EFFECTS OF MANAGEMENT TRAINING ON TRAINEES' LEARNING, JOB PERFORMANCE AND ORGANIZATION RESULTS: A META-ANALYSIS OF EVALUATION STUDIES FROM 1983-1997

By

JIPING ZHANG

Bachelor of Science
Tsinghua University
P. R. China
1970

Master of Science
Tsinghua University
P. R. China
1981

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF EDUCATION
May, 1999

EFFECTS OF MANAGEMENT TRAINING ON TRAINEES' LEARNING, JOB PERFORMANCE AND ORGANIZATION RESULTS: A META-ANALYSIS OF EVALUATION STUDIES FROM 1983-1997

Thesis Approved:

Martin Burlingame
Thesis Adviser

Gary J. Conti

CT Combs

Ramesh Sheda

Wayne B. Ballard
Dean of the Graduate College

ACKNOWLEDGMENTS

At the moment of finishing this research and my doctoral program, I wish to express my sincere appreciation to several people who have helped me go through the process.

First, I would like to thank several professors: Dr. Burlingame, my supervisor, whose trust empowered me, whose support and effective guidance led me reach the final goal with solid steps. Dr. Conti, my committee member, who volunteered his time to audit my coding results, whose "Anatomy of Dissertation" helped me and could be helpful to every graduate student. Dr. Sharda, a business professor and the outside member of my committee, who has been always supportive and guided me to learn the Internet and other educational technology. Dr. Coombs who taught at my research methodology class and served as my committee member. Many thanks also go to Drs. Miller, Nolan and Folks who gave me support, and taught me knowledge. Most important, these professors set up an example for me as a student-oriented and service-oriented educator.

Second, I would like to thank my peer classmates in the doctoral program. Their hard working and life-long learning spirit inspired me to keep my dream alive, their support, help and sharing encouraged me to overcome difficulties, their diversified background, knowledge and expertise enriched my learning.

Third, I would like to thank the authors of the primary studies included in this study. Without their quality research, the meta-analysis would not be possible. Many thanks also go to the Oklahoma State University Library. The librarians' knowledge, skill and excellent service provided me materials necessary for this study.

I sincerely thank Dr. Joan Barrick, a retired teacher, who taught English to so many foreign students, and who edited all the papers I wrote during my doctoral program, including this dissertation.

At last, I would like to thank my husband, Dr. Zhou Jinfu who has been supportive to my academic endeavor, and has been an intelligent learning partner. I appreciate very much the support from my parents and other family members in Beijing, from my children—Zhou Yi (Mike) and Zhou Xu (Molly). I am proud to become the first doctor in my Zhang family. I hope that my effort will inspire the next generation in the family to achieve their goals through education.

TABLE OF CONTENTS

Chapter Page

Chapter Page

LIST OF TABLES

FIGURE

CHAPTER I

INTRODUCTION

Work organizations are experiencing and influenced by demographic evolvement, accelerated technology advancement and information explosion, as well as economic globalization. The economy has been reshaped by post-cold-war political world structure, technological innovation, and intensive market competition driven by ever demanding customers who request high quality, timeliness, variety, and customized products and services. As reactions and/or pro-actions to external factors of change, the workplace has adapted and transformed itself through organizational restructuring and development, work process reengineering, and job/task re-design (Bassi et al., 1996; Gordon, Morgan, & Ponticell, 1994; Nelson & Quick, 1996; Rummler & Brache, 1995; Swanson & Torraco, 1995).

Training and development have historically played an important role in work organizations to maintain and update both the management and the workforce to meet the requirements of current and future job performance needs (Harris & DeSimone, 1994; McLagan et al., 1989; Miller, 1996; Nadler, 1990; Swanson & Torraco, 1995). In the world of fast-paced changing technology and intensive global competition, workplace performance constantly requires new knowledge, skills and attitudes. What is happening in the workplace today has been placing an increasing challenge and premium on training and development in work organizations (Katzell, 1989).

American organizations believe strongly in the critical importance of effective managers, whose quality and performance may determine the organization's survival and prosperity (Wexley & Baldwin, 1986). They also believe that managerial knowledge, skills and abilities can be learned and improved (Keys & Wolfe, 1988; Saari et al., 1988; Wexley & Baldwin, 1986). Management training aims at job performance improvement of managers through formally organized instruction of the management-related knowledge, skills and attitudes. Together with management education and on-the-job experience, management training constitutes one dimension of management development (Wexley & Baldwin, 1986), which is one of the major categories of training and development in work organizations (Campbell, 1971; Goldstein, 1980; Latham, 1988, Tannenbaum & Yukl, 1992; Wexley, 1984).

McLagan (1989) pointed out that the dual nature of the workplace in the 90's was "simultaneously tougher, and more human" (p. 49). In the ruthless marketplace, competent and committed people are the most competitive advantage of an organization. More and more organizations realize that their people are their most precious asset. When heading into a new century, "training seems to have finally arrived as an endeavor that is acknowledged to contribute to the bottom line in organizations. ... and high-level managers in many organizations now consider training to be a strategic function that helps the organization fulfill its mission and reach its goals" (Gegne & Medsker, 1996, p. v). The insight and strategy of investing in people are evidenced by the continuous and steadily increasing investment of work organizations in their workforce and management development.

A 1992 survey conducted by *Training* magazine found that the dollars spent on training by companies were more than that spent on all post-secondary education in the U.S. in the same time period (cited by Zemke, 1994). The 1995 survey of *Employer-Provided Training* conducted by the U. S. Bureau of Labor Statistics, a benchmark investigation for training in work organizations, indicated that the number of employees trained and the money spent on training grew in most organizations in the mid-1990s. Based on this survey, American Society for Training & Development (ASTD) estimated that all U. S. organizations spent a total of $55.3 billion on training in 1995 (Benson, 1996). Data from the *Human Performance Practices Survey (HPPS)*, which was conducted jointly by ASTD and the Times Mirror Training Group, Development Dimensions International, the Forum Corporation, and the U. S. Department of Labor in 1997, suggested a similar figure for 1996. Estimates from the HPPS and the Benchmarking Forum suggested that the trend of increasing investment in training continues (Bassi & Van Buren, 1998).

American business spends billions of dollars each year to train and develop their employees (Clement, 1981). Phillips (1990) estimated that U.S. businesses spend at least $4 billion each year on management training and development programs alone. Management and executive development expenditures amounted to one quarter of the total training and development expenditures. The 1998 ASTD *State of the Industry Report* (Bassi & Van Buren, 1998) indicated that 93% of organizations offered Management-Supervisory skills courses, which was the second most popular course type offered by most employers; 63% of organizations offered executive development

courses. In terms of the percentage of the total training time, management-supervisory skills accounted for 12%, and executive development accounted for 3 per cent.

Nature of the Problem

Are work organizations' training dollars effective and worthwhile? In a meta-analysis which accumulated research findings of 177 training evaluation studies from 1960-1993, Bennett (1995) discovered that "training was more effective than expected" (p. iii). The HPPS survey aforementioned revealed strong correlation between a company's performance and its workplace learning and development investment and practices (Bassi & Van Buren, 1998). However, knowledge about the effect of training in organizations in general, and the effect of the management training in particular, is far from an adequate understanding of the complex system (Campbell et al. 1970; Clement, 1981; Hilbert, Preskill & Russ-Eft, 1997; Wexley & Baldwin, 1986). "Published and unpublished literature on the effectiveness of managerial training has produced conflicting results and left more unanswered questions than definitive statements concerning the effectiveness of managerial training" (Burke & Day, 1986, p. 232).

Four types of training outcomes, i.e., reaction, learning, behavior and results (Kirkpatrick, 1959, 1994) are commonly accepted and widely used by training researchers and practitioners as criteria for evaluating training effects. Reaction refers to how participants feel and think about the training and measures how well they liked the program. Learning refers to what the participants have learned from the program and measures the extent to which trainees have improved or increased their knowledge or

skills from the training. Behavior refers to changes of participants' performance in their job role and measures whether and how much trainees are transferring or applying their new learning to the workplace. Results refers to changes in organizational variables such as cost, productivity, and turnover and measures the effect of training on organizational achievement or business results.

It has always been a concern for training researchers and practitioners to demonstrate the effect of management training not only in measures of participants' satisfaction and learning, but also in behavior change on the job, and of contributions to the organization's bottom-line. The continued increase of training investment has fueled the call for accountability.

In addition to the five notable successive reviews on training and development in work organizations (Campbell, 1971; Goldstein, 1980; Latham, 1988; Tannenbaum & Yukl, 1992; Wexley, 1984) that discussed management training as a special program area or training function, two review articles focused on published empirical evaluation research studies on management training (Campbell et al., 1970; Clement, 1981). The authors of the latter two articles brought some progress to the qualitative review. First, they only selected empirical studies with some kind of experimental control—control group or both pre- and post-training measures. Second, as much as possible, they tried to include studies in their review, and used descriptive statistics to give a profile of the evaluation studies on management training at that time. There were 73 evaluation studies published before 1970 included in Campbell's review, and 26 published during the 1970s in Clement's. However, the information from the studies reviewed was basically inadequate, and led to no conclusion on which specific program or technique

leads to greater or lesser changes for certain attitude or behavior measured by certain criterion. Major critical shortcomings pointed out by Campbell et al. (1970, p. 323-325) included:

1. Almost exclusive reliance of management development research on internal criteria, i.e., participants reaction and learning. There is no simple relationship between an attitude change and a change in job performance. Attention must be given to linkages with the organization's goals.

2. Most studies examined a relatively small range of content and techniques of training, and used few well-researched measures.

3. There is a lack of comparative studies and research studies incorporating measures of individual differences, and organization climate factors.

4. Most researchers depended exclusively on the statistical significance as an indicator of judging success or failure of training. Few investigators attempted to say very much about the practical or theoretical significance of the magnitude of the changes they observed.

Clement (1981) used a similar format to Campbell's for his review. He pessimistically stated that management training evaluation practices had not improved much during the 1970s.

The relation between managerial training and the acquisition of managerial skills was much clearer when Burke and Day (1986) conducted the first quantitative literature review on management training studies by applying meta-analytical techniques to 70 studies that covered a 32-year period (1951-1982). They reported, overall, that "different methods of managerial training are on the average moderately

effective in improving learning and job performance" (p. 243). This is the first meta-analysis on management training programs that goes beyond other literature reviews "by quantitatively evaluating the degree to which the effectiveness of managerial training generalizes across settings for various training content areas, training methods, and outcome measures" (p. 243)

"Meta-analysis is the application of statistical procedures to collections of empirical findings from individual studies for the purpose of integrating, synthesizing, and making sense of them" (Wolf, 1986, p. 5). An individual study normally has unique study features, such as treatment, research design, and sample characteristics. The study result might be mediated by such study characteristics. The traditional qualitative literature review gives a general picture of the state of science at a given point in time. However, it is associated with the reviewer's bias of selection of studies, judgment and weighting of the findings of studies, and even misleading interpretations. Qualitative reviews, even those that employed descriptive statistics such as those by Campbell and Clement, could not accumulate information from significant tests, especially when the results were in conflict. By applying statistics to the data set of primary studies, the meta-analysis can accumulate research findings across studies, and examine characteristics of studies as potential explanations for disparate or consistent results. With statistical tools, meta-analysis can interpret information about the practical or theoretical significance from the magnitude of the changes given in each individual study.

It has been thirteen years since Burke and Day's (1986) first meta-analysis on management training evaluation studies was published. Many changes have been

happening in the training field since then. Both researchers and practitioners have been making various kinds of efforts to improve training effectiveness, such as training needs assessment, transfer of training, instructional psychology and learning process, implementation and maximizing technology in training, linking of training with performance improvement and organizational goals, and cost-benefit analysis of training programs. Many empirical evaluation studies have been conducted in the work setting to discover the effect of training on job performance and the organization results. However, the overall state of the evaluation on management training in the past 15 years is not clear; and many questions are remained to be answered through a quantitative review. For example, has any progress been made in management training evaluation research in either quantity or quality since 1983? What is this progress, if any? What are the findings of those research studies on training effectiveness?

Purpose of the Research

The purpose of this study is to apply meta-analysis procedures to the 1983-1997 collection of published (including dissertations) empirical evaluation studies with control to find out the magnitude of the effect of management training on trainees' learning, job performance and organization results. To achieve this purpose, three specific research objectives are established:

1. To summarize the characteristics and delineate a profile of empirical evaluation research studies on management training in the 15-year time period of 1983-1997.

2. To estimate the magnitude of the training effect of management training programs on trainees' learning, job performance and organizational results by calculating the respective average effect size of five measurement criteria.

3. To conduct moderator analysis on measurement criteria, training content, training method, training needs assessment and favorable condition of transfer of training, in order to explain the variability of the magnitude of training effect.

Definition of Terms

Management Training Intervention

It refers to formally organized training/learning activity for managerial personnel (executives, managers/administrators, supervisors) for the purpose of improving their managerial capacity.

Management Training Program Contents

This study adopted the classification of management training program contents that differentiates various training programs into ten major categories. They are: general management programs, human relations/leadership programs, self-awareness programs, problem-solving/decision-making programs, rater training programs, motivation/values training programs, technical skills, entrepreneurial skills, ethical decision making, and others.

Training Method

Based on information of training methods employed in primary studies, such as traditional (lecture, discussion), case study, behavior-modeling, computer-supported-training, the present study employed a thee category classification of training method: cognitive methods mainly, behavioral methods mainly, and combination of cognitive and behavioral methods.

Measurement Criteria of Training Outcomes

Reaction, learning, behavior and result are used as measurement criteria of training outcomes (Kirkpatrick, 1959, 1994) in most training evaluation studies. This study excluded reaction criterion since it is not the ultimate purpose of management training, and is not related to the other three outcomes (Dixon, 1990, Noe & Schmitt, 1986, Holton, 1996). By following Burke and Day's (1986) example of combining source of data (subjective vs. objective) with outcomes criteria (learning, on-the-job behavior and result), there are five measurement criteria of training outcomes classified in this study. They are subjective learning (SL), objective learning (OL), subjective behavior (SB), subjective result (SR), and objective result (OR). The behavior on the job is hardly measured by an objective standard. The hard data that measure the on-the-job behavior, such as accuracy of work, actually measure the result of the behavior rather than the behavior itself. So the criterion of objective behavior is omitted. Four of the five criteria are congruent with that used by Burke and Day (Burke & Day, 1986, p. 233, 237).

