hypothesis One

To answer research hypothesis one, the reservations sales agents who completed the training improved their knowledge of the content and required skills (Level 2), the study examined the average call quality assessment scores two months before and the average scores two months after the training intervention. This assessment measured the agents' knowledge and skills in handling calls. Utilizing the Hotel's scoring criteria (see Appendix B), the call center supervisors randomly reviewed a selection of each reservations sales agent's recorded calls and conversations each month. The score is calculated on a 100-point scale. For this particular variable, eight (8) out of the 69 agents' call scores were unavailable. Therefore, the n for this variable was 61 instead of 69 for all other variables.

The call scores before training ranged from 53 to 97 (M = 84.2), while after training they ranged from 66 to 98 (M = 87.7). As shown in Table 5, the call score mean increased significantly by 3.52 points from pre to post training , p = .001, with a medium effect size of .46. The significant improvement in the call score supports hypothesis one that the reservations sales agents who completed the training improved their knowledge

of content and required skills in handling calls (Level 2). Therefore, hypothesis one is accepted.

Table 5

Knowledge and Skills Variable and Statistical Results for Hypothesis One

Call Score	Mean	Standard Deviation	*t*	*p*-value	Effect Size (*d)*
Pre	84.18	9.41	3.60	.001**	0.46
Post	87.68	7.21			

** $p < .01$. n = 61.

Findings Pertaining to Hypothesis Two

To answer research hypothesis two, the reservations sales agents who completed the training improved their job performance (Level 3), the variables to be examined are the call conversion ratio, the average talk time per call, and the average processing time per call as the measurements of productivity.

The call conversion ratio is the ratio of the total number of reservations booked divided by the total number of received calls. Call conversion is the industry wide measure for reservations sales agents (HSA International, 2007; Ismail, 2002). Because all incoming calls are routed randomly to the agents, every agent has an equal opportunity to convert each inquiry call into a confirmed reservation.

The ratios ranged from .148 to .577 (M = .319) before training and from .238 to .445 (M = .340) after training. As shown in Table 6, the mean increase in the call conversion ratio of .021 was significant, p = .001, d = .41.

Table 6

Job Performance Variables and Statistical Results for Hypothesis Two

Variables	Mean	Standard Deviation	*t*	*p*-value	Effect Size (*d*)
Conversion (%)					
Pre	.319	.055			
			0.034	.001**	0.41
Post	.340	.043			
Average Talk Time (sec)					
Pre	279.25	57.12			
			1.56	.124	0.19
Post	284.29	59.37			
Average Processing Time (sec)					
Pre	31.47	13.68			
			.37	.710	0.04
Post	31.06	12.46			
Number of Calls (#)					
Pre	1097.65	282.56			
Post	1021.77	290.47			
Number of Reservations (#)					
Pre	346.52	94.53			
Post	345.65	102.34			
Average (Talk + Processing) Time /Per Call (sec)					
Pre	310.33	58.03			
Post	312.85	59.81			
Total Talk Time (sec)					
Pre	298674.36	69195.92			
Post	280451.14	74702.33			
Total Processing Time (sec)					
Pre	33874.14	16306.20			
Post	31449.97	15095.56			
Total Talk + Processing Time (sec)					
Pre	332548.50	74232.49			
Post	311901.12	81100.41			

** $p < .01$. n = 69.

The call conversion ratio is important in Level 3, job performance, because the call conversion ratio is not only a job performance measurement, but also a business survival indicator. To successfully convert an incoming call into a confirmed reservation, reservations sales agents have to apply their knowledge about hotel properties, the services, the destinations, etc., and also their listening, interpersonal, and relationship skills. This affirms how vital the call conversion is as a key job performance measurement for reservations sales agents, and a key indicator for the call center's success. The higher the conversion ratio means more confirmed reservations, and a more productive reservations center. The significant improvement in the call conversion ratio supports hypothesis two that the reservations agents who completed the training improved their job performance (Level 3), i.e., made significantly more confirmed reservations.

Time usage is also a key job performance measurement of productivity (Attia, 1998; Cascio, 2000; Kirkpatrick & Kirkpatrick, 2006; Lockwood, 2001). Cascio (2000) and Lockwood (2001) indicated that the length of time employees spend doing specific tasks should be measured to identify the results/outputs. Data for this study are collected on the time each reservations agent spends on each telephone conversation (average talk time), and also the time to process the information (average processing time). The time is recorded and reported in seconds.

The average talk time ranged from 176 seconds to 468.5 seconds ($M = 279.25$) before training and from 177.50 seconds to 476.50 seconds ($M = 284.29$) after training. The mean increase in the average talk time of 5.04 seconds was not significant, $p = .124$, $d = .19$.

The average process time ranged from 3.50 seconds to 71.50 seconds (M = 31.47) before training and from 4.00 seconds to 65.00 seconds (M = 31.06) after training. The mean decrease in the average processing time of 0.41 second was not significant, p = .710, d = .04.

The average talk time and average processing time are important in Level 3, job performance, because they measure the job performance on the length of time the reservations sales agents spend on each call. To efficiently use the time, the reservations sales agents should minimize their talk time and processing time so they would be able to handle more incoming calls in any given shift.

In addition to call conversion ratio, average talk time, and average processing time, all important for hypothesis two, other measurements of time usage associated with those variables are also reported here. The number of calls received ranged from 302 to 1761 (M = 1097.65) before training and from 206 to 1676.50 (M = 1021.77) after training. The total talk time ranged from 96,271.00 seconds to 431,742.50 seconds (M = 298,674.36) before training and from 56,247.00 seconds to 465,970.50 seconds (M = 280,451.14) after training. The total processing time ranged from 2,501.50 seconds to 84,640.50 seconds (M = 33,874.14) before training, and from 2,096.00 seconds to 79,370.00 seconds (M = 31,449.97) after training. The decrease in total talk plus processing time after training was marginally significant, p = .051, which is similar to the decrease in total talk time, p = .056. The decrease in processing time after training had p = .108, which is almost marginally significant. The average time to handle a call (talk plus processing time) before the training intervention was 310.72 seconds, and the average time to handle a call after the training intervention was 315.35 seconds, an

additional 4.63 seconds. As a result, hypothesis three was accepted in part with conversion ratio improved significantly, but time usage did not show significant improvement.

Findings Pertaining to Hypothesis Three

To answer research hypothesis three, the reservations sales agents who completed the training contributed to increased organizational impact (Level 4), five measurements were conducted. First, the total time saved per call for each agent was calculated as the difference between the average talk and processing time per call before the training and the average talk and processing per call after the training. As stated in the finings pertaining to hypothesis two, the average time to handle a call (talk plus processing time) before the training intervention was 310.72 seconds, and the average time to handle a call after the training intervention was 315.35 seconds, an additional 4.63 seconds.

Second, the total employee wages saved per 1,000 calls were calculated by multiplying the total time saved per call times the agents' average hourly wage. As a result, the average time to handle a call after the training actually increased 4.63 seconds. It was a total of 4,630 seconds increase for 1,000 calls and a $16.385 cost of wages.

The third measurement was total sales, which was calculated by multiplying the number of room nights by the average daily room rate (ADR). As shown in the following Table 7, the sales ranged from $42,015.50 to $352,200.42 ($M$ = 197667.72) before training and from $42,690.96 to $401,443.85 ($M$ = 201622.70) after training. The number of bookings ranged from 76 to 560 (M = 346.52) before training and from 70 to 653.50 (M = 345.65) after training. The number of room nights ranged from 170 to 1287.00 (M = 751.22) before training and from 156.00 to 1459.00 (M = 760.64) after training.

Fourth, the measure of sales/call ratio, which was the total sales divided by the number of calls received. Before training the sales per call ranged from $47.34 to $275.18 (*M* = $180.54), and from $127.43 to $270.99 (*M* = $197.17) after training. The mean increase in the sales per call was $16.63 (*SE* = $3.89), and was significant, $p < .001$, $d = .51$ as shown in Table 7. The median increase in sales per call was $18.11, with 71% of the agents improving their sales per call.

Table 7

Organizational Impact Variables and Statistical Results for Hypothesis Three

Variables	Mean	Standard Deviation	*t*	*p*-value	Effect Size (*d*)
Total Sales ($)					
Pre	19,7667.72	57,981.58			
			.54	.591	0.06
Post	20,1622.70	65,732.74			
Sales per Call ($)					
Pre	180.54	34.03			
			4.27	<.001**	0.51
Post	197.17	28.38			
Bookings (#)					
Pre	346.52	94.53			
			.08	.936	0.01
Post	345.65	102.34			
Room Nights (#)					
Pre	751.22	218.46			
			.34	.736	0.04
Post	760.64	240.89			

** $p < .01$. n = 69.

Kirkpatrick and Kirkpatrick (2006) indicated sales-per-call is a crucial criterion for measuring sales training programs results. The sales-per-call is determined by the total sales divided by the total number of calls received. In sum, the significant mean

increase and the improvement in the sales per call support hypothesis three that the reservations sales agents who completed the training contributed to increased organizational impact (Level 4).

Regarding the fifth measurement, cost of training/sales ratio, according to the HR Director, the cost of training materials (workbook, handouts, etc.), was $399 dollars per agent. Based on the average $12.74 hourly wage, the two and a half day, 20 hours, of training, the per agent wage cost was $254.80. The fee for the learning coach was $420 dollars per training program. With a maximum of 12 agents per session, the learning coach fee per agent was $35. Thus, the total cost of training per agent was $688.80 dollars as shown in Table 8.

Table 8

Costs of Training Intervention

Items	Total Cost	Cost Per Agent
Training Materials (workbook, handouts, etc.)		$399.00
Employee Wages		$254.80
Learning Coach Fee (maximum 12 agents per session)	$420.00	$35.00
TOTAL		$688.80

Regarding the cost of training/sales ratio, it was calculated as follows. First, the total improvement in sales is determined by the difference between the total sales before ($57,981.58) and after the training (65,732.74), or $7,751.16. This is divided by the number of agents that were trained, 69, for the average gain in sales for each agent, $112.34, a 13.37% increase. Finally, this amount is divided by the total cost of training per agent, $688.80, for a ratio of 1/.163. This means that, for every dollar spent for the

training, the average sales for each agent was $1.163 above the average sales per agent before the training for the first two months after their training. The total amount for the 69 agents is $80.247 for the two months. Projecting the sales per agent for 12 months, assuming the average amount of sales remains the same, the per agent average sales was $6.978, and $481.482 for all 69 agents. This demonstrates a significant organizational impact of the training investment.

In summary, the significant improvement in the sales per call supports hypothesis three, that the reservations agents who completed the training contributed to increased organizational impact (Level 4), i.e., made significantly more sales. Thus, hypothesis three is accepted.

Findings Pertaining to Hypothesis Four

To answer research hypothesis four, employee learning (Level 2) and job performance (Level 3) will predict organizational impact (Level 4), the differences from pre to post training on the learning, performance and impact variables were utilized for correlations and hierarchical regression analyses.

As shown in Table 9, increases in sales per call were significantly associated with conversion ratio increases, $r = .82$, $p < .001$, and with increases in average talk time per call, $r = .34$, $p = .007$. For an additional increase of one minute in talk time, the average increase in sales per call was $36 (SE = $11).

Table 9

Correlations of Organizational Impact Change from Pre to Post with Changes in Employee Learning and Job Performance Variables for Hypothesis Four

Variables (Post – Pre)	Sales/Call Increase (Post – Pre)	
	Pearson Correlation (*r*)	*p*-value
Employee Learning		
Call Score	-.097	.455
Job Performance		
Conversion	.819**	<.001
Average Talk Time	.341**	.007
Average Processing Time	.027	.839

** $p < .01$. n = 60.

Hierarchical regression analyses were performed to test that Level 2 employee learning (call score) and Level 3 job performance (conversion, average talk time, and average processing time) will predict Level 4 organizational impact (sales per call). The call score variable was examined for Level 2. Three additional job performance variables (conversion, average talk time, and average processing time) were also examined for Level 3. The results are shown in Table 10.

In the first block entered into the regression equation, Level 2, call scores ($\beta = .237$, $p < .05$) contributed unique variance to the prediction of sales per call ($R^2 = .027$, $p < .05$) in the regression equation. In the second block entered into the regression equation, after controlling for call score, conversion ($\beta = .903$, $p < .001$) contributed additional variance to the prediction of the increase in sales per call ($R^2 = .809$, $p < .001$) in the regression equation. On the other hand, average talk time and average processing time did not make a statistically significant contribution to the regression equation. Thus, hypotheses four was supported in this model. These findings suggest that sales per call

can be predicted by call score and conversion. Overall, the regression model explained 83.6% of the variability of increase in sales per call.

Table 10

Summary Hierarchical Regression Analysis with Employee Learning and Job Performance, Predicting Sales per Call

	Sales per Call Model			
Variable	β	*SE*	R^2	Sig. *F* Change
Step 1				
Level 2, Employee Learning				
Call Scores	.237*	.016	.027*	.043
Step 2				
Level 3, Job Performance				
Conversion	.903**	.311		
Ave. Talk Time	.126	.094		
Ave. Processing Time	.246	-.056		
Block			.809**	.000
Total adjusted R^2			.836**	

Note. * $p < .05$, ** $p < .001$. *F* value for Block 1 was 5.00 and 65.47 for Block 2

Although two other variables (average talk time and average processing time) were not significant in the model, the hypothesis is accepted.

Summary

The results from the analyses mostly support the hypotheses in this study. The significant improvement in the call score supports hypothesis one that the reservations

sales agents who completed the training improved their knowledge of content and required skills in handling calls (Level 2). Hypothesis two was accepted in part as there was significant improvement in call conversion, but there was no significant improvement of time usage. The significant improvement in the sales per call supports hypothesis three that the reservations agents who completed the training contributed to increased organizational impact (Level 4), i.e., made significantly more sales. Lastly, findings support hypothesis four, that Level 2 and Level 3 variables can be used for predicting Level 4 organizational impact. Chapter 5 discusses the results and implications of these findings. Recommendations are also given for future research and practice.

CHAPTER 5

SUMMARY, CONCLUSIONS, AND IMPLICATIONS

This study examined Kirkpatrick's training evaluation model (Kirkpatrick & Kirkpatrick, 2006) by assessing a sales training program conducted at an organization in the hospitality industry. The research question and four hypotheses, as stated in the previous four chapters, served as the foundation and purpose of this study. They also served as the guides for the summary of the study, discussions, and implications for future studies addressed in this chapter.

Summary of the Study

The study was implemented at a leading luxury hotel and the data were retrieved from its human resource department and central reservations center's database. The specific time period to be studied was the two-and-a-half-year period of January 2005 to May 2007. As stated previously, the emphasis on this particular period was due to the hotel's shift of business practices in 2005 from focusing primarily on customer service to a focus on promoting sales while still maintaining a high standard of customer service. A new training program was launched to implement the new standards and practices to meet their new objectives.