Subjective learning (SL). This criterion includes those measures completed by trainees or their trainer in the form of personal opinion and judgment on what (knowledge, skills, attitudes) and the extent that participants have learned during or by the end of training, e.g., an attitude survey completed by the trainee on leadership.

Objective learning (OL). This criterion includes those measures against objective means or standard on what knowledge, skills, attitudes and the extent that participants have learned during or by the end of training, e.g., a knowledge test; an expert's assessment on a video-taped interview done by the trainee.

Subjective behavior (SB). This criterion includes those measures on trainees' on-the-job behavior perceived or observed by themselves, peers, subordinates, or supervisors. The behavior being measured must be those which happened after training, and on-the-job situation.

Subjective result (SR). This criterion includes those measures of organization results perceived by respondents, not reported by organization hard record, e.g., trainee's subordinates' job satisfaction, or commitment to the organization, or group effectiveness perceived by subordinates.

Objective result (OR). This criterion includes those measures of organization results reported by organization hard record, e.g., turnover rate, revenue, productivity, and accuracy of work.

Training Needs Assessment

Training needs assessment is a systematic investigation or other efforts to identify training needs of participants at organizational, process/team, and individual levels. It

also includes explicit training objectives related to job performance improvement, or efforts to linking training with organizational strategy and goals.

Favorable Condition of Transfer of Training

It refers to favorable personal and environmental factors that promote training participants to apply their learning to job situations. These factors include transfer motivation, opportunity to use, peer support, supervisor support, and positive personal outcomes.

Meta-Analysis

Meta-analysis is analysis of analyses, i.e. the statistical analysis of the findings of many individual analyses (Glass et al. 1981, p. 12). It combines evidence across studies through a set of specially designed statistical procedures (Hedges & Olkin, 1985, p. 13).

Effect Size

Effect size is a common metric to represent the magnitude of the treatment effect. It measures the relationship between two variables that had been investigated. The computation of Effect Size is "the mean difference between experimental and control groups divided by within-group standard deviation" (Glass et al. 1981, p. 102).

Moderator Analysis

It is one of the major functions of meta-analysis. It determines the factors that would explain the variance in the magnitudes of the relationships between two variables. Such factors are called moderator variables as they moderate the magnitude of the treatment effect (Rosenthal, 1991).

Description of Variables

In primary studies included in the meta-analysis, the independent variable is the specific management training intervention implemented in that research. The dependent variable is the specific training outcomes measurement, such as reaction, learning, behavior, and result. When calculating the effect size of an individual study, the measurement of the dependent variable (training outcomes) is transferred into the common metric of effect size.

In this meta-analysis, the independent variable is the common research domain, i.e., management training to managerial personnel. The dependent variable is the magnitude of training effect expressed by the effect size. When conducting moderator analysis to identify moderator variables that explain the variance of the training effect among studies, the independent variables are the selected study characteristics such as measurement criteria, training content, training method, training needs assessment, and favorable condition of transfer of training. The dependent variable is the training effect in terms of the combined effect size.

Significance of the Study

This study is another meta-analysis on the effectiveness of management training conducted 13 years apart from the first of its kind by Burke and Day (1986). It covers empirical studies on management training evaluation which were published during a 15-year time period from 1983 to 1997.

This study gives new evidence of effect of management training on trainees' learning, job performance and organization results. It is an effort to partially answer the unsolved question of "which specific program or technique leads to greater or lesser changes for certain attitude or behavior measured by certain criterion", which was raised by Campbell et al. (1970) about three decades ago. In addition, the moderator analysis discovers the impact of training needs assessment, favorable condition of transfer of training, and combined cognitive and behavioral methods on the management training effectiveness. These findings reflect progresses on enhancing training effectiveness that have been made by the training field in the past two decades. They add more understanding and useful knowledge to the complex of management training effectiveness.

The present study tried to include important features and similar research questions of previous review papers (Burke & Day, 1986; Campbell, 1970; Clement, 1981), in order to compare the research findings. By using the quantitative approach of literature review, this study expands the current stock of meta-analysis on training effectiveness, especially for management training programs.

Limitations of the Study

The present study has several limitations. First, the literature search is limited to several major databases, such as ERIC, PsycINFO, Dissertation Abstract, and ABI. There must be management training effectiveness studies on other databases that were missed by this study. So the publication bias and incompleteness of the population is always a limitation for a meta-analysis.

Second, the data set of this meta-analysis is not as big as the previous one conducted by Burke and Day (1986). This mainly reflects the amount of the training effectiveness studies that meet the inclusion criteria set by the present study. However, there should be several more studies which could be included into the analysis if the time period for conducting this study were longer or the availability of certain studies and missing data in publication could be improved. Limited by the number of data points in a category, statistics procedures could not be applied to some subgroups. For those categories with small number of data points, the generalization of the finding was speculative.

Third, there were possibilities of capitalizing on chance when conducting statistical tests in an exploratory mode (Hunter et al., 1982, Hedges et al., 1989). This meta-analysis examined the influence on training effects of five study features of the primary studies. Since inferential statistics were applied to many variables of a data set, the interpretation of findings should be based on theoretical framework and should consider the practical significance as well.

In addition to these specific limitations, the common limitation from the implementation of a meta-analysis should be realized. The meta-analysis result is

influenced by the series of personal decisions made by the meta-analyst along with the analytical process. Although there are some common rules to follow, the inclusion criteria, the literature search process, the unit of analysis, and the information pulled out from each primary study are all subject to the individual reviewer. The various combinations of these personal decisions shape the analytical process which differentiates from others, and influences the result that it will reach. Though the researcher of the present study has followed some common rules in making decisions when it is possible, this study could not escape from this common limitation.

While Abrami et al. (1988) found notably different conclusions reached on a same research domain (the validity of student ratings in their case) by different meta-analysts, they believed that the great promise of meta-analysis was to attempt to make the process of reviewing as scientific as the conduct of the primary research. By improving the implementation of the process, meta-analysis should, and could be more precise, objective, and repeatable.

Organization of the Study

The first chapter is an introduction to this study. The background information provided context of the problem—effectiveness of the management training. The section of Description of Variables explains the independent variable and dependent variable in a primary study, the meta-analysis, and in the moderator analysis respectively, which are different but closely connected.

Chapter II is the review of relevant literature. Besides the common review on major concepts of management training, the sections of Training Needs Assessment,

and Favorable Conditions of Transfer of Training introduce two important study features that are employed as moderator variables in meta-analysis. The section of Introduction of Meta-Analysis is a brief overview of this quantitative analytical technique on primary studies, which lays a foundation for Chapter III.

The methodology chapter illustrates the specific meta-analytical procedures selected and employed by the present study. A series of decisions are made explicit and public. After enough methodological preparation, research questions and hypotheses are given in this chapter.

Chapter IV reports in detail the data analysis process and the research findings. The present study only reports the findings related to the established questions and hypotheses.

The last chapter summarizes major findings from the meta-analysis, then draws conclusions and discusses implications for both training practitioners and training researchers. Several suggestions for conducting meta-analysis are given at the end of the chapter.

CHAPTER II

REVIEW OF THE LITERATURE

This chapter reviews and synthesizes literature on management development, training in organization, and meta-analysis. The first four sections cover the management training, and discuss the definition, objectives, target audience, training contents and methods in order. The following two sections pay special attention to training needs assessment and favorable condition of transfer of training. The next two sections focus on training evaluation and the effectiveness of management training programs. The last section is an introduction to meta-analysis. Four meta-analytical studies on training effectiveness are reviewed and compared as examples to help understand the concepts and methodology.

Concept and Definition of Management Training

Management training is an area of practice that falls into both categories of management development and training in organizations. The field of industrial and organizational psychology treats management training as one of the major training contents, or one type of training programs of training in organizations (Campbell, 1971; Goldstein, 1980; Tannenbaum & Yukl, 1992). Management education, management

training, and on-the-job experience are defined in management literature as the three major components of management development (Wexley & Baldwin, 1986).

Training in Organizations

As a field of practice, training in work organizations has been known as "training in business and industry" (McGehee & Thayer, 1961), or "technical training" (Swanson & Torraco, 1995). In 1983, ASTD (McLagan) defined "training and development" as one of three major functions of the Human Resource Development (HRD)—Identifying, ensuring, and—through planned learning—helping develop the key competencies that enable individuals to perform current or future jobs.

Training is different from education. Swanson and Torraco (1995) cited from Dooley (1945): "Education is for rounding-out of the individual and the good of the society; it is general, provides background, increases understanding. Training is for the good of plant production—it is a way to solve production problems through people; it is specific and helps people to acquire skill through the use of what they learned" (p. 2). The history of training is as long as the history of the civilization of human society. Human beings began to create, and pass onto others, knowledge and skills even in the stone age (Swanson & Torroco, 1995, Miller, 1996). Harris et al. (1994) described briefly the history of training as one of the most important functions of the human resource development (HRD) field.

Goldstein (1980) defined training as "the acquisition of skills, concepts, or attitudes that results in improved performance in an on-the-job environment" (p. 230). Wexley's (1984) definition of training was "a planned effort by an organization to

facilitate the learning of job-related behavior on the part of its employees" (p. 519). The term "behavior" was used in the broad sense to include any knowledge and skills acquired by an employee through practice.

Training as a field of study was introduced early in the history of psychology and scientific management. According to Swanson and Torroco (1995), Taylor's principles of scientific management, which was published in 1912, discussed both selection of the best workers and extensive training. Munsterberg's focus on training in 1913 appeared in most early industrial and organizational psychology textbooks. McGehee and Thayer's (1961) classic text, *Training in Business and Industry*, was a systematic treatment of the major issues in training and development, as a topic in industrial and organizational psychology.

Management Development

Management development is a broad term describing any process by which managerial knowledge and skills are attained. In a review article, Wexley and Baldwin (1986) defined management development as "the whole, complex process by which individuals learn, grow, and improve their abilities to perform professional management tasks" (p. 277). Consistent with this definition, they viewed management education, management training, and both planned and unplanned on-the-job experiences as all being potentially important inputs for a manager's development.

As subsets of management development, management education refers to the acquisition of a broad range of managerial knowledge and general conceptual abilities in formal classroom settings in degree-granting institutions (Keys & Wolfe, 1988,

Rothwell, 1984, Wexley & Baldwin, 1986), while management training refers to the formally organized instruction—conducted internally or externally and specific to those already in the ranks of management. The focus is "on improving a narrow range of management-related knowledge, skills and attitudes" (Educational Research Encyclopedia, 1992, p. 763). These specific managerial skills, or self-awareness or motivation could be immediately applicable in a particular organizational setting (Wexley & Baldwin, 1986).

According to Harris et al. (1994), on-the-job experiences are planned or unplanned learning activities that a manager can experience through the daily work to gain new knowledge and information, and to develop and enhance skills and abilities. Coaching, mentoring and job rotation are several examples.

Importance of Management Development

The ability of the enterprise to survive and grow depends on the success of management development activities. According to Schmidt et al. (1971), financial analysts evaluate a company's overall program of management development as a basic factor for investment purposes.

Wexley and Baldwin (1986) pointed out that in today's competitive environment, the quality and performance of an organization's managers might determine its very survival. The idea that managerial knowledge, skills, and abilities can be learned and improved is widely and increasingly accepted. Many American organizations expend huge amounts of time, money, and energy to develop their managerial talent.

Management training is a coping strategy that helps managers do their jobs better. According to Curry (1992), management training aims to develop the individual by providing job-specific and state-of-the-art management knowledge and skills. By developing a working knowledge of the functional areas of the company and an understanding of the corporation as a whole, the trained managers are expected to become leaders of the company. Individual growth will enlarge the pool of promotable talent to guarantee continuity of leadership of the organization. Ultimately, management training is one of the means to improve corporate results. The increased demand for accountability, pervasive concern for cost constraint, and the pressures of globalization fueled the growth of management training and development.

Managerial Population and Target Audience of Management Training Programs

The managerial population covers a broad range of personnel in terms of level of authority and other characteristics. A common classification is to divide the whole managerial population into three levels: executive, middle management, and first-line management (Keys & Wolfe, 1988; Schmidt et al., 1971). Since different segments of the managerial population have different training needs, various management training programs with specific objectives were designed to fulfill special needs for different segments.

Considering the rank and authority of managerial personnel, management training and development programs are designed differently for executives, middle managers, and supervisors. *The Encyclopedia of Education* (1971) reviewed

management training programs from this perspective by discussing executive programs and middle management programs as major contents. The target audience of executive programs is those personnel from middle management up through the president or top executive of the organization. These programs provide guided growth and planned training activities to those who perform the executive function in an organization. The objectives of the executive training include: (a) provide leaders who are able to continue the growth of the organization; (b) encourage executives to grow as persons and increase their capacity to handle greater responsibility; (c) improve the multilevel performance of managers in the jobs they now perform; (d) help to sustain effective performance of executives throughout their career; and (e) provide a basis for measuring growth. Keys and Wolfe (1988) pointed out a major change in many firms' executive development programs from informal on-the-job training to formal training due to the growing complexity and globalization of the business world.

Different from the executive programs, middle management programs aimed at middle managers in an organization including personnel at all levels of authority between the vice-presidential level and the first, or foreman, level. Middle management personnel are concerned with intermediate and short-range organizational goals, with carrying out top management directives, and with the motivation and direction of the first line of management. They are concerned with individual functions. The objectives of the middle management programs include: (a) to make middle managers more effective on their present jobs; (b) to broaden their knowledge, skills, insights and attitudes to make them generalists from specialists; and (c) enlarge the pool of

promotable managerial talent to ensure succession of leadership. Keys and Wolfe (1988) discussed the middle management programs and the first-line management programs separately in differentiating their training needs.