The population for this study was reservations sales agents from the hotel's global reservations center. During the study period, there were 335 reservations sales agents employed in this global reservations center. The reservations sales agents who completed a sales training program or intervention, and for whom complete pre-and-post training data were available, were the sample of this study. The number of available agents was 69.

The training intervention was a two and one-half day classroom-based comprehensive course for reservations sales agents (see Appendix A for the course schedule). This hotel chain within the luxury hotel segment provides high standard customer service, which has long been the main objective for the hotel's reservations center. However, due to the changes in business climate and the continuous increasing competition, the hotel recognized the significance of the reservations sales agents who deliver the first impression to their customers and have direct impact on their bottom lines. Consequently, a new training program/intervention has been delivered to the reservations sales agents since 2005 to provide the skills and ultimate performance to meet their new business objectives. Considering the nature of hotel operations and the demands for providing services 24 hours a day, there were only about 10 agents being scheduled for each training session. The training was conducted by the same learning coach (facilitator) despite the time the sessions were scheduled.

Length of employment was found to be significantly different when comparing the study group and the remaining group of reservations sales agents. This was reflected in the job titles as there were more inexperienced GRC Reservations Sales Agents than Senior Reservations Sales Agents. It implies that the Hotel wanted newly hired agents went through the training first.

Discussion

This study was guided by the research question: "Do the data from a training program implemented at an organization in the hospitality industry support the theories of Kirkpatrick's evaluation model (Kirkpatrick & Kirkpatrick, 2006)?" The data were available for examining the three higher levels of Kirkpatrick's evaluation model. Four

research hypotheses were the guides for the data to be collected and analyzed, and to ultimately answer the basic research question.

Research Question

Do the data from a training program implemented at an organization in the hospitality industry support the theories of Kirkpatrick's evaluation model (Kirkpatrick & Kirkpatrick, 2006)? The results of this study supported the four hypotheses and, therefore, also supported the basic research question. The data from the training program implemented at an organization in the hospitality industry, and described in this study, supported the theories of Kirkpatrick's evaluation model. The detailed results for each hypothesis will now be discussed.

Research Hypotheses

Hypothesis one (H_1). Employees who completed the training improved their knowledge of the content and required skills (Level 2). To answer research hypothesis one, the study examined call quality assessment scores two months before and two months after the training intervention. This assessment measured the agents' knowledge and skills in handling calls. The significant improvement of the call score supported hypothesis one, that the reservations sales agents who completed the training improved their knowledge of the training content and required skills in handling calls (Level 2).

Hypothesis two (H_2). Employees who completed the training improved their job performance (Level 3). To answer research hypothesis two, the variables examined were the call conversion ratio, the average talk time per call, and the average processing time per call. The call conversion is the ratio of the total number of reservations booked divided by the total number of received calls. Call conversion is the industry wide

measure for reservations sales agents (HSA International, 2007; Ismail, 2002). Because all incoming calls are routed randomly to the agents, every agent has an equal opportunity to convert each inquiry call into a confirmed reservation. The call conversion ratio is important in Level 3 job performance because the call conversion ratio is not only a job performance measurement, but also a business survival indicator. To successfully convert an incoming call into a confirmed reservation, reservations agents have to apply their knowledge about hotel properties, the services, the destinations, etc., and also their listening, interpersonal, and relationship skills. The higher the conversion ratio means more confirmed reservations, and a more productive reservations center. The significant improvement in the call conversion ratio partially supported hypothesis two, that the reservations sales agents who completed the training improved their job performance (Level 3), i.e., they made significantly more confirmed reservations. However, while the agents with the lowest conversion increased after the training (1/.148 to 1/.238), the agents with the highest conversion decreased after the training (1/.577 to 1/.445).

In addition to the call conversion ratio, time usage is also a key job performance measurement of productivity. The study examined the time each reservations sales agent spends on each telephone conversation (average talk time), and also the time to process the information (average processing time). The study found that both average talk time and average processing time were not significantly different before and after the training. Thus, hypothesis two was partially accepted for improvement in conversion but not in time usage.

Hypothesis three (H_3). Employees who completed the training contributed to increased organizational impact (Level 4). To answer research hypothesis three, five

measurements were conducted. First, the total time saved per call for each agent was calculated as the difference between the average talk and processing time per call before the training and the average talk and processing per call after the training. As stated in the finings pertaining to hypothesis two, the average time to handle a call (talk plus processing time) before the training intervention was 310.72 seconds, and the average time to handle a call after the training intervention was 315.35 seconds, an additional 4.63 seconds.

Second, the total employee wages saved per 1,000 calls were calculated by multiplying the total time saved per call times the agents' average hourly wage. As a result, the average time to handle a call after the training actually increased 4.63 seconds. It was a total of 4,630 seconds increase for 1,000 calls and a $16.385 cost of wages.

The third measurement was total sales, which was calculated by multiplying the number of room nights by the average daily room rate (ADR). As shown in the following Table 7, the sales ranged from $42,015.50 to $352,200.42 (*M* = 197667.72) before training and from $42,690.96 to $401,443.85 (*M* = 201622.70) after training. The number of bookings ranged from 76 to 560 (*M* = 346.52) before training and from 70 to 653.50 (*M* = 345.65) after training. The number of room nights ranged from 170 to 1287.00 (*M* = 751.22) before training and from 156.00 to 1459.00 (*M* = 760.64) after training.

The fourth measurement of the mean increase in the sales per call ratio was significant. The median increase in sales per call was $18.11, with 71% of the agents improving their sales per call, which supports hypothesis three, that the reservations agents who completed the training contributed to increased organizational impact (Level 4).

The fifth measurement of the cost of training/sales, the total cost of training per agent was $688.80 dollars and the average gain in sales for each agent was $112.34, a 13.37% increase. Thus, the ratio was 1/.163. For every dollar spent for the training, the average sales gain for each agent was $1.16 above the average sales per agent before the training for the first two months following their training. The total amount gained for the 69 agents is $80.25 for the two months. Projecting the sales per agent for 12 month, assuming the average amount of sales remains the same, the per agent average sales was $6.98 and $481.48 for 12 months. This demonstrates a significant organizational impact of the training investment, and the acceptance of hypothesis three.

Hypothesis four (H_4). Employee learning (Level 2) and job performance (Level 3) will predict organizational impact (Level 4). To answer research hypothesis four, the differences from pre to post training on the learning, performance and impact variables were utilized for correlations and multiple regression analyses. The study found that the increases in sales per call were significantly associated with the improvement of call score, conversion ratio, and average talk time. A hierarchical regression predicting increase in sales per call from pre-training to post-training from increases in call score and conversion were significant. Call score ($\beta = .237$, $p < .05$) contributed unique variance to the prediction of sales per call ($R^2 = .027$, $p < .05$) in the regression equation. After controlling for call score, conversion ($\beta = .903$, $p < .001$) contributed additional variance to the prediction of the increase in sales per call ($R^2 = .809$, $p < .001$) in the regression equation. These findings suggest that sales per call can be predicted by the call score and conversion. On the other hand, average talk time and average processing time did not make a statistically significant contribution to the regression equation. These

findings suggest that sales per call can be predicted by call score and conversion. Overall, the regression model explained 83.6% of the variability of increase in sales per call. Although two other variables (average talk time and average processing time) were not significant in the model, hypothesis four was accepted.