Wexley and Baldwin (1986) discussed management training from a special target groups' perspective—women, small business owners/entrepreneurs, and international managers. In spite of the scientific fact that women are basically no different from men in managerial talent and potential, the percentage of female managers and administrators (no more than 5-10%) is much smaller than the percentage of female workers (over 43% of the American work force). To overcome both internal and external barriers for women to become managers, some special contents, such as career awareness education, identification and removal of stereotypical prejudgment and behaviors, as well as mentoring, coaching, and positive role modeling should be added to management training for women. Small business owners and entrepreneurs work in a complex and challenging environment, and face special problems. They have been neglected by training researchers and practitioners until several institutes designed and implemented special executive development programs for them. Intercultural awareness is a special problem faced by international managers. More attention is needed for appropriate content and methods of intercultural training to help managers identify their own cultural paradigms and understand both general and specific information about host countries.

The identification of training needs of all the segments of the managerial population is closely related to the recognition and definition of effective managers, and is strongly influenced by the competency-based movement in management education

popularized in 1980's, and the AACSB (American Assembly of Colleges and Schools of Business) outcome measurement project (Baldwin & Padgett, 1994; Keys & Wolfe, 1988; Wexley & Baldwin, 1986). Katz (1974) suggested that the special skills a manager should have depended on his or her administrative level. He claimed in his theory that a successful manager should have technical, human, and conceptual skills. The conceptual skills are most important to executives, while technical skills are most important at the lower level. However, human skills are needed at all managerial levels. According to Baldwin and Padgett (1994), Katz's theory was supported by some later studies in which the importance of various managerial activities was rated by managers at different levels. However, they also pointed out that much traditional research, such as taxonomies developed, and many programs implemented, were mainly focused on the common skill requirements of management jobs and did not specifically differentiate between managerial skill requirements according to hierarchical level. Blakely, Martinec and Lane (1994) conducted a study of 155 organizations. They found out that greater emphasis was placed on technical skills at lower management levels and on entrepreneurial skills at senior management levels, which supported Katz's theory. They also found that organizations with growth strategies focused on more management development areas than those with stability or retrenchment strategies. However, there were no differences in the focus on ethical decision making or technical-skills training among the three strategies.

Contents of Management Training Programs

The contents of management training cover a very broad range of topics. Those most frequently taught, and written about by researchers and practitioners include managerial motivation, leadership, decision making, and supervisory interpersonal skills. Wexley and Baldwin (1986) claimed that it was impossible for them to categorize the plethora of subjects falling under the broad umbrella of management training. The following section introduces three classifications of contents of management training.

Campbell, Dunnette, Lawler III, and Weick Jr. (1970) categorized management training programs into five major categories (p. 288):

1. General management and supervision programs, the broadest type of development effort including material on human relations, labor relations, labor economics, company policies and practices, supervision, leadership, decision making, et cetera;

2. General human relations programs, focused on human relations problems of supervision, attitudes toward employees, and communication problems;

3. Problem solving and decision making; emphasized teaching generalized problem-solving and decision-making skills;

4. Laboratory education, referred to a distinct training content as well as a unique method which some form of *T* (training) group is the prime ingredient. The other ingredients may consist of short lectures, group exercises designed to illustrate problems in interpersonal or inter-group behavior, role-playing sessions, and the like; and

5. Specialized programs; designed to achieve very specialized objectives.

Based on Campbell et al.'s categorization, Burke and Days (1986) expanded and redefined six categories of training content areas for their meta-analysis on effectiveness of managerial training. The six training content areas are (p.233):

General management programs. This is the broadest type of development effort and typically includes material on labor relations, management theory and practice, company policies and procedures, labor economics, and general management functions. The primary goal of these training programs is the teaching of facts, concepts, and skills.

Human relations/leadership programs. The content of these programs is narrower than that of the general management programs category in that the focus is on human relation problems of leadership, supervision, attitudes toward employees, and communication.

Self-awareness programs. The content of these programs is on understanding one's own behavior and how one's behavior is viewed by others, identifying the so-called games people play, and learning about one's strengths and weaknesses. Typical training methods are sensitivity training, laboratory training, T-groups, and transactional analysis.

Problem-solving/decision-making programs. The emphasis of these programs is on teaching generalized problem-solving and decision-making skills that would be applicable to a wide range of work problems that managers encounter.

Rater training programs. In these programs managers are trained to minimize errors when they are observing and evaluation their subordinates.

Motivation/values training programs. The content of these programs deals with increasing managers' motivation or modifying managers' values or attitudes.

In investigating the effects of management level and corporate strategy on the contents of management development programs, the questionnaire by Blakely, Martinec, and Lane (1994) contained nine training content areas. They expanded the areas from six to nine by adding the following three ones (p. 10):

Technical skills. These programs focus on job-specific technical skills.

Entrepreneurial skills. Training emphasizes risk taking and creativity.

Ethical decision making. Programs deal with the internal or external standards or codes of conduct used to govern the behavior of individuals or groups.

Methods of Management Training Programs

There are many ways to classify training methods according to various attributes such as the major purpose of the method, the training situation where the method is being used, the information technology imbedded in the method, et cetera.

A widely accepted classification of training methods for management training was set up by Campbell et al. (1970), which classified various training methods into two major categories: information presentation and simulation. Lectures, programmed instruction, and audiovisual means are techniques for information presentation, and case studies, role playing and business games are techniques used in the simulation methods. The information presentation techniques, that emphasize the teaching of facts, concepts, attitudes, or skills without requiring simulated or actual practice on the job, are mainly used for cognitive learning. According to Wexley and Baldwin (1986), behavior

modeling was becoming a popular technique in the second category as it was a research-based method for improving the human-relations skills of managers (p. 281).

Harris and DeSimone (1994) divided training methods into two types according to the training situation: on-the-job training and the classroom training. Each type consists of several techniques. The techniques used for on-the-job training included job instruction training, job rotation, coaching, and mentoring. Those used in classroom training included lecture, discussion/conference, audiovisual, experiential techniques (case study, business games, role-playing, behavioral modeling), and computer-based training (computer-aided instruction, intelligent computer-aided instruction). The advantage of this classification is that it is aligned with the real training situation, and distinguished the training situation and technology imbedded.

Along with the development of information technology and its application in teaching and learning, we can add a new training situation of the virtual classroom, and some new techniques in each situation. Table 2.1 lists training methods and techniques, which is expanded from the table of Harris and DeSimone (1994, p. 138).

In management training, several training techniques are unique, and/or are used widely. They include:

Behavioral modeling. It is one of the most popular methods for both interpersonal and technical skills training, which started in the 1960s and continued into the present.

Table 2.1

Training Methods and Techniques

Training Methods	Techniques
On-the-job Training	Job instruction training
	Job rotation
	Coaching
	Mentoring
Classroom Training	Lecture
	Discussion/Conference
	Audiovisual
	static media (e.g., books)
	dynamic media (e.g., film/video)
	telecommunication (teleconferencing, video-conferencing, multimedia on computer)
	Experiential techniques
	case study
	business games
	role-play
	behavioral modeling
	Computer-based training
	computer-aided instruction
	intelligent computer-aided instruction
Virtual Classroom Training	Internet-based-training (can include text, multimedia, discussion forum, group work)
	Email
	CD-ROM
	Internet
	Intranet

Source: Harris and DeSimone (1994, p. 138)

"The method involves learning some simple rules or key steps of the behavior and seeing a demonstration of the behavior to be learned (or the model), which trainees may then imitate or practice" (Russ-Eft, 1997, p. 105). Behavioral modeling is different from the lecture method, which tries to give people information and improve their knowledge with the expectation that they may change attitude and behavior after gaining the relevant knowledge. It is also different from experiential learning, which aims primarily at improving trainees' attitudes. Behavioral modeling attempts to change participants' behavior directly, and is a so-called "no-trial learning" (cited by Russ-Eft, 1997, from Bandura). Due to this unique feature, behavioral modeling has been widely used by business and industry in various programs and for different levels of managers. After reviewing extensively the research literature, Russ-Eft (1998) concluded that behavioral modeling is an effective training method in terms of reactions to training, participants' knowledge gains, changes in on-the-job performance, and bottom-line results or hard measures of organizational performance.

Leader Match. This special training technique based on Fiedler's contingency theory of leadership is used in leadership training. Its notion is that effective leadership needs a match between the leader's style and the situation he or she faces (Harris et al., 1994). The training method includes the identification of trainee's leadership style and a diagnosis of the situation, as well as the skills to modify the situation so that it becomes favorable to his style.

Sensitivity training. This is a distinct training method that is used in the self-awareness training. The classic model of sensitivity training is a group meeting without

an agenda in which participants discuss topics dealing with the "here and now" of the group process (Burke & Day, 1986).

Due to the multiple nature of the training skill/task of the management training, one single training technique is hardly adequate to attain the objectives of managerial training. In practice, most training programs employed more than one training technique. Burke and Day (1986, p. 233) listed "lecture/group discussion with role playing or practice" and "multiple techniques" (use of three or more training methods) paralleled with five other single techniques. Their meta-analysis found out that the "lecture with discussion and either role playing or practice" is more effective in general than other single techniques by showing a sizeable effect size (0.66 in subjective learning, 0.93 in objective learning, 0.34 in subjective behavior) and positive lower bound credibility values in all three applicable criteria of measurement. The "multiple technique (3 or more)" had effect size of 0.81 on objective learning, and 0.52 on objective result. The results also suggest that the effectiveness of the multiple technique with respect to objective result criterion generalizes across situations (p. 242).

The higher effectiveness of the multiple techniques could be partially explained by the better match of the teaching method and the teaching contents. Bennett (1995) concluded in his meta-analysis that "different training methods were found to be effective for different skills and tasks" (p. 125). Harrison (1992) found that the combination of cognitive and experiential methods was more effective than the cognitive only and the experiential only methods in cross-cultural training. He argued that the combination of methods increased trainee learning and behavioral outcomes

because it included all three sequential stages of skill development, and so was in conformity with the learning principles (cf. Gegné & Medsker, 1996).

It is interesting that two meta-analyses (Bennett, 1995, Burke et al., 1986) found that lecture is very effective in teaching both cognitive and interpersonal skill/tasks. The effectiveness of managerial behavior-modeling training technique with respect to subjective behavior criteria was high (effect size 0.78), and could be generalized across settings (Burke & Day, 1986). This finding is consistent with the aforementioned review resulting from Russ-Eft.

Training Needs Assessment

Overview

Most practitioners and academics agree that needs assessment, which normally serves as the first step to solve a performance problem or to design a training program, is important and critical. Ignoring the definitional nuances among different authors examined by Sleezer (1992), there are agreements on the conceptualization of the term needs assessment. It refers to the process to determine needs, including the steps and procedures of the process, such as identifying, prioritizing, selecting (for solution), and documenting/reporting, as well as the data gathering techniques employed to search for and integrate the information. Analyzing the causes of needs is normally a part of the process, especially when using a systems approach (Benjamin, 1989). Needs assessment is also problem-solving oriented. It "presents the opportunity to diagnose an organizational problem, and to prescribe a course of remediation" (Lewis & Bjorkquist, 1992, p. 35).

The literature suggests three stages of evolution for the needs assessment for training and several trends of development. From a very informal, intuitive approach to a systematic research approach, needs assessment has become an integrated component of performance improvement technology and human resource development.

Before 1961, training needs assessment was mainly done by an informal, intuitive approach. Moore and Dutton (1978) mentioned that in the 1950's only about one in ten companies reported systematic approaches to determine training needs. Management requests, observations and talks with supervisors were the most used techniques.

In 1961, McGehee and Thayer published their influential book *Training in Business And Industry*, and began a new , systematic research stage to determine training needs. They claimed:

> Training will not come of age until it abandons intuitive approaches to the solution of training problems ... Training, if it is to become an effective tool of management, must be a systematic, orderly procedure constructively applied to solutions of organizational problems and attainment of organizational goals ... Training, to be effective, must be backed up by careful and continuous research (p. 22).

Their integrated, three-level analysis designed to determine training needs—organization analysis, operation analysis and man analysis—laid a solid foundation for the theory of needs analysis. Since then, McGehee and Thayer's approach has been promoted by many writers and gradually adopted by practitioners (Goldstein, 1980, Latham, 1988, Moore & Dutton, 1978, Tannenbaum & Yukl, 1992, Wexley, 1984).

In 1983 the American Society for Training and Development (ASTD) published a study of the training and development field called *Models for Excellence*. "This study positioned training and development within the larger domain of human resources and provided a future-oriented description of the training and development field that could be used to select, manage, and develop training and development professionals" (McLagan & Suhadolnik, 1989). Needs assessment has been conducted to serve diverse purposes or focuses, such as providing information useful in making decisions about individual skill or developmental needs, organizational development needs, training program design, budget planning, or short or long range workforce staffing.

There are four trends in the development of needs assessment which have appeared in recent literature. First, the concept of need expands from present discrepancies to both the present and future-oriented discrepancies due to increasing competition, changing markets, and business itself. Second, there is a growing awareness of the importance of management support for training (Dodge, 1987, Sleezer, 1993). In practice, not only the management buy-in is critical to the success of needs assessment, and the following training intervention, but also the participants' buy-in. Third, the process of needs assessment is evolving from linear, concrete steps to a more dynamic, integrated process. This is evidenced by dozens of needs assessment models available in the training and development literature. Fourth, associated with the recent development of needs assessment expanding to performance technology and human resource development, there is a move from pure training needs assessment to more involved aspects of the organization, such as performance improvement, organization development, and career development (Phillips & Holton, 1995).

Three Levels Analysis

McGehee and Thayer (1961) first proposed and defined the three levels of analysis–organization analysis, operation analysis, and man analysis. Later Rummler and Brache (1990, 1995) expanded and described carefully this approach. It has been widely adopted and used by other researchers (e.g. Goldstein, 1991, Moore & Dutton, 1978, Sleezer, 1991, Swanson, 1994), and becomes a classic approach to conduct needs assessment for various purposes. The analyst should start the assessment and analyzing process from the organization level, emphasizing the organization's relationship with its market and the basic skeleton of the major functions that comprise the organization (Holton, 1995, Rummler & Brache, 1990, Swanson, 1994). The objective is to determine the organization condition, whether it meets the organization's established goals, and why. Organizational analysis provides information about where and when training is needed in an organization. For the process level, the analyst examines work, or tasks. They must go "beyond the cross functional boundaries that make up the organization chart to see the work flow–how the work gets done." (Rummler & Brache, p. 17). The objective is to determine the job performance condition, whether it meets the established standard of effectiveness and efficiency, and why. Process analysis identifies what should be the content of training in terms of knowledge, skill, and abilities employees must learn to perform the job effectively. Since any process or task is performed and managed by individuals doing various jobs, the objective of individual or person level analysis is to determine the individual expertise and job performance condition. It determines who needs training/learning on what. The three levels are connected and should be examined and understood with the system thinking.

According to Holton (1995), how to apply the three levels analysis depends on the situation and purpose of the needs assessment. "For some situations and/or purposes, it is appropriate for the needs analysis to address all three levels. However, for some other situations and/or purposes, it may only need two or even one level to be addressed" (p. 1).

In order to increase the effectiveness and efficiency of training, some other activities make a similar effort and achieve a similar result as needs assessment for training. One example is to explicitly state training objectives related to job performance improvement requirements. This is based on a good understanding of the process or work/task, as well as knowledge, attitude, skills of performance. Another example is to link HRD interventions, including training, with organizational strategy and goals (Tannenbaum & Yukl, 1992, Phillips & Rothwell, 1997). To achieve the linkage, the intervention must be designed and implemented in a way to address the organizational needs.

The Effect of Training Needs Assessment

Direct results of needs assessment were discussed and reported in the training literature. Holton (1995) summarized four possible results of needs assessments: information, priorities, management buy-in, and the recommended solutions and interventions. In the same casebook he edited together with Phillips, all but one of the 17 cases yielded results in all four categories.

However, there is little research reporting the effect of training needs assessment on training evaluation results. Bennett (1995) used training needs assessment as the

indicator for the quality of the implementation of a training program. He hypothesized that the extent to which a training intervention employed a systematic approach to needs assessment could influence the overall effectiveness of training. His meta-analysis on the effectiveness of training in organizations found out that implementation quality was a significant moderator of training effectiveness. Although only 7% of his total primary studies reported some need assessment activities, those training programs with need assessment were found to be markedly more effective than those without.

Transfer of Training

Concept and Definition

The ultimate purpose of training in organizations is to improve job performance and organization result. To achieve this, trainees must apply the knowledge and skills learned to their job situation. Tannenbaum and Yukl (1992) defined "transfer of training" as "the extent to which trainees effectively apply the knowledge, skills and attitudes gained in a training context back to the job" (p. 420).

Baldwin and Ford (1988) pointed out two distinctive concepts in understanding the transfer of training: generalization and maintenance. Generalization refers to the extent to which trained skills and behaviors are exhibited in the job situation. Maintenance refers to the length of time that trained skills and behaviors continue to be used on the job. The effectiveness of a training program is determined not only by the training program itself, but also by events that occur after a trainee returns to the job.

Supportive Organizational Climate

According to Rouiller and Goldstein (1993), a study conducted by Fleishman, Harris, and Burtt in 1955, the first of this kind, suggested that a supportive climate of the work environment is a factor in the transfer of learning to the job situation. Over the years, other studies (Baldwin & Ford, 1988, Rouiller & Goldstein, 1993, Bates, 1997) indicated that the supportive organizational climate influenced the transfer of training from the classroom to the job. They also identified major elements that make up the supportive organizational climate.

In a review article, Baldwin and Ford (1988) noted that supervisory support, such as reinforcement , modeling of trained behaviors, and goal-setting activities, could be important to affect the transfer process. Rouiller and Goldstein (1991) classified components of the organizational transfer climate into situational-cues and consequences that either inhibit or help to facilitate the transfer of what has been learned in training into the job situation. Situational-cues in the work environment included goal cues that remind trainees to use what they have learned, social cues that are influence exerted by supervisor, peers and/or subordinates behavior, and task and structural cues that come from the design and nature of the job itself. Consequences included positive and negative feedback and punishment. Rouiller and Goldstein (1993) hypothesized that the more positive the organizational transfer climate, the more likely it is that trainees will transfer key behaviors to the job that have previously been learned in training. In the empirical study they conducted with 102 assistant managers, the organizational transfer climate is significantly related to transfer behavior. So they

concluded that "a positive organizational transfer climate appears to be at least as important if transfer of training behavior is to occur" (p. 389).

Bates (1997) measured nine components of the organization transfer climate: (a) supervisor support refers to the extent to which supervisors reinforce and support use of learning on the job; (b) opportunity to use refers to the extent to which trainees are provided with or obtain resources and tasks on the job enabling them to use the skills taught in training; (c) transfer design refers to the extent to which training has been designed to give trainees the ability to transfer learning to job application and the training instructions match the job requirements; (d) peer support refers to the extent to which peers reinforce and support use of learning on the job; (e) change resistance refers to the extent to which the prevailing group norms are perceived by the trainee to resist or discourage using new skills; (f) supervisor sanctions refers to the responses made by supervisors which oppose or discourage the use of training on the job; (g) personal outcomes–positive, refers to the degree to which applying training on the job leads to outcomes that are positive payoffs for the individual; (h) personal outcomes–negative, refers to the degree to which applying training on the job leads to outcomes that are negative for the individual; (i) content validity refers to the extent to which the trainees judge the content of the training to accurately match the job (p. 105). In the same research, Bates defined performance utility as the measure of transfer motivation which influences the training outcomes directly. Among the nine above components, four (i.e., supervisor support, opportunity to use, peer support, and positive personal outcomes) were found to be positively correlated with performance utility, and three (i.e., negative personal outcomes, change resistance, and supervisor sanctions)

were found to be negatively correlated with performance utility. Correlation analysis of the association between predictor variables and performance ratings showed that only two environmental elements, peer support and supervisor sanctions, were significant.

The Effect of Transfer Climate

Several notable training transfer studies (Gist, et al., 1990, Rouillier & Goldstein, 1993, Tziner, et al., 1991, Wexley & Baldwin, 1986) provided empirical evidence to identify factors affecting transfer of training (Hilbert et al., 1997). However, the effect of transfer climate on training effectiveness is unknown.

One meta-analysis on training effectiveness (Bennett, 1995) hypothesized that favorable environment would mediate the training effect in a positive way. However, no empirical data available at that time to test it. In the above training transfer studies, the factors of transfer climate were independent variables, but not study features. This explained why there was no data for the moderator analysis of the training effectiveness meta-analytical study.

Evaluation of Training Programs

Considering the significant amount of money from the organizations, and the participants' time and learning efforts invested, it is not only reasonable and pragmatic but also ethical to develop effective training evaluation which provides some evidence of the effectiveness of training intervention on job performance and organization results (Curry, 1992, Hilbert, Preskill, & Russ-Eft, 1997).

Definition and Function of Training Evaluation

Goldstein (1980) defined training evaluation 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" (p 237). Wexley (1984) gave a similar definition of training program evaluation: "a set of procedures designed to systematically collect valid descriptive and judgmental information with regard to the ways in which a planned change effort has altered (or has failed to alter) organizational processes" (p.538).

All these definitions suggest two basic functions of training evaluation. The first basic function is to help decision-makers with the judgment of the value of the training program. As pointed out by Hilbert, Preskill, and Russ-Eft (1997), evaluation of training is thought to be the most appropriate method of demonstrating the value of Human Resource Development (HRD) interventions when HRD professionals are increasingly pressured to show how their efforts add value to the organization. Another basic function of training evaluation is to help trainers to improve the teaching and learning program. Information to help understand the training process, and outcomes of the program activity are all essential to achieve these basic functions.

Criteria of Training Evaluation

The training outcomes have a manifold nature, so it is important to establish multiple criteria that reflect the various instructional objectives and organizational goals. Campbell (1971) noted that the criteria chosen to evaluate training programs represented a value judgment which all concerned parties should agree upon before the

research began. Hilbert, et. al. (1997) called on "more collaborative approaches in conducting training program evaluations" since "the more stakeholders are involved in the evaluation's design and implementation, the greater the likelihood they will use evaluation results to make formative and summative program decisions" (p. 146).

Four criteria of evaluating training outcomes are most commonly accepted and widely used by training researchers and practitioners. Almost four decades ago, Kirkpatrick (1959) proposed four possible training outcomes: reactions, learning, behavior, and results. He originally conceptualized these four outcomes to be evaluated in a "four-step approach". The order of the evaluation steps reflects the time series of training outcomes that might occur, and is congruent with the degree of difficulty to evaluate them. So the four steps have also been called four stages, or four levels. Kirkpatrick (1994) finally called it the four-level model of evaluation. In his model, Level 1 of evaluation measures participants' reaction, i.e. to find out what the participants think and feel about the training, such as in what degree the participants enjoy the training, if the training environment is suitable and comfortable, and if the trainers are capable and credible. Level 2 of evaluation measures participants' learning, i.e. to find out how much the participants increased their knowledge, improved their skills or changed their attitudes as a result of the training. Level 3 of evaluation measures participants' behavior on the job, i.e. to find out whether and in what extent they used the knowledge and skills they have gained from training in their job roles. Level 4 of evaluation measures results, i.e. to find out the contribution of training to the business result and or/organizational goals.

According to Hilbert, Preskill, and Russ-Eft (1997), Kirkpatrick did not explicitly state the hierarchical nature of the four levels, but it has been understood and accepted that way. "Trainers have assumed that positive reactions (Level 1) are a prerequisite for learning (Level 2) to occur; behavior (Level 3) depends on learning (Level 2); and behavioral changes (Level 3) drive organizational results (Level 4)" (p. 112). However, the assumption has little research support. In recent years, some empirical studies have found some evidences against this assumption. Dixon (1990) concluded that higher trainees' reaction did not necessarily result in higher learning for them. Faerman and Ban (1993) conducted a study on a management training program designed to train leadership of first-line supervisors. They recommend measuring behavior changes as rigorously as possible, while not predicting it from participants' reaction to training, although positive reactions may increase the possibility of changing behavior on the job.

In spite of the many arguments about the relationship between reaction, learning, behavior, and results, as well as fierce criticism of the model (Holton, 1996), it has been widely accepted and used in practice and research that Kirkpatrick's four levels of evaluation are treated as types of training outcomes, or criteria of evaluation and categories of measurement. The series of review articles on training (Campbell, 1971, Goldstein, 1980, Tennenbaum, 1992), all treated Kirkpatrick's typology as the prevalent framework for categorizing training criteria. In a most recent training evaluation review article, Hilbert, Preskill, and Russ-Eft (1997) still used reaction, learning, behavior, and result—four possible training outcomes—to organize and report training evaluation studies.

Effectiveness of Management Training Programs

There is no simple conclusion on the effectiveness of management training programs. The knowledge on this issue is still insufficient to get an adequate understanding of the whole picture. This section summaries three important review articles on effectiveness of management training programs.

Campbell et al. (1970)

Campbell et al. examined research studies on management training evaluation published before 1970. The reviewers selected 73 studies that employed a control group, or used both before and after training measures when the control group was absent.

Target trainee. Studies included in this review focused on managers, supervisors, and administrators, but not on nurses, teachers, salesmen, and college students who have no managerial responsibilities.

Training contents. The reviewers used five major categories to classify contents of management training programs: general management or general supervision programs, general human relations programs, problem-solving and decision-making programs, laboratory education programs, and specialized programs.

Evaluation criteria. Campbell et al. used internal and external criteria to group evaluation studies. Studies with internal criteria attempted to measure and demonstrate some change in behavior relevant to the training itself, while studies with external criteria were directly concerned with changes in job behavior. Criteria such as general opinions regarding the training program, attitude measures, and tests of decision-

making ability belonged to the internal category. Objective measures of unit and manager performance in the job situation, turnover or grievances in a manager's unit, and ratings of job performance by superiors, peers, or subordinates were examples of external criteria. The problem associated with the internal criteria is whether the behavior changes observed in training have anything to do with management effectiveness. Campbell et al. pointed out that any expectation of a simple relationship between an attitude change and a change in performance is folly (p. 322).

Experimental control. The reviewers only selected studies that made a great effort toward experimental control to reduce the ambiguity in the link between the training activity and the criteria. If the study used experimental and control groups with either a before and after measure or just an after measure, it was identified as "some control". If the study employed only before and after measures but no control group, it was identified as "few controls".

Major findings. Neither too optimistic, nor too pessimistic, the reviewers were middle-of-the-roaders in regard to the impact of management development programs:

1. "Research on the effects of management development has demonstrated significant effects, and it has made a contribution to knowledge" (p.326). About 80 percent of the studies in the general management and general human relations categories showed significant results on most of the criteria used.

2. The most negative conclusion of this review was the almost exclusive reliance of management development research on internal criteria. Seventy-one percent of total studies reviewed (52 out of 73) used internal criteria. In general management

and general human relations programs, the proportion was even higher than that with 83 percent (29 out of 35).

3. The majority of the studies examined only a relatively small range of content and teaching techniques, and used only a few well-researched measures. Human relations courses were over-represented, and focused on a particular kind of attitudinal content. The most frequently used methods were combinations of lectures, conferences, role playing, and T groups. Relatively little research attention has been given to business games, the in-basket, or any of the on-the-job techniques.

4. There was a very small number of research studies on problem-solving and decision-making skills (one study used external criteria but few controls, three studies used internal criteria with some controls), and their results were generally negative. There was a lack of comparative studies and research incorporating measures of individual differences. A few studies had pointed to the importance of the managerial climate for the post training behavior of trainees, but no further study was made.

5. The problem under research is "much more complex than the efforts to attack it to date" (p. 326). There is basically no conclusion which can be drawn on which specific program or technique leads to greater or lesser changes for certain attitudes or behavior measured by a certain criterion. "Trying to assess the effects of management training on an organization by administering a narrow range of criteria before and after a relatively short-term teaching effort made up of a number of techniques which are not differentially understood can convey only so much information" (p. 326).

6. Most researchers depended exclusively on the statistical significance as an indicator of judging success or failure of training. "Few investigators attempted to say

very much about the practical or theoretical significance of the magnitude of the changes they observed" (p. 324).

Clement (1981)

Following the Campbell et al. review (1970), Clement (1981) looked at management training evaluations published in the1970's. Again, only meaningful data—26 studies that employed a control group or at least pre-training and post-training measures—were included. Clement used Campbell et al.'s review as a baseline to compare and summarize the progress during the 1970's.