Implications for Theory, Research, and Practice

Despite the criticisms and the development of other comprehensive evaluation models, Kirkpatrick's model is still being widely utilized due to its simplicity and practicality (Kirkpatrick & Kirkpatrick, 2006; Twitchell, 1997). From the findings and conclusions of this research, some recommendations and implications for human resource theory development, research, and practice are presented.

Implications for Theory

Training evaluation has been debated and discussed for decades since Kirkpatrick initiated the concept of evaluation and the model of evaluation in 1959. It is evident in the literature that the needs for HRD accountability and results continue to grow. One of the greatest challenges is creating, developing, and using evaluation methods.

Due to the common misunderstandings of time constraints, personnel, belief in the value of the evaluation process, and the complexity for higher levels of evaluations, many HRD efforts still emphasize the lower levels of evaluation of Kirkpatrick's model. In addition, the concerns of what the financial impact evaluations should be present barriers for higher levels of evaluations. Nonetheless, Kirkpatrick's four levels evaluation model still serves as an effective guide for conducting training evaluation even though it has been a half of a century since its debut. As stressed by Kirkpatrick (1959b; Kirkpatrick & Kirkpatrick, 2005, 2006), evaluation of the behavior (job performance) is

more complicated, difficult, and time-consuming than the evaluation of reaction to the training and evaluating what was learned (Levels 1 and 2). Consequently, Kirkpatrick believed that Level 3 is the forgotten level. Lots of time, energy, and expense are put into Levels 1 and 2 because these are the levels that they have the most control over. However, executives are interested in level 4, and that is as it should be. Therefore, it leaves Level 3 out there on its own with no one really owning it.

The main objective of this study was to demonstrate whether a sales training program in the hospitality industry supported Kirkpatrick's evaluation model. This study supported his theories by implementing all four levels of evaluation as fully as possible at an organization. The implementation on only Levels 1 and 2 will not be a valid predictor of Levels 3 and 4. Implementing just the higher levels will not validate the learner's reaction (Level 1) or learning (Level 2) either.

It should be noted again that most of the evaluation models found in the literature are generally based upon Kirkpatrick's four levels (Bomberger, 2003; DeSimone & Harris, 2002; Werner & DeSimone, 2005; Goldwasser, 2001). Kirkpatrick's model, is outcome and objective-oriented and focuses on determining the effectiveness of a program. In other words, it is a summative evaluation model, which only takes place after the training program has been conducted in order to assess the merit and worth of the training program, and provide a summary report of the training outcomes for consideration of its continuation and/or its improvement. However, as argued by Kirkpatrick, based on the evaluation results, decisions to continue or alter the training program can be made accordingly. The summative evaluation results can turn into formative evaluation for instrument development, future program improvements, and/or

modifications (Kirkpatrick & Kirkpatrick, 2006). As demonstrated in this study, when done thoroughly, Kirkpatrick's summative evaluation model has a strong theoretical base that is valid and implementable. It would be strengthened with a viable formative evaluation system and theoretical base. This could be an area warranting further research.

An assumption in the literature indicates that the levels of Kirkpatrick's model are sequential. Level 1 is the lowest level on the hierarchy. While Level 2 can predict Level 4 outcomes, the prediction is enhanced by Level 3 performance data. The findings from this study indicated that learning occurred (Level 2) in the training, job performance improved (Level 3), and organizational results (Level 4) were achieved. This seems to reflect assumption that the four levels are sequential. While Level 2 is confirmed to be able to predict Level 4 outcome, adding Level 3 increases the predictability. Further research is recommended to examine the sequential relationships among the four evaluation levels of the Kirkpatrick (1959a) model as found in the literature (Alliger & Janak, 1989). That is, favorable trainee reactions help in assuring learning that assist in applying the learned skills to the job, which finally lead to favorable results in the individual and organizational levels. More research is still needed to further test Kirkpatrick's theory to its full extend.

Implications for Research

This study provides the groundwork for additional research into the effectiveness of training programs of Kirkpatrick's model as a whole, and also for each level, particularly Level 3. The findings from this study support the main body of literature and Kirkpatrick's theories of his evaluation. The research findings and the empirical links are addressed as the following.

Level 1. Participant reaction (Level 1) evaluation provides a basis for developing a balanced set of measures as long as data are provided that can improve facilitation and program implementation, and if there is predictive value in the measures. The measurement instruments usually request comments about the training content, materials, instructors, facilities, delivery methodology, etc. Kirkpatrick (Kirkpatrick, 1959a; Kirkpatrick & Kirkpatrick, 2005, 2006) strongly recommended obtaining candid responses by using anonymous reaction sheets where the trainees are not required to identify themselves or sign the forms. Holton (1996), one of the most critical of Kirkpatrick's model, contends that reactions should not be considered a primary outcome of training, believing that favorable reactions and learning are not necessarily related (Holton, 1996; Holton & Naquin, 2004). Kirkpatrick emphasizes that Level 1 is important because positive reactions to a training program may encourage employees to attend future programs. In contrast, negative comments about the program may discourage learners from attending and/or completing the program.

In this study, the sales training was mandatory for the reservations sales agents. Level 1 evaluation was not possible due to the data not being available. The inability to acquire Level 1 data for this study presented a challenge to examine the employees' reactions relate to Levels 2, 3, and 4, and to provide complete recommendation for program improvement. And because favorable reactions to training do not, by itself, guarantee that learning (Level 2), performance (Level 3) has occurred, Kirkpatrick stressed that many organizations and HRD professionals are overlooking the importance of Level 1 evaluation (Kirkpatrick, 1959a; Kirkpatrick & Kirkpatrick, 2005, 2006). This might have been the case for the Hotel. While interests in accountability and higher levels

of evaluation grow, future research of training program should still conduct Level 1 evaluation. It also a key source on how to improve future training programs (Kirkpatrick, 1998, p. 17). Level 1 evaluation should be included to thoroughly examine its predictability for Levels 2, 3, and 4.

Level 2. Kirkpatrick's Level 2 is content evaluation, the examination of whether employees changed attitudes, improved knowledge, and/or increased skills as a result of participating in the program (Kirkpatrick & Kirkpatrick, 2006). It is evident in the literature that Level 2 evaluations are still one of the most popular forms to evaluate the training program effectiveness despite research that does not support that acquired knowledge and skills equates to behavioral changes on the job performance (Bersin, 2003; Strunk, 1999). However, Kirkpatrick stressed that evaluating learning is important. Without measuring learning, no change in behavior can be validated.

Therefore, one of the major reasons for measuring learning is to determine whether learning is transferable to the job. In this study, the reservations sales agents' call score was used for measuring Level 2 learning performance. The positive improvement of learning was detected and helped to explain and predict Levels 3 and 4 results. The implications for the future research are to continue measuring Level 2 performance.

Level 3. It measures employees' job performance by determining the extent to which employees apply their newly acquired knowledge and skills on the jobs. This level is critical, as it addresses the issue of learning transfer. If employees cannot apply what they learned to their job, the training effort cannot have an impact on the organizational results (Level 4). No results can be expected unless a positive and measurable change in behavior (performance) occurs.

In this study, the identified job performance variables for reservations sales agents were call conversion, average talk time, and average processing time. The research findings indicated that call version improved significantly after the training. However, the time usage of both average talk time and average processing time did not show significant improvement. Nonetheless, it still demonstrated partial job performance improvement and established the link between Levels 2 and 4.