Evaluation criteria. Clement used the same method of categorizing studies as internal criteria and external criteria. Studies used internal criteria to measure trainee reactions toward a course or the trainee learning that had occurred. These outcomes were usually measured within the course before the trainee had returned to the job. Studies employing external criteria measured results outside the course, such as improvements in the trainee's job performance and /or in the results for the organization. These outcomes tell more about the effectiveness of a training course than do reactions and learning. Major progress had been made in employing external criteria in the 1970's. About 58% of the post-1970 studies had focused on external outcomes, while only 29% of the pre-1970 studies did so.

Experimental control. The most disappointing finding is that the proportion of the 1970's evaluation studies using control group or both pre- and post-training measures (68%) is 19 percent less than the before-1970 studies (87%).

Other findings. Other findings included:

1. The effectiveness of various training methods was different in achieving different training goals. For example, the case method was better than other methods in developing problem-solving skills, while programmed instruction was considered better for retention of knowledge (Carroll, Paine, & Ivancevick, 1972, cited by Clement, 1981). According to Campbell et al. (1970), a study which compared the relative effects of two or more training methods against the same criteria was very valuable from both the scientific and the organizational point of view. However, only 4 of the 73 studies they reviewed (5%) did so. In the 1970's, more training researchers were concerned with the comparisons of relative effectiveness, and more studies compared the relative effectiveness of two or more training methods. The proportion increased from 5% to 19% (5 out of 26 studies), but it is "still too few to warrant making generalizations about which training method is best for a given objective" (p.10).

2. In spite of the strong argument that there is no "one best way" to manage, and no "one best way" to train—that the appropriate training method also depends upon the nature of the individual trainee—there were fewer research studies in the 1970's (8%) than before 1970 (11%) that examined the influence of individual differences.

3. Among the many variables that the appropriate training method depends upon, the organizational environment is an important one. Research to study the influence of the organizational environment on the outcomes of management training is very important (p. 11). However, very few studies, both before 1970 (7%) and in the 1970's (4%), focused on this issue.

In general, the reviewer was pessimistic about the management training evaluation. Not only did he claim that "evaluation practices have not improved much since 1970", but also he predicted that "evaluation may continue to play a lesser role in management training" in the 1980's (p. 12). He warned that the potential for growth of management training without accountability is just a proliferation of ineffective programs. "It simply makes no sense for American business to spend billions of dollars on training and development programs and almost nothing to determine their effectiveness" (p. 12).

Burke and Day (1986)

Burke and Day (1986) conducted the first meta-analysis to study the effectiveness of management training. They located 70 studies in the period of 1951-1982 according to the specific inclusion criteria: (a) involved managerial or supervisory personnel, (b) evaluated the effectiveness of training program(s), and (c) included at least one control or comparison group.

Training contents and methods. The researchers categorized studies into six content areas that were similar to those defined by Campbell et al. (1970): (a) general management programs, (b) human relations/leadership programs, (c) self-awareness programs, (d) problem-solving/decision-making programs, (e) rater training programs, and (f) motivation/values training programs. Training methods were classified as: (a) lecture, (b) lecture/group discussion, (c) leader match, (d) sensitivity training, (e) behavioral modeling, (f) lecture/group discussion with role playing or practice, and (g) multiple techniques.

Evaluation criteria. The authors of this study built the subjectivity-objectivity dimension of measurement into three of the four of Kirkpatrick's (1959, 1994) evaluation criteria: learning, behavior, and results. This procedure resulted in four evaluation criteria of this study: (a) subjective learning: what the participants learned about knowledge, attitude, and skills that are assessed by themselves or observed by others during or by the end of the training program; (b) objective learning: what participants learned about knowledge, attitude, and skills that are measured by objective means (e.g., standard tests) during or by the end of the training program; (c) subjective behavior: changes in on-the-job behavior perceived by trainees, peers, or supervisor; and (d) objective results: tangible results such as reduced costs, improved quality or quantity, promotions, and reduced number of errors.

Research findings. The meta-analysis resulted in 34 distributions of managerial training effects representing six training content areas, seven training methods, and four types of criteria. On average, the managerial training was moderately effective. Some specific findings included:

1. Regarding training content, all of the true mean effect size for the training content areas (human relations, general management, self-awareness, problem solving, motivation/values) were positive. On the basis of 21 studies measuring subjective learning criteria and 22 studies measuring objective learning criteria, the estimated true mean effect sizes were .34, and .38, with variances of .199 and .339 respectively. When measured against subjective behavior criteria, the estimated true mean effect size and variance were .49 and .344 respectively. Of all the 11 studies using objective results as criteria, an estimated true mean effect size was calculated as .67. In other word, the

mean of the experiment groups was more than one half of one standard deviation above the mean of the control groups. The estimated true effect size variances were also large (p. 323).

2. In four outcome criteria, the percentage of observed effect size variance accounted for artifactual effects (criterion unreliability and sampling error) over all studies is small to moderate (12.4%, 13.3%, 13.8%, and 22.0%). "Relatively larger amounts of unaccounted-for variance in these distributions may well be explained by other substantive variables (e.g., training method)" (p. 240).

3. Regarding specific training methods, there are several interesting and important findings. With the estimated true effect size 0.99, 39.3% of the observed variance accounted for artifactual effects, and a sizeable (.76) lower bound credibility value, behavior modeling was found to be a sound method for improving learning across situations as measured by subjective learning criteria. Evidence indicated that the method of lecture with discussion and either role playing or practice was very likely to generalize across situations using objective learning criteria. In subjective behavior criteria, the results confirmed that the training method (lecture, lecture with discussion, lecture with discussion and either role playing or practice, Leader Match, and behavioral modeling) can help explain the non-artifactual effects of variance. In addition to the result that all three lecture methods are likely to generalize across situations to some degree, the behavior modeling method is shown to generalize across settings. Results also suggest that the effectiveness of the Leader Match training method with respect to subjective behavior criteria generalizes across situations. In one

of the two distributions for objective results criteria, the multiple training techniques tended to generalize across settings.

Introduction to Meta-Analysis

Overview

Meta-analysis is quantitative synthesis and integration by applying statistical procedures to empirical findings of a set of individual studies which address the same research theme. According to Glass, the inventor of this research method, a meta-analysis seeks general conclusions across studies (Glass, McGaw, & Smith, 1981). Compared with a primary analysis which is the original process of data in a research study, meta-analysis is a secondary analysis of research data. It is a perspective that uses many techniques of measurement and statistical analysis to quantitatively summarize many individual studies (Glass, et al., 1981). The data set for meta-analysis is obtained from each individual study rather than from each individual subject as is done in the primary study. Compared to a traditional narrative literature review, meta-analysis is a quantitative review of literature (Wolf, 1986). The traditional narrative review is influenced by the subjective judgment, preferences, and biases of reviewers, and the differences of definitions, procedures, and samples of original researchers. Meta-analysis overcomes these limitations and weaknesses by collecting data from individual studies and synthesizing them more technically and statistically.

Though different statistical procedures applied, meta-analysis has two major functions. First, meta-analysis can combine research results of individual studies in the same research domain by using a common metric. It summarizes the empirical

relationship between two constructs. The two constructs, which form the common research theme of the meta-analysis, represent respectively the independent variable(s) and the dependent variable(s) in the individual studies. The accumulated, or combined result from a meta-analysis describes how the two constructs (e. g., management training intervention and trainee job performance) are related. Second, meta-analysis can examine the variability of research findings across studies which focused on the same relationship of interest. Is the variability a sampling error, or result of study features, or both? Meta-analytic researchers try to explain the variability by identifying mediate factors and possible interactions (Wolf, 1986, Hedges et al., 1989). These two functions are performed by two major meta-analytic approaches—combination of studies' results and comparison of studies' results.

In spite of much criticism, methods of meta-analysis have evolved dramatically and gained increasing popularity (Hedges, et al., 1985, Rosenthal, 1991, Hunt, 1997). The prolific studies on the same research domain in social science made it possible to conduct secondary analysis across many studies. The diversity of findings of individual studies provoke more meta-analysis to examine the variability.

Effect Size as a Common Metric

To accumulate and compare research findings from primary studies, either "effect size" or "significance level" of individual studies are analyzed. Glass and his colleagues' (1976, 1981) approach was to calculate effect size from primary studies and then combine them. Effect size is the standardized mean difference between experimental and control groups in the primary study, and tells the magnitude of the

observed effects of the treatment. The larger the standardized difference between the means, the greater the magnitude of the effect. Effect size is expressed on a standard normal metric *d*– the mean difference between experimental and control groups divided by within-group standard deviation. According to Mullen and Rosenthal (1985), another common index of effect size is the Pearson product moment correlation coefficient (*r*), and the associated proportion of variance accounted for (r^2). The larger the *r*, and the associated r^2, the greater the magnitude of the effect. Rosenthal (1976), independently from Glass, published the method of combining significance levels. Rothenthal's procedure allows the researcher to estimate the probability that the *p* levels of included studies "might have been obtained if the null hypothesis were true" (Mullen & Rothenthal, 1985, p. 9). Over the years, many researchers expanded and developed various meta-analysis procedures from Glass and Rothenthal's early contributions (e.g., Hunter, et al., 1982, Hunter & Schmidt, 1990, Hedges & Olkin, 1985).

Mullen and Rosenthal (1985) summarized various statistical procedures and techniques, and gave a simple but insightful frame of understanding them—"Generally, there are two study outcomes that can be analyzed (significance level and effect sizes), and there are two major analytic approaches (comparison of studies' results and combination of studies' results)" (p.3).

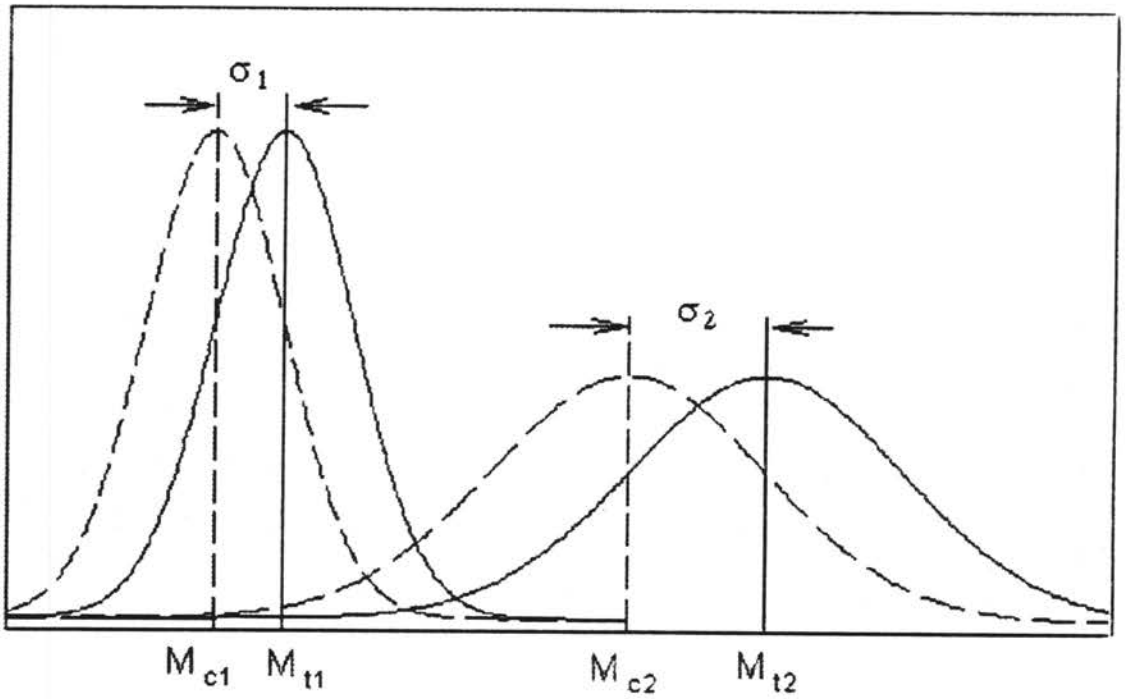
The present study focuses on the magnitude of the effect of management training programs that is observed by the difference of outcome measurements among treatment and control groups. So the standardized mean difference *d* is used to be the index of the effect size. Formula (2-1) and Figure 1 illustrate the concept and meaning of the effect size.

$$ES = \frac{M_t - M_c}{\sigma} = \frac{\overline{X_t} - \overline{X_c}}{S}$$

(2-1)

Where ES stands for effect size, M_t and M_c represent the mean of outcome measurement of the experimental and control group respectively, and σ is the standard deviation of the distribution. Since true parameters of the population are normally unknown, they are estimated by corrispanding sample statistics: $\overline{X_t}, \overline{X_c}$, and S.

The mean difference between treatment and control groups (numerator in the Formula) tells the magnitude of the treatment effect of a study. However, they are not comparable among studies (see Figure 1), since the standard deviation of each study is not equal. Therefore, the ES is then standardized by dividing the mean difference with its standard deviation, and becomes comparable and combinable among studies. As illustrated in Figure 1, if the effect size of a training intervention *d* equals 1, i.e., the mean group difference is one standard deviation, it means that the learning result of a 50^{th} percentile student in the treatment group is equivalent to the learning result of a 84^{th} percentile student in the control group. In education, this is quite a remarkable achievement.



Source: Adapted and expanded from Glass et al. (1981, Figure 2.2, p. 29)

Figure 1. Illustration of the Effect Size as a Common Metric.

Procedures and Special Considerations

To ensure the rigor of the meta-analysis as the same as that applied to the original studies, there is a series of procedures of collecting, processing, and reporting data. The procedures normally include formulation of research problem, collection of data, analysis and interpretation of data, and report of data. In the research problem formulation stage, special considerations and decisions of a meta-analysis include (a) select the common research domain or theme, (b) define major constructs and operations of the research theme, (c) select major study features based on theoretical and conceptual framework. Several special issues in the data collection stage are

(a) specifying inclusion criteria based on the problem formulation, (b) locating studies, (c) coding study features, (d) calculation individual outcomes. In the data analysis stage, the analyst needs to select specific statistical technique to combine and compare research findings across studies.

Several important considerations are discussed in the following part using four meta-analytical studies on training effectiveness as examples. The four meta-analytical studies are: Burke and Day's (1986) meta-analysis on management training effect, Chen's (1994) meta-analysis of 25 studies on cross-cultural training effectiveness, Bennett's (1995) accumulation of 177 training evaluation studies from 1960 to 1993. Lai's (1996) integration of 18 studies on leadership training programs.

Research domain and constructs. In meta-analysis of training effectiveness, the common research domain is the examination of the relationship between training intervention and its outcomes. Depending on the focus of a particular meta-analysis, the constructs could be broad or relatively narrow. Bennett (1995) treated training effectiveness in general, so in his study the construct of training intervention was very broad to include almost any training activities in work organizations. Burke and Day (1986) aimed at the effectiveness of management training, so the construct of training intervention was narrowed to include only management training programs. Chen (1994) and Lai (1996) were more focused on one type of management training activity, i.e. cross-cultural training and leadership training respectively. The degree of the broadness of construct was totally dependent on the researcher's interest–at what level he or she would like to conduct the study. The formulation of the research domain will influence the inclusion criteria of the meta-analysis.

Inclusion criteria. According to Abrami et al. (1988), "inclusion criteria is the description or operational definition of your research population" (p. 156). Even in the same research domain, there will be significant differences in effect size associated with inclusion criteria. The first type of inclusion criteria is based on the operational definition of the research domain and constructs. The second type of criteria constitutes boundaries, such as publication year, organization setting, program type and trainee classification. It makes the research both meaningful and manageable. Another type of inclusion criteria reflects technical requirements of meta-analysis, such as control group, pre- and post-test, empirical data to calculate effect size.

Inclusion criteria reflect the researcher's standards, some are looser than others (e.g., Bennett, 1995), to select studies. Normally, the meta-analysts state explicitly their inclusion criteria. Table 2.2 includes the inclusion criteria of the four meta-analytical studies under discussion.

Criteria of training outcomes. Training outcome is the dependent variable of training effectiveness studies, and so the meta-analysis. In more focused meta-analysis, such as that of Chen (1994) and Lai (1996), the dependent variables are very specifically defined. They are close to the original definitions in primary studies. When the constructs are broader, such as in the cases of Burke and Day (1986) and Bennett (1994), the dependent variables of training outcomes are defined in general categories as reaction, learning, behavior and results. Each category can contain many specific defined dependent variables of primary studies. It is very unique that Burke and Day built the subjectivity-objectivity dimension of measurement into the outcome

dimension, and got four dependent variables of their meta-analysis: subjective learning, objective learning, subjective behavior, and objective results.

Unit of analysis. In a meta-analysis, the researcher's decision on the unit of analysis will influence the data analysis procedure, and should be explained clearly. The most logical selection of the unit of analysis seems to be individual study, so a mean effect size could be calculated for each primary study. However, only Chen (1994) can do this because her several dependent variables were so similar in nature that a composite effect size is meaningful. Lai (1996) calculated an average effect size for each measurement criterion in a study. Bennett (1995) used the independent data point, and took out outlier data points to ensure the quality of the data set. He aggregated data collected from the same group of subjects to find the average, and got 474 independent data points from an initial data set of 1219 data points. Burke and Day (1986) calculated one effect size for each outcome variable to enter the overall analysis.

Independent variables for moderator analysis. All meta-analysts had their target factors based on a theoretical framework and literature review when they tried to explain the variability of the mean effect size. This is a confirmatory approach rather than an exploratory approach. Burke and Day (1986) did moderator analysis on outcomes measurement criteria, training content areas and training methods. They suggested that future researchers assess the influence of the trainer's experience and qualifications on the effectiveness of training. Bennett (1995) took implementation quality (training needs assessment), training methods, skill/task characteristics, trainee characteristics, study design/methodology rigor, and evaluation criteria as major factors that influence the effectiveness of training. He expected environmental favorability as a

moderator factor, but, unfortunately, there was no data available at that time. Chen and Lai conducted multi-regression to find out the contributions of their selected moderator factors.

Meta-analysis procedure. All four studies used effect size as the common metric to present the magnitude of effect of training. However, different meta-analysis procedures were employed in each study. Burke and Day (1986) estimated true mean effect size and variance for criteria groups and for training content/method subgroups. Using formulas provided by Hunter et al. (1982), the effect size was corrected for criterion unreliability, and weighted by sample size. In order to test significance, they established lower bound credibility value for each estimated population effect size. Chen (1994) tested the significance of the combined effect size, and conducted a multiple regression to investigate the relationship between study characteristics and the effect size. In Bennett's (1995) study, mean sample-weighted effect size (ES) for each level of the moderator variables was calculated. Subset meta-analysis was conducted to explain variance. A moderator was identified if the effect size variance was lower in the subsets than for the factor as a whole, and/or if the average effect size varied from subset to subset. Lai (1996) employed Hedges et al. (1989) approach. Calculation for a combined effect size was performed only to a homogeneous group, which passed the homogeneity test. Table 2.2 is a summary and comparison of these four meta-analyses on training effectiveness.

Table 2.2

Comparison of Four Meta-Analyses on Training Effectiveness

Author	Burke & Day	Chen	Bennett	Lai
Publication	Journal article, 1986	Masters thesis, 1994	Ph.D. dissertation, 1995	Ph.D. dissertation, 1996
Data	70 studies 1951-1982	25 studies	177 studies 1960-1993	18 studies 1965-1994
Research Domain	effectiveness of managerial training	effectiveness of cross-cultural training	effectiveness of training in organziations	effectiveness of leadership training progams
Inclusion Criteria	published and unpublished; involved managerial or supervisory personnel; evaluated the effectiveness of one or more training programs; include at least one control or comparison group	quantifiable data regarding cross-cultural training effectiveness; with control group, without control group but reported pre- and post training effectiveness data (using pre-training data as the control group data)	investigate effectiveness of training (rater training excluded); report sample size and statistics to calculate ES.	evaluation research of leadership training for education administrators; experimental or quasi-experiment with control group design; with measures to calculate effect size; year 65-94; in USA.
Evaluation Criteria (dependent variable)	subjective learning, objective learning, subjective behavior, objective results	personal adjustment (attitudinal and behavior) job performance of trainees composite (multiple dependent variables from one study, and compute a composite ES)	reaction learning behavior result	perceptions of change in leadership skills; perceived changes in social/org climates; changes in followers' performance; perceptions of the leader's roles
Independent Variables (for moderator analysis)	training content areas (one variable with 6 categories), training methods (one variable with 7 categories)	methodological char. (3 variables: Statistical equation of group; use of control group, # of trainees) training method (one variable: analytical, experiential, integrated) training char. (3 variables: Training time length, training time range, training situation) measurement char. (2 variables: time of outcome measurement, feature of outcome measurement) type of dependent variable publication status	implementation quality (TNA) training methods skill/task characteristics trainee characteristics study design/methodological rigor evaluation criteria environmental favorability (no data)	Study design (experiment or not, sampling method, subjects assigning, measurement design, instrument type) Participants (work unit, sample size of exp/contr group in posttest, # of subjects) Training char. (Training focus, source, program scale, planning agency, credit program? NA? Trainer type, time frame, duration of training, type of training method) Outcome measurement (posttest timing, evaluation level, instrument type, source of effect size, value of effect size

Table 2.2 (Continued)

Author	Burke & Day	Chen	Bennett	Lai
Publication	Journal article, 1986	Masters thesis, 1994	Ph.D. dissertation, 1995	Ph.D. dissertation, 1996
Data	70 studies 1951-1982	25 studies	177 studies 1960-1993	18 studies 1965-1994
Research Questions	1. Across studies with respect to various criteria, how effective is managerial training? 2. For each type of criterion measure, what is the relative effectiveness of different types of managerial training? 3. For each type of criterion measure, what is the relative effectiveness of different managerial training methods and combinations of methods?	N/A	1 What is the overall effectiveness of training? 2 examine factors that influencing training effectiveness (TNA, measurement criterion, training methods, skill/task char. Trainee char. Research methodological rigor)	1. What are the study features of evaluation research on leadership training effectiveness using experimental or quasi-experimental design? 2. What is the overall effect size of leadership training described in the studies synthesized? 3. To what extent do selected study features contribute to the variability of leadership training effects?
Unit of Analysis	an effect size was calculated for each outcome variable (one study can have several ESs)	one study: calculate a composite ES for one study (Rosenthal & Robin, 1986).	independent data points (474 data points, or effect size from 177 studies) when data points are non-independent, get the average; identify and reject outlier.	the average effect size of each measurement criterion in a study, overall effect size Hedge 1989
Meta-analysis Procedure	calculate effect size and variance of primary studies (formulas used: Glass et al. 1981, Hunter, et al. 1982, Schmidt, Hunter, & Pearlman, 1982); estimate true mean effect size and variance for criteria groups and content/method subgroups (corrected for criterion unreliability, and weighted by sample size, Hunter, et al. 1982 formula); establish lower bound of credibility value to test significance;	descriptive statistics for categorical variables (nominal scale); convert r, t, F into ES using Glass's formula (1980); test significance of mean effect size; conduct multiple regression to identify the relationship between study characteristics and ES;	accumulating ES across studies: mean sample-weighted ES for each level of the independent variables was calculated; subset meta-analysis to explain variance (a moderator is identified if: the effect size variance is lower in the subsets than for the factor as a whole; and/or the average effect size varies from subset to subset);	descriptive counting and percentage for N=18 heterogeneous effect sizes; homogeneity test; for the N=12 homogeneous effect sizes: descriptive analysis; calculation of overall ES; multiple regression for relationship between study characteristics and ES.
Special Issues	collect and code sample size and criterion reliability to correct reported results for artifacts such as sampling error and attenuation.		identify outlier; grouping data points by category	homogeneity test, combine effect sizes only for a group that passes the homogeneity test

Summary

This chapter reviewed literature relevant to the present study. The first four sections dealt with management training. The topics reviewed were typical and the contents were relatively stable for years. There are a few new content areas added to the list of management training programs. Training methods were enriched by the development of technology. There was a trend of using multiple training techniques, or combination of cognitive and behavioral methods in the management training.

The next part of the review synthesized research literature on training needs assessment and transfer of training. These are important progress happened in the training field in the past two decades. Without a systematic needs assessment to link training to organizational problems at various levels, training would never play a strategic role in the organization. Two training outcomes, trainees' job performance and organization results occur after the training program is finished. They depend on not only the quality of the training program itself, but also on the extent in which trainees apply their learning to the job situation. Therefore, the training effects on trainees' job performance and organization results are dependent variables of the transfer of training. The importance of training needs assessment has been well stated in the literature, and the major factors influencing the transfer of training have been studied and identified by several notable studies. However, the magnitude of effect of these two factors on trainees' learning, job performance and organization results remains unknown.

The last part of the review was an introduction to the methodology of meta-analysis. Four real cases were used to illustrate how the methodology was applied to various situations of training effectiveness research.

CHAPTER III

METHODOLOGY

This study applied meta-analysis procedures to the 1983-1997 collection of published (including dissertation) experimental evaluation studies with control group to find out the magnitude of the effect of management training on trainees' learning, job performance and organization results. It also explained the variability of effect size through moderator analysis.

The research methodology employed served the purpose and objectives of this study. Unlike narrative literary reviewers who normally use private subjective rules in the integration of a body of research, meta-analysts are forced to make their subjective rules public and explicit (Mullen & Rosenthal, 1985). This chapter discusses each major step in the meta-analysis process of this study, and explains why the author makes these choices and decisions.

Formulation of Research Problem

Identification of Research Theme

The research theme of this study was the effect of management training on trainees' learning, job performance and organization results. In order to conduct a meta-analysis, the target primary studies had to have a common focus on this research theme

(or domain). They were studies that investigated and reported outcomes of management training programs on the learning, behavior and/or result levels.

There were two constructs in this research theme, training intervention and training outcomes. Training intervention in this study was defined as management training. If a training program's target audience was managerial personnel (executives, managers/administrators, supervisors), and the major training objective was to improve their management capability, the study was within the range of the construct of "management training". Another construct of this study was training outcome, which was broken down into four measurable operations: reaction, learning, behavior and result. This study paid attention to outcomes beyond the reaction level. To be included, a study had to investigate and report changes in trainees' learning (knowledge, skills, attitude) at the end of the training program, behavior on their job role (changes of trainee job performance), and/or organization results, which could be a contribution to bottom-line, increase of productivity, and decrease of cost.

Description of Variables

In a primary study included in the meta-analysis, the independent variable was the specific management training intervention that was implemented in that research. The dependent variable was the specific training outcomes measurement, such as reaction, learning, behavior, and result. When calculating the effect size of an individual study, the measurement of the dependent variable (training outcomes) was transferred into the common metric of effect size.

In the meta-analysis, the independent variable was the common research domain –management training to managerial personnel in this case. And the dependent variable was the magnitude of training effect expressed by the effect size. The moderator analysis investigated the relationship between study characteristics and study outcomes in order to explain the variance of the training effect across studies. So the independent variables were the selected study characteristics, such as measurement criteria, training content, training method, training needs assessment, and favorable condition of transfer of training. And the dependent variable was the training effect in term of effect size.

Research Questions and Hypotheses

To achieve the objectives of finding out the magnitude of the effect of management training and identifying moderator variables, three research questions and two generally expressed null hypotheses were proposed:

Research Question 1: What are the major characteristics of empirical evaluation research studies on management training in the 15-year time period of 1983-1997?

Research Question 2: What is the magnitude of training effect measured by each of the five criteria (SL, OL, SB, SR, OR) of management training programs? Under each criterion, what is the magnitude of training effect of management training programs that teach different contents or employ different methods?

Since different criteria measured different training outcomes, the researcher of the present study expected that the whole data set would be heterogeneous, so that a meaningful overall effect size (the grand d_{++}) of management training would not be available. However, through grouping and sub-grouping the data set by moderators,

several groups and/or subgroups would reach homogeneity and produced an average effect size (d_+) for its own category.

After a meaningful d_+ was calculated for a specific group or subgroup, a null hypothesis was designed to test its significance: Is the effect size a true difference between the experimental group and the control group, or just by chance? In other words, is the effect size caused by the treatment—management training, or just the sampling error? The general expression of the null hypothesis to test the significance of an average effect size d_+ is:

H_{01}: There is no difference in the training effect measured by X criterion of managerial personnel who receive management training that teaches Y content (or that is taught by Z method) and those who do not receive the management training that teaches Y content (or is taught by Z method).