Level 4. It is the most important and also the most challenging level to assess (Werner & DeSimone, 2005; Kirkpatrick, 1960b; 1998; Phillips, 1996a). It is critical for programs designed to influence impact measures such as output, quality, cost, and time (Phillips, 2003a). It is also frequently found in the literature is that the most important barrier to training evaluation is all the costs related to training. As identified in this study, those costs could be the training materials, the employees' salaries, and learning coach's (facilitator's) fee. In many other cases, there would be more costs involved in training investments such as travel, accommodations, facility usage, etc., and many stakeholders perceive training investment is too costly.

ASTD's latest 2009 report estimated that U.S. organizations spent $134.07 billion on employee learning and development in 2008 (Paradise & Patel, 2009). The average annual expenditure per employee in the ASTD's sample organizations increased to $1,103 per employee in 2007, an increase of 6 % from 2006 (Paradise, 2008). The finding in 2008 was slightly down 3.8 % from the 2007 level to $1,068 (Paradise & Patel, 2009). While many may consider the individual reservations sales agent's training cost of $688.80 was high, it was still less expensive while comparing to the ASTD finding of 2007 level of $1,068.

In addition, Kirkpatrick and Kirkpatrick (Kirkpatrick & Kirkpatrick, 2005, 2006) stressed that obtaining objective measures, such as sales per trainee or sales to quota, to measure results is administratively infeasible and difficult, because factors other than the salesperson's efforts can have an influence on sales volume. However, the attempt is still crucial as the Level 4 results are often used to justify the existence of the training department and to decide whether to continue or discontinue training programs. In this study, the sales per call as identified as the Level 4 result showed a significant increase after the training.

As shown in this study, by examining the sales increase after the training and comparing the sales against the cost of training, the results not only demonstrate the value, but also validate the program. A single use or snapshot result may not be reliable, but continued refinement of the process can increase its credibility as a part of the evaluation. The framework developed through this research should be considered for further research.

Critique and problematic assumptions of Kirkpatrick's Evaluation Model. The first assumption frequently found in the literature is that the levels are arranged in ascending order and the model is hierarchical in nature. Therefore, the higher levels are more valuable and important than the lower ones. With this notion, many HRD professionals purport to skip the lower levels of evaluations and focus on the higher levels of evaluations. This is questionable, as shown in the empirical review that few reported studies have addressed Levels 3 and 4. Also, Kirkpatrick (1959a; Kirkpatrick & Kirkpatrick, 2005, 2006) contends that it is a serious mistake to bypass Levels 1 and 2

evaluations and only conduct Level 3 and 4 evaluations. This will easily lead to the wrong conclusions about the effect of each level and the training program's overall result.

The second assumption is that the four levels of evaluation are causally linked. Based on this assumption, many researchers and HRD professionals presume that positive reactions are the prerequisite for learning to occur. Once learning has occurred, desired behaviors will change and ultimately lead to positive organizational results (Alliger & Janak, 1989; Hilber, Preskill & Ress-Eft, 1997; Kirkpatrick & Kirkpatrick, 2005, 2006). However, Holton (1996) strongly claimed that Kirkpatrick's model failed to demonstrate the causal relationships between the levels.

The second assumption leads to the third assumption, that the four levels are positively intercorrelated. If these two assumptions were true, it would be sufficient just to evaluate whether employees have positive reactions (Level 1) to the training program, from which it could be assumed they learned from the training, they ultimately would improve their job performance, and positively contribute to the organizational results. Addressing these assumptions, Kirkpatrick (1959a), and Kirkpatrick and Kirkpatrick (2005; 2006) emphasized that there is no guarantee that a favorable reaction to the training program assures learning, positive behavioral change, and favorable organizational results. This is why it is important to evaluate both reaction (Level 1) and learning (Level 2) in case no change in behavior (Level 3) occurs.

Although two Level 3 variables identified in this study did not show significant changes and contribute to the organizational impact, the study still provided a thorough evaluation of Kirkpatrick's model. The implications represent professional training situations in many organizational settings.

In this study, Level 2, learning, did occur, Level 3, job performance, did improve, and it resulted in Level 4, a positive organizational impact. Organizational results (Level 4) were detected, and were associated with the employees' acquired knowledge and skills (Level 2) and changes of behaviors that lead to job performance improvement (Level 3). In other words, Level 2 (call score) and Level 3 (conversion) can be used to predict Level 4 (sales/call).

Limiting evaluation to one particular level might not provide an adequate picture of the overall effectiveness of any training program. As interest in accountability and results grow, emphasis may be placed on enhancing current evaluation practices at the higher levels of evaluation for even the smallest organizations. The implementations and findings from this study should be considered and generalized to any business that emphasizes every level of evaluations. In this study, the comparisons were made two months before and after the training intervention. The decision was made based on the consideration of seasonality factor that occurs in the hospitality industry. Two months were utilized for further examining the training effectiveness while avoiding the seasonality variable. The recommendation for future research would be extending the length of study to detect whether the performance changes over a longer period of time. Also, further research into the relationships among the four levels for training is still needed and recommended. An experimental design study, and/or a meta-analysis, and/or a study that examines the possible interactions between the variables identified in this study are recommended. In addition, qualitative research could provide insight into various problems, such as identifying some of the underlying factors that account for the weak but statistically significant relationships found in this study. Qualitative research

may also be helpful to identify variables that have not yet been considered or quantitatively tested.

Implications for Practice

The implementations and findings from this study should be an encouragement for the hospitality industry to further investigate their training endeavors in different segments and areas. Every business should consider implementing Kirkpatrick's evaluation model by identifying their unique critical levels of performance, eliminating or modifying ineffective programs, ensuring training dollars are spent wisely, and enhancing the impact of the organization.

Because sales training is a very complex process, a single level of measurement of sales training will not provide a comprehensive picture of the program. Similar studies should be considered at different hotel chains across different regions of the world. Within a hotel chain this study could be replicated in other units such as airlines reservations centers, hotel sales departments, catering or banquet departments, event planning, food and beverage department, etc.

Besides replicating a similar study with similar sales training program, different delivery methods and scheduling formats should also be considered for future research. Since the emphasis in today's hospitality industry is on both productivity and service, the reservations sales agents have limited time for attending days-long training. Future research could investigate whether the same material is being placed in an on-line format or blended format remains just as effective and whether going through the entire training via smaller sessions make any difference.

A comprehensive evaluation of sales training programs, as demonstrated in this research, is difficult to conduct. Despite these difficulties, the sales training program evaluations can and should be performed as was demonstrated in this study. With the advancement of computer technology and the acknowledgement of the importance of data acquisition and management, every hospitality business should collect performance data on different levels so comprehensive analysis can be performed. As demonstrated in this research, both individual and organizational performance data could be recorded and collected. This minimizes the concerns often found in the literature that training evaluations are complex and infeasible. More studies of effective practices are needed to document processes and procedures for designing and implementing these evaluations.

Concluding Remarks

This research was an initial attempt to develop an extensive evaluation system to assess a training program in a hospitality organization. The objective was to provide the first fully implemented study to investigate correlations among all levels of Kirkpatrick's model as they relate to a sales training course. Although it was not the objective of this study to provide instruments that could be used for all types of training, the assessment of these particular instruments could provide insight for other training professionals attempting to design effective evaluation instruments in their particular field. While this study hopefully contributed to the research of effective training programs, more research is needed to fully understand the drivers for increased accountability and the conditions under which appropriate evaluation can take place.

REFERENCES

Abernathy, D. (2003). *A guide to online learning service providers.* Retrieved December 10, 2003, from http://www.learningcircuits.org/oct2000/abernathy.html

Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology, 42*, 331-342.