Research Question 3: Are the selected study characteristics--measurement criteria, training content, training method, training needs assessment, and favorable condition of transfer of training—moderator variables that mediate the magnitude of training effect of management training programs?

The general expression of the null hypothesis to test the significance of the difference between averaged effect sizes d_{+i} is:

H_{02}: There is no difference in the magnitude of training effect of management training programs that are measured by different outcomes criteria (or that teach different contents, or that are taught by different methods).

Procuring Studies

Individual studies included in a meta-analysis are similar to research subjects in a primary study. The study allocation process is comparable with the population identification and sampling process. There are two major decisions and steps involved in procuring studies for a meta-analysis. The first is to decide the inclusion criteria which draw the boundaries and standards for the literature search. The second is to plan and conduct the literature search to identify the population according to the inclusion criteria. The entire primary studies which the meta-analysis is based on could be either the population or a random sample of it. Since the size of the targeted population of this study was moderate, the whole population identified through the literature search was included into the meta-analysis.

Inclusion Criteria

Several special considerations were used to form the inclusion criteria of this study. The first consideration was the connection with previous meta-analytical studies on management training. Only one study was found that of Burke and Day (1986), which included management training evaluation studies from 1951-1982. Therefore the starting year of this study was set at 1983.

The second consideration was the scope of the training program that was evaluated by the included research studies. There were diverse viewpoints, such as training contents, trainee rank, and work-settings to classify training programs. In previous review articles on management training, Campbell et al. (1970) focused on managerial, supervisory, and administrative roles, and did not include studies

examining the training effect on nurses, teachers, salesmen, or college students. Clement (1981) selected 26 management training studies for review from the same scope as Campbell's. The first inclusion criterion of Burke and Day's (1986) meta-analysis was "the study involved managerial or supervisory personnel" (p. 233). The present study was consistent with these authors. So those studies with nurse, sales person, military instructor and college student as trainee were excluded. However, the present study did not limit the type of organization to business and industry. It included also studies that involved formal managerial training/learning activities happened in military, government, hospital, school and university.

The third consideration was related to the research design and control. Results of training effect studies were meaningless if the researcher did not have some kind of control on the variables (Campbell, 1971, Clement, 1981, Burke & Day, 1986). One of Burke and Day's inclusion criteria was that "it included at least one control or comparison group" (p. 233). The author of this study decided to include only true experimental and quasi-experimental studies with a control group. The one-group pretest-posttest design and two static-group comparison (no random assignment of treatment to groups, and no pre-test), which belonged to pre-experimental design (Gay, 1996), were excluded.

According to Gay (1996), experimental research is the only type of research that can truly establish cause-effect relationships, because of the direct manipulation and control of variables. The true experimental designs control for nearly all sources of internal and external invalidity. Two important and common characteristics of all experimental designs are random assignment of subjects to groups, and the presence of

a control group. However, sometimes it is just not possible to randomly assign subjects to groups. In many cases, the best effort that a researcher can make in a work setting is to conduct a quasi-experiment on managerial personnel. One of the most frequently used quasi-experimental designs is the nonequivalent control group design (Gay, 1996). Although it does not involve random assignment of subjects to groups, it should randomly assign treatment to groups. In addition, two groups are both pre-tested in order to control as much as possible the pre-existing differences.

In recent years, an increasing trend in training research is to conduct studies in a work setting rather than in a laboratory or college classroom (Russ-Eft, 1997). So it is not surprising to see many quasi-experimental designs in the primary studies.

Based on the above considerations, the inclusion criteria of the present study are empirical evaluation studies which: (1) examine effects of training programs for managerial personnel (including supervisor, manager, executive, administrator), (2) report training outcomes at the learning, behavior and/or organization result levels, (3) use true experimental or quasi-experimental design, (4) report experimental and control group sample size, (5) report statistical data, such as group mean and standard deviation, or other inference statistics of r, t, F, to allow calculation of effect size, and (6) were published between 1983-1997.

Literature Search

According to Mullen and Rosenthal (1985), "a meta-analysis could be useless and even counterproductive if the meta-analysis is conducted upon an incomplete and

biased data base" (p. 16). The source of studies includes: (a) articles published in professional journals, (b) books/chapters in books, (c) theses and dissertations, (d) papers presented at conferences, (e) archives, (f) technical reports, and (g) unpublished sources (unpublished manuscripts, manuscripts under editorial review).

There is no doubt that utilizing the full range of sources of studies will produce a more comprehensive, representative, and more accurate meta-analysis. However, in practice, the researcher is always subject to certain constraints and limitations that make perfect completeness impossible.

Publication bias has been discussed by many meta-analysis books (Glass, et al., 1981; Hedges, et al., 1989; Hunt, 1997; Mullen & Rosenthal, 1985; Rosenthal, 1991), which stated concerns that publication policies may be biased in favor of reporting significant findings. Therefore, strict reliance on published sources could produce inflated estimates of the overall levels of significance and/or overall effect size. Similar concern is discussed under the "File Drawer" problem which refers to the "possibility that unknown, unpublished studies might exist, whose results fail to support the pattern established by published findings" (Mullen & Rosenthal, 1985, p.17). However, the unpublished study is just an unknown variable in most cases since there is no way of knowing the whole picture. Campbell et al. (1971) only included published studies in their base-line review on management training, and hoped that "if someone carries out a well-designed study, he will publish the results, regardless of the outcome" (p. 321). Burke and Day (1986) reported that their studies "were located by conducting computer searches of the ERIC and PsychINFO indexes and by scanning bibliographies of the published and unpublished sources" (p. 233). However, they did not report the

percentage of the unpublished studies.

Many organizations, especially large-scale ones which have conducted internal training evaluations operated by either internal staff or outside consultants, have not published the results. However, there is no good system to track organizations' internal training evaluation studies. One cannot know how much has been done and how well done the several available unpublished studies are. Furthermore, among unpublished studies, many were conducted for purposes other than research. The inclusion of unpublished studies thus may introduce other unexpected bias. So this researcher decided to include only theses and dissertations as the so-called "unpublished" studies because (a) a thorough search on this sub-population can be conducted, and (b) most of these studies are carefully designed and conducted research work. Actually, theses and dissertations are published studies as they are on the *Dissertation Abstract* database and are accessible to everybody.

The literature search process of the present study started with the electronic databases of Educational Research Information Center [ERIC], PsycINFO, Dissertation Abstracts, and [ABI]. During the research process of this study, the OSU Library updated gradually its information databases from CD-ROM to the Internet, which is more complete and most up-dated. Accordingly, this researcher transferred the devices of the literature search of this study.

The key words used for search included: management training, management development, effectiveness, evaluation, experimental, and control group. Abstracts of the search results were printed out and read to determine whether they were worthwhile for further examination.

References of major qualitative review articles on management training and training evaluation (e.g., Hilbert, et al., 1997, Latham, 1988, Tannenhaum & Yukl, 1992, Russ-Eft, 1998, Wexley, 1984) were another source to find management training evaluation studies. A manual search of this source gave several new and repeated results.

A unique source for the present study was the three meta-analytical studies on training effectiveness (Bennett, 1995, Chen, 1994, Lai, 1996). The inclusion criteria of these three meta-analyses were either broader or narrower, but all had more or less overlaps with the present study in training contents, trainee, or time period of publication. Chen (1994) and Lai (1996) meta-analyzed the effectiveness of cross-culture and leadership training, which overlapped with managerial training in contents. Bennett (1995) accumulated effects of training in work organizations, which included management training for managerial personnel. A manual search for the list of their primary studies generated extra studies. Seven journal articles on training effectiveness which were included in Bennett's meta-analysis met inclusion criteria of the present study. One of the seven articles was also on the list of Chen. Four dissertations which studied leadership training effect on administrators in education from the study population of Lai (1996) were included in this study.

The Social Science Citations Index [SSCI] is used moderately through the OSU librarian (the university needs to pay on-line time for searching) to trace publications on two major review articles on management training (Campbell, et al., 1970, Burke & Day, 1986).

Forty-seven (47) studies were obtained from the OSU Library, or through the interlibrary loan service. After reading carefully, the author dropped 11 studies that did not really meet the inclusion criteria. Efforts were made to acquire extra data from the author of the primary study if it met the criteria but did not report enough data for calculating the effect size. Two efforts were successful, but 5 studies were dropped due to inadequate data.

Finally, 31 studies, including 19 journal articles, 10 dissertations, 1 conference paper, and 1 book were identified as the population of the present meta-analysis. They were all coded with the same coding sheet which will be discussed in the next section.

Coding Research Findings and Study Characteristics

This process is to obtain data on research findings and a set of selected study characteristics from each individual study. The process of coding information from each individual study in a meta-analysis is similar to the process of collecting data from each subject in a primary study.

In order to calculate effect size, mean of both experiment and control group, the standard deviation should be coded from the original study. Some studies report only the result of inference test, and it can be converted into effect size by using relevant formulas.

Choice of Study Characteristics

Meta-analysts use data on study characteristics for descriptive statistics, and moderator analysis. The relationships between study characteristics and training

outcomes may suggest which variable (if any) mediates reported effect size, or significance levels of training outcomes.

As a part of problem formulation, Hedges et al. (1989) pointed out that the researcher had to decide whether the review was to be confirmatory or exploratory. If the review is exploratory in nature, then the researcher may code a large number of study features, as in the approach used by Glass et al. (1981). Otherwise, the researcher will only code important study features that are contributing factors to the variance of research outcomes suggested by theoretical or conceptual frameworks and previous studies. The statistical procedures are for testing hypotheses formulated beforehand. Both Hunter et al. (1981) and Hedges et al. (1989) pointed out that when conducting many tests in an "exploratory" mode, there was a tendency to "capitalize on chance".

In the present study, the researcher selected five study features (measurement criteria of training outcomes, training content, training method, training needs assessment, and favorable condition of transfer of training) based on literature review and previous studies. The hypothesis test was confirmatory.

Some other straight or qualitative attributes of the individual studies were also coded. These study features included: date of publication, publication form (journal article, book, dissertation), ranking of subjects (executive, manager, administrator, or supervisor), location of study (laboratory vs. work setting), research design, assignment of subjects (random vs. nonrandom), control group, and pre-, posttest. These data were used to delineate a profile of the research studies on the effectiveness of management training.

Code Book and Coding Sheet Development

The coding sheet is an instrument for coding information and data from each primary study. It was developed based on the code book, which decided the scale of measurement (nominal, ordinal, and interval or ratio scale) and assigned a code to each category or level of a variable. First, the author developed a code book by referencing the check list of code book from Hedges et al. (1989, p.19). Then a coding sheet was developed based on the code book. Every primary study was coded by using the same coding sheet. The coding sheet was attached as Appendix A.

In the present study, the coding process of three major study features involved subjective judgment, and was unique. The first study feature of these three was the training method. Unlike Burke and Day (1986), and Bennett (1995), there were only three categories of this variable in this study—cognitive methods mainly, behavioral methods mainly, and combination of cognitive and behavioral methods. The coder first coded various training methods such as traditional (lecture, discussion), role playing, behavior-modeling, that were reported in the primary study, and then classified them as one of the three categories.

There was little direct information on training needs assessment (TNA) and favorable condition of transfer of training (TOT) in training evaluation studies. Bennett (1995) found that only 7% of his primary studies reported training needs assessment activities, and no study reported favorable conditions of transfer of training. Although straightforward information was deficient, there were a lot of scattered information in evaluation studies that suggested or implied training needs assessment activities and favorable conditions of transfer of training. These fragmentary information was caught

up and recorded on the coding sheet. The total score a study received then was converted to one of the three ranks of the variable.

For example, one of the primary studies (Earley, 1987) examined the effect of an intercultural training of a manufacturer organized for its low-level management personnel. The company was sending the participants "to Seoul, South Korea for three months to teach the production and manufacturing techniques that several assembly plants being set up there would be using" (p. 688). "Their participation in the study was a mandatory part of their preparation for traveling overseas" (p. 688). These information suggested that the training was conducted according to the organizational needs. So the study received 1 score for organizational analysis on training needs assessment. The description of training material development and delivery showed that personnel representatives previously worked in Korea who were familiar with the culture participated in material development and training. This implied that the task analysis was well conducted. Trainees participated several experiential learning activities, such as role playing, simulation and field work, and received feedback about their attitude and behavior. This process had a similar function as the individual analysis. Thus, the study received a score of 3, which was then converted to the highest rank—"Adequate TNA was conducted or reported (score 3 and more)" (see the coding sheet). For the variable of favorable condition of transfer of training, at least three factors could be checked out—transfer motivation and opportunity to use due to the real overseas assignment, and supervisor support by the company's positive attitude toward the training. So this study received at least a score of 3, which was then converted to the highest rank—"Strong favorable condition of transfer of training (score 3 and more)".

Coding Procedures and Reliability of Coding

The author of this dissertation coded all the primary studies with the coding sheet.

An important advice from Hedges et al. (1989) about "over-specify categories at the coding stage rather than to under-specify" was listened and followed since "categories can always be grouped at the analysis stage" (p. 20).

A quality assurance procedure is to check the reliability of coding. Normally an inter-rater agreement will be obtained at an acceptable level if there are more than one coders coding the primary studies. In the case of a single coder, an auditing conducted by an expert on the coding results served as an effective and efficient check of consistency. According to Guba (1981), auditing was the most useful and feasible way to demonstrate consistency in naturalistic inquiring. He argued that "one criterion for the necessity and sufficiency of a categorical set is its reproducibility by another competent judge. ... a second judge should be able to verify that the categories derived by the first judge make sense in view of the data pool from which the first judge worked and that the data have been appropriately arranged into the developed category system" (p. 122).

One of the committee members, Dr. Conti, volunteered this service. He audited three journal articles which were randomly selected from the 21 that formed the major sub-population of the total primary studies for this research. He read the articles, checked the coding results, gave alternative answers to several items, and pointed out a few points that need clarification. After an in-depth discussion, a total agreement was reached between the researcher and the auditor on major coding results—classification

of training content, training method, criteria of outcome measure, ratings on training needs assessment and favorable condition of transfer of training, research design characteristics, and sample size. Dr. Conti also suggested to add a category of "not applicable" to certain items, and his suggestion was accepted.