American Society for Training and Development (ASTD). (2007). Retrieved December 15, 2006, from http://www.astd.org

American Society for Training and Development (ASTD). (2009). *The value of training: Making training evaluations more effective*. Alexandria, VA: ASTD Press.

Arthur, W., Bennett, W., Edens, P. S. & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology, 88*(2), 234-245.

Attia, A. M. (1998). Measuring and evaluating sales force training effectiveness: A proposed and an empirically tested model. (Doctoral dissertation, Old Dominion University, 1998). *Dissertation Abstract International, A59/09,* 175.

Barrow-Britton, D. B. (1997). Formative evaluation of a computer based interactive multimedia presentation for adult education in gaming. (Doctoral dissertation, Northern Arizona University, 1997). *Dissertation Abstract International, A58/10,* 208.

Bassi, L., Benson, G. & Cheney, S. (1996, November). Top ten trends. *Training and Development*, *50*, 28-42.

Bassi, L., & Van Buren, M. (1998). The 1998 ASTD state of the industry report. *Training and Development, 52*(1), 21-43.

Bassi, L., & Van Buren, M. (1999, January). The 1999 ASTD state of the industry report. *Training and Development*, 2-27.

Benabou, C. (1996). Assessing the impact of training programs on the bottom line. N*ational Productivity Review, 15*(3), 91-98.

Bersin, J. (2003, June). *E-learning analytics*. Retrieved September 6, 2006, from http://www.learningcircuits.org/jun2003/bersin.htm

Bledsoe, M. D. (1999). Correlations in Kirkpatrick's training evaluation model. (Doctoral dissertation, University of Cincinnati, 2000). *Dissertation Abstract International, A60/07,* 54.

Bomberger, D. W. (2003). Evaluation of training in human service organizations: A qualitative case study. (Doctoral dissertation, The Pennsylvania State University, 2003). *Dissertation Abstract International, A64/12,* 162.

Bowers, B. (2007, February). *Smith Travel Research*. Retrieved August 10, 2007, from http://www.nbta.org/NR/rdonlyres/9826D473-2B40-4260-8382-A6D23A0F0EE3/0/FF07BowersBobby.pdf

Brinkerhoff, R. (1981). Making evaluation more useful. *Training Development Journal*, *35*(12), 66-70.

Brinkerhoff, R. (1989). *Evaluating training programs in business and industry*. San Francisco, CA: Jossey-Bass.

Brinkerhoff, R. O., & Gill, S. J. (1994). *The learning alliance*. San Francisco, CA: Jossey-Bass.

Brinkerhoff, R. O. (1983). The success case: A low-cost, high-yield evaluation. *Training and* Development, *37*(8), 58-61.

Brinkerhoff, R. O. (1987). *Achieving results from training.* San Francisco, CA: Jossey-Bass.

Bromley, P., & Kitson, B. (1994, January). Evaluating training against business criteria. *Journal of European Industrial Training*, *18*(1), 10-14.

Bushnell, D. S. (1990). Input, process, output: A model for evaluating training. *Training and Development Journal, 44*(3), 41-43.

Caffarella, R. (1988). *Program development and evaluation: Resource book for trainers*.New York: John Wiley & Sons.

Cascio, W. (1989). Using utility analysis to assess training outcomes. In Goldstein and associates (Eds.). *Training and Development in Organizations* (pp. 63-88). San Francisco, CA: Jossey-Bass.

Cascio, W. (2000). *Costing human resources: The financial impact of behavior in organization.* Cincinnati, OH: South-Western College Publishing.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Creswell, J. W. (2005). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (2nd ed.). Upper Saddle River, NJ: Pearson Education .

Delerno, J. (2001, September). *Raise the bridge or lower the river – Internet based training opportunities for the hospitality industry*. Retrieved June 6, 2003, from http://hotel-online.com/News/PR2001_3rd/Sept01_IBT_Delerno.html

DeSimone, R. L., & Harris, D. M. (2002). *Human resource development* (3rd ed.). Orlando, FL: The Dryden Press.

DeVeau, P. M. (1995). Utilization of multimedia computer technology in corporate training and development programs: A survey study. (Doctoral dissertation, University of Bridgeport, 1995). *Dissertation Abstract International, A56/08,* 155.

Dick, W., & Carey, L. (1996). *The systematic design of instruction* (4th ed.). New York: Longman.

Driscoll, M. (2001, August). *Strategic plans from scratch*. Retrieved September 12, 2006, from heep://www.learningcircuits.org/2001/aug2001/driscoll.htm

Farrell, D. (2005). What's the ROI of training programs? *Lodging Hospitality, 60*(7), 46.

Feiertag, H., & Hogan, J. (2001). *Lessons from the field: A common sense approach for effective hotel sales.* Brentwood, TN: JM Press.

Flynn, G. (1998, November). Tool: The nuts & bolts of valuing training. *Workforce, 77*(11) 80-85.

Freitag, J. D. (2006a, January). *Smith Travel Research*. Retrieved August 10, 2007, from http://www.amginc.com/HSMAI/ResortConference/2006/06Presentations/ResortConf-smith%20travel.ppt

Freitag, J. D. (2006b, July). *Smith Travel Research*. Retrieved August 10, 2007, from http://www.krisam.com/pdf/Smith_Travel_Research_Presentation-Jan_Freitag.pdf

Freitag, J. D. (2006c, August). *Smith Travel Research*. Retrieved August 10, 2007, from http://www.hotelsmag.com/archives/2007/03/web_sr/VOIC_orlando_1003.pdf

Gagné, R. M., & Medsker, K. L. (1996). *The conditions of learning: Training applications*. Orlando, FL: Harcourt Brace & Company.

Galvin, J. C. (1983). Evaluating management education: Models and attitudes of training specialists. (Doctoral dissertation, Northern Illinois University, 1983). *Dissertation Abstract International, A44/05,* 146.

Gay, L. R. (1996). *Educational research: Competencies for analysis and applications* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Gay, L. R., & Airasian, P.W. (2002). *Educational research: Competencies for analysis and applications* (7th ed.). Englewood Cliffs, NJ: Prentice Hall.

Gilley, J. W., Eggland, S. A., & Gilley, A. M. (2002). *Principles of human resource development* (2nd ed.). Cambridge, MA: Perseus Books Group.

Goldstein, I. (1986). *Training in organizations: Needs assessment, development, and evaluation*. Pacific Grove, CA: Brooks/Cole.

Goldwasser, D. (2001, January). Beyond ROI. *Training, 38*(1), 82-90.

Hackett, B. (1997). The value of training in the era of intellectual capital: A research report. *The Conference Board*, Report number 1199-97-RR, 5-30.

Hall, B. (2001). Corporate drivers of e-learning. In Mantyla & Woods (Eds.), *The 2001/2002 ASTD distance learning yearbook* (pp. 171-173). New York, NY: McGraw-Hill.

Hart, P. H. (1992). Interactive video: Strategic Implications for self-directed training. (Doctoral dissertation, Northern Illinois University, 2002). *Dissertation Abstract International, A53-04,*111.

Hilbert, J., Preskill, H. & Russ-Eft, D. (1997). Evaluating training. In Bassi, L. & Russ-Eft, D. (Eds.). *What works: Assessments, development, and measurement*. Alexandria, VA: American Society for Training and Development.

Holton, E. F., III. (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly, 7*(1), 5-29.

Holton, E. F., III. (2005). Holton's Evaluation Model: New Evidence and Construct Elaborations. *Advances in Developing Human Resources, 7*(1), 37-54.

Holton, E. F., III, & Naquin. S. S. (2004). New Metrics for Employee Development. *Performance Improvement Quarterly, 17*(1), 56-80.

Honeycutt, E., Ford, J. & Rao, C. P. (1995). Sales training: Executives' research needs. *Journal of Personal Selling & Sales Management, 15*(4), 67-71.

Hospitality Service Alliance (HSA) International (2007). Retrieved August 1, 2007, from http://www.hsa.com

Hudson, S. M. (2002). A qualitative study of learner-centered training: A corporate view of Web-based instruction. (Doctoral dissertation, Memphis State University, 1992). *Dissertation Abstract International, A63-06,* 141.

Ismail, A. (2002). *Front office operations and management*. Florence, KY: Cengage Delmar Learning.

Jackson, T. (1989). *Evaluation: Relating training to business performance*. San Diego, CA: University Associates.

Kaufman, R., & Keller, J. M. (1994, Winter). Levels of evaluation: Beyond Kirkpatrick. *HRD Quarterly, 5*(4), 371-380.

Kauffman, R., Keller, J. & Watkins, R. (1996). What works and what doesn't work: Evaluation beyond Kirkpatrick. *Performance & Instruction, 35*(2), 8-12.

Kidder, P., & Rouiller, J. (1997). Evaluating the success of a large-scale training effort. *National Productive Review, 16*(2), 79-89.

Kim, I. Y. (2006). Evaluating an instructor training program in a church setting. (Doctoral dissertation, University of Southern California, 2006). *Dissertation Abstract International, A67/10,* 114.

Kirkpatrick, D. L. (1959a). Techniques for evaluating training programs: Reaction. *American Society for Training and Development Journal*, *18*, 3-9.

Kirkpatrick, D. L. (1959b). Techniques for evaluating training programs: Learning. *American Society for Training and Development Journal*, *18*, 21-26.

Kirkpatrick, D. L. (1960a). Techniques for evaluating training programs: Behavior. *American Society for Training and Development Journal*, *19*, 13-18.

Kirkpatrick, D. L. (1960b). Techniques for evaluating training programs: Learning. *American Society for Training and Development Journal*, *18*, 28-32.

Kirkpatrick, D. L. (1996, January). Great ideas revisited: Revisiting Kirkpatrick's four-level model. *Training & Development, 50*(1), 54-57.

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

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2005). *Transferring learning to behavior: Using the four levels to improve performance*. San Francisco, CA: Berrett-Koehler Publishers.

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: The four levels* (3rd ed.). San Francisco, CA: Berrett-Koehler Publishers.

Kraiger, K., Fords, J. & Salas, E. (1993, February). Application of cognitive, skill-based, and affective theories of learning outcomes to methods of training. *Journal of Applied Psychology, 78*, 311-328.

Lanigan, Mary Louise. (1997). Applying the theories of reasoned action and planned behavior to training evaluation levels. (Doctoral dissertation, Indiana University, 1997). *Dissertation Abstract International, A58/03,* 123.

Larsen, N. B. (1985). Implementation and meta-evaluation of an experimental method for evaluating an administrator training program. (Doctoral dissertation, Western Michigan University, 1985). *Dissertation Abstract International, A47/01,* 128.

Lockwood, S. L. (2001). Enhancing employee development: Development and testing of a new employee orientation protocol. (Doctoral dissertation, California School of Professional Psychology - San Diego, 2001). *Dissertation Abstract International, A62/03,* 166.

Lomanno, M. V. (2005, October). *Smith Travel Research*. Retrieved August 10, 2007, from http://www.hospitalitynet.org/file/152002287.ppt

Mager, R. F. (1984). *Preparing instructional objectives* (2nd ed.). Belmont, CA: David S. Lake Publishers.

McNeil, K., Newman, I., & Kelly, F. J. (1996). *Testing research hypotheses with the general linear model*. Carbondale, IL: Southern Illinois University Press.

Merriam, S. B., & Simpson, E. L. (1995). *A guide to research for educators and training of adult education.* (2nd ed.). Malabar, FL: Krieger Publishing Company.

Newman, D. R., & Hodgetts, R. M. (1998). *Human resource management: A customer-oriented approach*. Upper Saddle River, NJ: Prentice-Hall, Inc.

Newman, I., & Newman, C. (1994). *Conceptual statistics for beginners* (2nd ed.). Lanham, MD: University Press of America.

Newman, I., Newman, C., Brown, R. & McNeely, S. (2006). *Conceptual statistics for beginners* (3rd ed.). Lanham, MD: University Press of America.

Nickols, F. W. (2005, Feb.). Why a stakeholder approach to evaluating training. *Advances in Developing Human Resources, 7*(1), 121-134.

Paradise, A. (2007). *The ASTD 2007 state of the industry report.* Alexandria, VA: American Society for Training and Development.

Paradise, A. (2008). *The ASTD 2008 state of the industry report.* Alexandria, VA: American Society for Training and Development.

Paradise, A., & Patel, L. (2009). *The ASTD 2009 state of the industry report.* Alexandria, VA: American Society for Training and Development.

Phillips, J. H. (2000). Evaluating training programs for organizational impact: Five reports. (Doctoral dissertation, Wayne State University, 2000). *Dissertation Abstract International, A61/03,* 187.

Phillips, J. J. (1991). *Handbook of training evaluation and measurement methods*. Houston, TX: Gulf Publishing Company.

Philips, J. J. (1996a). ROI: The search for best practices. *Training & Development, 50*(2), 42-47.

Philips, J. J. (1996b). Was it the training? *Training & Development, 50*(3), 28-32.

Philips, J. J. (1996c). How much is the training worth? *Training & Development, 50*(4), 20-24.

Phillips, J. J. (1998, August). The return-on-investment (ROI) process: Issues and trends. *Educational Technology, 38*(4), 7-14.

Phillips, J. J. (1999). *HRD trends worldwide: Shared solutions to compete in a global economy*. Boston, MA: Butterworth-Heinemann.

Philips, J. J. (2003a). *Return on investment in training and performance improvement programs* (2nd ed.). Philadelphia, PA: Elsevier Science & Technology.

Phillips, P. P. (2003b). Training evaluation in the public sector. (Doctoral dissertation, The University of Southern Mississippi, 2003). *Dissertation Abstract International, A64/09,* 215.

Pine, J., & Tingley, J. C. (1993, February). ROI of soft-skills training. *Training, 30,* 55-58.

Plant, R., & Ryan, J. (1992, October). Training evaluation: A procedure for validating an organization's investment in training. *Journal of European Industrial Training, 16*, 22-38.

Posavac, E. J., & Carey, R. G. (1997). *Program evaluation methods and case studies* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.

Rivera, R. J., & Paradise, A. (2006). *The ASTD 2006 state of the industry report.* Alexandria, VA: American Society for Training and Development.

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

Setaro, J. (2001, June). *Many happy returns: Calculating e-learning ROI.* Retrieved December 6, 2004, from http://www.learningcircuits.org/2001/jun2001/Elearn.htm

Shelton, S., & Alliger, G. (1993, June). *Who's afraid of level evaluation?: A practical approach.* Training & Development, 43-46. Retrieved February 15, 2007, from http://arapaho.nsuok.edu/~philljam/proposals/Level_4_Evaluation.doc

Speizer, I. (2005, July). *Training's holy grail: ROI*. Retrieved September 8, 2006, from http://www.forkforce.com.archive/feature/24/10/90/241092_printer.php

Spitzer, D., & Conway, M. (2002). *Link training to your bottom-line.* Infoline. Alexandria, VA: ASTD.

Strunk, K. S. (1999). Status of and barriers to financial impact evaluations in employer-sponsored training programs. (Doctoral dissertation, University of Arkansas, 1999). *Dissertation Abstract International, A60/06*, 148.

Stufflebeam, D. (1983). The CIPP model for program evaluation. In G. Madeus, M. Scriven & D. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human service evaluation* (pp. 117-142). Boston, MA: Klewer Nijhoff.

Swanson, R. A. (2001). *Assessing the financial benefits of human resource development*. Cambridge, MA: Perseus Publishing.

Swanson, R. A., & Gradous, D. B. (1988). *Forecasting financial benefits of human resource development.* San Francisco, CA: Jossey-Bass.

Tabachnick, B., & Fidell, L. (2000). *Computer-assisted research design and analysis*. Upper Saddle River, NJ: Pearson Education.

Tanke, M. L. (1999). *Human resources management for the hospitality industry* (2nd ed.). Albany, NY: Delmar Publishers.

Tidler, Karen Louise. (1999). Evaluation of continuing medical education using Kirkpatrick's four levels of evaluation. (Doctoral dissertation, The University of New Mexico, 1999). *Dissertation Abstract International, A61/01,* 140.

Tung, F. C. (1998). Factors that impact the implementation of multimedia in hotel training: A survey study. (Doctoral dissertation, University of Nebraska - Lincoln, 1998). *Dissertation Abstract International, A60/01,* 208.

Twitchell, S. (1997). Technical training program evaluation: present practices in United States business and industry. (Doctoral dissertation, Louisiana State University and Agricultural and Mechanical College, 1997). *Dissertation Abstract International, A58/09,* 152.

Ulrich, D. (1997). Measuring human resources: An overview of practice and a prescription for results. *Human Resource management, 36*(3), 303-320.

Van Buren, M. E. (2001). *The 2001 ASTD State of the industry report.* Alexandria, VA: ASTD.

Van Buren, M. E., & Erskine, W. (2002). *The 2002 ASTD State of the industry report*. Alexandria, VA: ASTD.

Warr, P., Bird, M. & Rackham, N. (1970). *Evaluation of Management Training*. London, England: Gower Press.

Warr, P. B., & Bunce, D. (1995). Trainee characters and the outcomes of open learning. *Personnel Psychology*, *48*, 347-375.

Werner, J. M., & DeSimone, R. L. (2005). *Human resource development* (4th ed.). Mason, OH: Thomson South-Western.

Wertz, C. (2005). Evaluation of CLAD training in northern California. (Doctoral dissertation, University of Southern California, 2005). *Dissertation Abstract International, A66/06,* 108.

Yaw, D. C. (2005). An evaluation of e-learning in industry at Level Three based upon the Kirkpatrick model. (Doctoral dissertation, Indiana State University, 2005). *Dissertation Abstract International, A66/12,* 136.

Appendix A

Schedule of the Training Program

Situational Selling, Focus on the Customer

Situational Selling, Focus on the Customer

Course Overview

Hotel's reputation and success is measured in many ways but the first impression our customers receive is often delivered by our front line telephone Reservation Sales Agents. Our ability to connect with our customers is what sets us apart. This program was designed to enhance the performance of telesales professionals at the Hotel's Global Reservation Centre.

Day 1 - Introduction to Situational Selling

Attitudes for Success

Unit 1 - Having a strong belief in self – works on the premise that an agent will sell as well as they feel

Topics covered: High Self-Esteem

Positive Self-Talk

Positive Force for Others

Clear Values

Unit 2 - Being Goal Oriented

Topics covered: SMART Goals

Positive Affirmations

Creative visualization

Skills for Success

Unit 3 - Pre-Call Planning

Topics Covered: Identify the different Market Segments

Identify Features and Benefits that fit each of those market segments

Tele-time Management

Day 2

Unit 4 - Cultivating

Topics Covered: Methods of Communication

Impact of non-verbal communication in tele-sales

Connecting with your customers by phone

Voice Quality

Positive statements

Unit 5 - Discovering

Topics Covered: The use of questions

Different types of questions

Effective Listening

Planning questions strategy

Skill Practice – Roll Play #1

Unit 6 - Presenting Recommendations

Topics Covered: Making a recommendation

Benefit Statements that work

Day 3

Unit 7 - Confirming

Topics Covered: Why people buy

Gaining commitment

Dealing with customer responses

Handling Buyers Concerns

Finishing the Call

Skill Practice Roll Play #2

Assignment

Appendix B

The Hotel's Call Quality Scoring Criteria

Appendix B

Call Quality Assessment for:
For the month of:
Team Leader: Kim Ayles & Gena Richard

1) Sales Techniques	***WDC***	***RYH***	***WDC***	***RYH***	***CLL***	***Numerator***	***Denominator***
a) Has the guest stayed with us before? (prior to offering room types and rates) (1 pt)						*0*	*0*
b) If repeat guest - no hotel or destination overview needed *New guest - offer to create a mental picture of hotel and destination (1 pt)*						*0*	*0*
c) Were the caller's needs identified (room and rate/reason for travel) and were suitable options provided based on these needs? (1 pt)						*0*	*0*
d) Was an appropriate room description offered?						*0*	*0*
e) Were benefits used to capture the sale? (1 pt)						*0*	*0*
f) Was the sale asked for? (1 pt)						*0*	*0*
g) Were buyers concerns overcome? (1 pt)						*0*	*0*
h) Was dining and activities reservation recommended? (1 pt)						*0*	*0*
i) Was cross-selling explored? (1 pt)						*0*	*0*

2) Professional Behaviors							
a) Promoting the Brand (1 pt) ***(Close with the hotel name)***						*0*	*0*
b) Professional Attitude:						*0*	*0*
1. Confidence (Knowledge and pride in product) (1 pt)	*Y*	*Y*	*Y*	*Y*	*Y*		
2. Energy Level (Tone, Pitch, Inflection, courteous phrases) (1 pt)	*Y*	*Y*	*Y*	*Y*	*Y*		
3. Customer Focus (Actively Listening and personalizing the conversation) (1 pt)	*Y*	*Y*	*Y*	*Y*	*Y*		
c) Using the caller's name efficiently and discreetly (1 pt)							
3) Accuracy (1 pt)						***0***	***0***

Total: ***#DIV/0!***

Call Feedback:

Call 1:

Call 2:

Call 3:

Call 4:

Call 5:

VITA

YA-HUI ELEGANCE CHANG

Place of Birth	Hsin-Chu, Taiwan
1995	Bachelor of Science Applied Science – Foodservice Management School of Human Ecology Fu-Jen Catholic University, Taipei, Taiwan
1995-1996	Assistant Manager Ponderosa Steakhouse, Taipei, Taiwan
1996-1998	Graduate Teaching Assistant School of Hospitality Management Florida International University, Miami, Florida
1998	Master of Science Hotel and Foodservice Management School of Hospitality Management Florida International University, Miami, Florida
1998-2003	Director of Distance Learning Hospitality Services Alliances International Sunrise, Florida
2001-Present	Doctoral Candidate Adult Education and Human Resource Development Leadership and Professional Studies College of Education Florida International University, Miami, Florida
2003-2008	Graduate Assistant, Training and eFolio Coordinator College of Education Florida International University, Miami, Florida
2009-Present	Instructor College of Hospitality Management Lynn University, Boca Raton, Florida