In the case of a single coder, other methods for preventing error and enhancing coding quality included checking computations by using alternative formula (e.g., effect size calculation), seeking multiple evidences (e.g., sample size), and getting agreement with the original author(s) (e.g., the type of research design, the positive /negative direction of the effect). In general, the reliability of coding was acceptable for this study.

Data Set for Meta-Analysis

This study used the modern methods of meta-analysis illustrated by Hedges et al. (1985, 1989). Before applying any statistic procedure of meta-analysis, a quality data set must be established based on the coded information from individual studies. There are three major decisions and steps in preparing the data set. The first is to calculate estimate of effect size and its variance from each individual study. The second involves decision on unit of analysis. In order to get independent data points, those that are dependent must be combined before getting into the data set. The third step is to identify outliers or extreme data points.

Effect Size Estimate and Variance Computation

Effect size is the most commonly used numerical index of study outcome. The effect size of management training programs is defined as the difference between the population mean criterion scores for treatment and control groups expressed in standard deviation unit. In practice, effect sizes are estimated from data of samples. When the sample sizes are not large, an unbiased effect size estimator can be produced by using the correction factor provided by Hedges & Olkin (1985).

The unbiased effect size estimate d is the product of J and g. In the formula, g is the uncorrected estimate of effect size, and J is a multiplier for correcting the bias of estimation caused by small sample size, which depends on the degrees of freedom for the standard deviation used in the formula of effect size. According to Hedges et al. (1989), when degrees of freedom is above 50, the correction factor is between 0.99 and 1, and so can be ignored. Relevant formulas are listed below (cf. Hedges et al. p. 26):

$$g = \frac{\overline{X_t} - \overline{X_c}}{S} \tag{3-1}$$

$$d = J * g \tag{3-2}$$

$$J = 1 - 3/(4m - 1) \tag{3-3}$$

$$m = N_t + N_c - 2 \tag{3-4}$$

g is the uncorrected effect size estimate,

$\overline{X_t}$ is the mean of the treatment group with sample size of N_t

$\overline{X_c}$ is the mean of the control group with sample size of N_c

S is the pooled standard deviation of treatment and control group

m is the degree of freedom of S

N_t and N_c are the sample size of the treatment group and the control group

J is the correction multiplier, and d is the corrected (unbiased) effect size estimate.

The sampling variance is measured by the square of the sampling standard error. A very accurate approximation to the sampling variance of the effect size estimate d is

$$S_d^2 = \frac{N_t + N_c}{N_t N_c} + \frac{d^2}{2m} \qquad (3\text{-}5)$$

According to Hedges et al. (1989), "the sampling standard errors of individual effect size estimates provide weithts for optimally combining effect sizes across studies and in addition provide information for computing standard errors of combined effect sizes" (p. 35). In case of t or F test results reported in the study, converting formulas were found in Hedges et al. (1985, p.30-p.33), Glass et al. (1981. P. 128), Hunter and Schmidt (1990, p. 273).

Unit of Analysis and Independent Data Point

Meta-analytic procedures assume independence of the units of analysis (Mullen & Rosenthal, 1985). The objective of selecting unit of analysis is to ensure obtaining the maximum amount of independent data available from primary studies, while eliminating non-independent data. In the present study, an effect size estimate was the average of effect sizes of each evaluation criterion for each independent experimental group in a study. If there was more than one measurement criterion used in a study, such as learning, behavior, or result, the effect size were calculated separately for each measurement criterion since the outcomes being measured were quite different in nature. If the outcome of the same criterion from the same experimental and control

groups was measured more than once after the training intervention (e. g., measured by self, by subordinates and/or by supervisor with similar instrument, or measured more than one time), these data should be averaged to get one effect size since they were not independent. If there was more than one pair of experimental and control groups in a study, each pair should get its own effect size, since it was independent data from different subjects. By doing so, one primary study might contribute more than one independent data point for accumulation.

The average of dependent effect sizes can be calculated according to the rules provided by Hedges et al. (1989), who called this operation "combining correlated effect size estimates" (p. 37).

Outlier Identification

Extreme data point will distort the study result, so should be identified and rejected. Hedges et al. (1986, p. 39) suggested to conduct a pre-heterogeneity test to the data set to identify those data points with suspicious high Q value, which measured the heterogeneity. The meta-analyst then were well advised to examine the relevant studies. It helped find errors in coding and calculation, or even suspect or mistake in the original study. If a study with high Q value had special study features, and its data were reliable, it was kept in the data set by the present meta-analysis. This process involved judgement of the researcher.

Meta-Analytical Procedures

After rejecting outliers, the data set is ready to apply statistical procedures of a meta-analysis. There are two major analytic approaches to perform the two major functions of a meta-analysis--combining research results of individual studies in the same research domain by using a common metric; examining the variability of research findings across studies (Wolf, 1986, Hedges et al., 1989).

Combination of Effect Size Estimates Across Studies

One objective of meta-analysis is to combine estimates of effect size from individual studies to get an overall average, an estimate of the treatment effect magnitude for the population. The effect size estimates in the data set are independent. "It is well-known that independent measurements can be averaged properly only if each measurement is weighted by the inverse of its variance of error" (Hedges, et al. 1989, p. iii). "The power and sensitivity of meta-analysis comes from the fact that this combined estimate will have smaller standard error than any of its parts" (p.36).

When $d_1, d_2, ..d_k$ are k independent effect size estimates and $S_1, S_2, ...S_K$ are their standard errors (standard deviation of sampling errors), the weighted average effect size is

$$d_{+} = \frac{d_1 / S_1^2 + d_2 / S_2^2 + ... + d_k / S_K^2}{1/S_1^2 + 1/S_2^2 + ... + 1/S_K^2} \qquad (3\text{-}6)$$

The standard error of the weighted average is

$$S_{d_{+}} = 1/(1/S_1^2 + 1/S_2^2 + ... + 1/S_K^2)^{1/2} \qquad (3\text{-}7)$$

The assumption here is that the underlying (population) effect size is identical in all of the studies. It is also called the assumption of homogeneity of effect sizes. Otherwise, "the representation of the results of a set of studies by a single estimate of effect magnitude can be misleading" (Hedges & Olkin, 1985, p. 147).

The present study paid special attention to the homogeneity assumption of the formula for combining effect size estimates across studies. An average effect size was calculated only to those groups and subgroups that passed the homogeneity test.

If the homogeneity assumption was not met, the data set then was grouped by similar characteristics. The grouping process stopped when an acceptable level of subgroup homogeneity was obtained.

Analysis of Heterogeneity

The analytical procedure of comparison of studies' results is also called analysis of heterogeneity. It serves as the test of homogeneity as well. Effect size estimate of individual studies could be differ by many times their sampling standard errors. The variation among studies is partially caused by sampling error, and partially caused by differences in studies. A heterogeneity analysis could be conducted to study the non-sampling variation. The statistic measure of heterogeneity is Q.

Suppose that $Q_1, Q_2, \ldots Q_K$ are the individual squared deviations, then the formula for the heterogeneity statistic Q is

$$Q = Q_1 + Q_2 + \ldots + Q_k = ((d_1 - d_+)/S_1)^2 + ((d_2 - d_+)/S_2)^2 + \ldots + ((d_k - d_+)/S_k)^2 \quad (3\text{-}8)$$

The Q statistic is distributed as a chi-square variable with (n-1) degree of freedom (n is the number of the data points of the distribution). If the Q value of a group of independent data points is smaller than the critical value of the chi-square distribution (df=n-1) at a selected significant level (e.g., α=0.05), it failed to reject the null hypothesis that all studies share a common population effect size. Therefore, the group under examination passes the homogeneity test (please note that the standard is not very restrict as the type II error is large if the null hypothesis is false). If a group of effect sizes has a significant Q statistic, the homogeneity null hypothesis is thus rejected at the α=0.05 significant level. The meta-analyst may owe the excess variability to certain study characteristics of that the effect size estimates are computed. Then the data points of effect size are sub-grouped according to study characteristics in order to achieve homogeneity of subgroups. Theoretically this sub-grouping process could be continued until the group homogeneity is achieved. In practice the number of data points will limit it to one or two turns. In the present study, the effect size data points were grouped by measurement criteria first, and then by training content and method respectively. When examining the impact of training needs assessment, data points were grouped by three levels of this moderator variable. Since the possible moderator of favorable condition of transfer of training only influence the behavior and result measurement, data points of learning criterion were excluded from that specific distribution.

An Analogue Weighted ANOVA

The two above procedures are employed through an analysis of variance for effect sizes—an analogue weighted ANOVA (Hedges & Olkin, 1985, Chapter 7, p.147-165).

In the weighted ANOVA, the dependent variable is the effect size, the reciprocal, squared sampling standard errors are used as weights, and the factors in the analysis of variance are those selected study characteristics that are most likely to explain the variance among studies (Hedges et al., 1989, p. 43). Statistics Q_T, Q_B, Q_W represent the "total fit", "between-group fit" and "within-group fit" to the model respectively.

$$Q_T = Q_B + Q_W \tag{3-9}$$

The sub-grouping process is stopped at where the Q_W is small or non-significant. The average effect size for the *i*th homogeneous subgroup was the weighted mean for that specific category. Q_B represents the extent to which the effect sizes differ among subgroups. Its distribution is that of a chi-square statistic with (k-1) degrees of freedom (k is the number of groups). If Q_B exceeds the critical value, the hypothesis that the average effect sizes are equal across groups is rejected at the selected significant level.

Test of Null Hypotheses

An average effect size is meaningful only when the assumption of all studies sharing a common population effect size is true at an acceptable level. After a d_+ and its standard error S_{d+} are calculated, an approximate confidence interval for the population effect size δ is determined upon them. It tests the null hypotheses (generally expressed as H_{01} in this study) of whether the management training makes a real difference on participants' learning, on-the-job behavior and/or organization results.

When more than one group or subgroup generates meaningful d_+ and S_{d+}, and the between-group heterogeneity Q_B is of significant value, it indicates that the average effect sizes are not equal across groups. This suggests that the dividing factor is a potential moderator. The comparisons among groups (Hedges & Olkin, 1985, p. 159) is conducted to test the null hypotheses (generally expressed as H_{02} in this study) of whether the study characteristic makes a real difference on the magnitude of effect size.

Summary

This methodology chapter described step by step how the meta-analysis was conducted in this study. Following the general sequence of conducting a normal research such as problem formulation, data collection and evaluation, data analysis and interpretation, every special step and decision for a meta-analysis was specified and explained.

An Analogue weighted ANOVA applied the two major meta-analytical procedures in an integrated manner. The grouping and sub-grouping process were based on the conceptual understanding about the independent and dependent variables, while the statistics and the numbers were only tools to help either reject or confirm the researcher's hypotheses and judgments. Therefore, an average effect size was obtained only when the group or subgroup passed the homogeneity test.

Later in Chapter IV, the two generally expressed null hypotheses H_{01} and H_{02} were specified according to the results of combination of effect sizes, and then were tested to determine significance. The results of testing the null hypotheses derived from H_{01} determined whether the average effect size d_{+I} (estimate of the magnitude of the

true population training effect) was a significant training effect, but not sampling error. The results of testing the null hypotheses derived from H_{02} determined whether the identified factor was a significant moderator variable of the magnitude of training effect, but not occurred by chance.

CHAPTER IV

ANALYSIS OF DATA SET FROM THE STUDY POPULATION

Data Set for Meta-Analysis

The literature search guided by the inclusion criteria identified 47 studies. Thirty-one studies with adequate information were actually coded. Altogether 163 estimates of effect size were calculated from the 31 studies. Those that were dependent (i.e., estimates of effect size calculated from repeated measures by self, supervisor, subordinates, or expert of the same pair of experimental and control group) were averaged to become an independent data point. The 31 primary studies generated a total of 68 independent data points. The pre-heterogeneity test for outlier identification then was performed on these independent estimates of effect size. Seven data points with extreme Q value (from 7.981 to 93.352) that came from 3 studies were eliminated. There was no cutting line for the Q value. Another 13 data points with a Q value higher than 5 were re-examined. The studies that generated them had special study features, and the original statistical data were clear and understandable. These 13 data points were kept in the data set. After this process, a total of 61 data points from 29 studies formed the data set for this meta-analysis.

Profile of Management Training Evaluation Studies

Research Question 1: What are the major characteristics of empirical evaluation research studies on management training in the 15-year time period of 1983-1997?

The study population consisted of 29 management training evaluation studies, of which 18 (62%) were journal articles and 9 (31%) were doctoral dissertations. The other two studies were a conference paper and a book. Seventeen studies (59%) were published in the 1980s, and the other 12 (41%) were published in the 1990s. Major characteristics such as publication year, author(s), organization type, trainee, and program content and method of these studies were listed in Table 4.1. The profile of the study population was described as follows.

Organization Setting and the Trainee Position

Among the 29 studies, 14 (48% of the study population) evaluated management training programs that were conducted in business and industry for supervisors (6 of 14), managers (6 of 14), and executives (2 of 14). They generated data points for a total of 33 (54% of the data set), and 10, 21, and 2 for respective trainee position. The second biggest organization setting is education, which constitutes 34% of the study population. Nine studies evaluated training effect of educational administrators (school principals, university department heads, and vice presidents), and one study evaluated training effect of managers in education organization. A total of 14 data points (23%) were generated from education organization. The other studies involved military, government, and hospital, and they accounted for 18% of the study population and 23% of the data set (see Table 4.2).

Table 4.1

Management Training Evaluation Studies Included in Meta-Analysis

Year	Author(s)	Settings	Trainee	Sample Size	Training Contents	Training Methods	Training Outcomes	Criteria & Indep. Data Point
1983	Larson, J. E.	Public sector (hospital)	middle level managers	31-T 31-C	group leadership and self-management	traditional, case study, exercises, role playing	reaction, learning, managerial behavior, work-group effectiveness,	SR (2)
1983	Smith, J. F., Boshoff, A. B., & de V. Vissor, M.	Business and industry (manufacturing)	production supervisors	34-T 34-C	criterion-referenced instruction on reducing negative critical incidents	programmed instruction	results: number of negative critical incidents, productivity	OR (2)
1984	Davis, B. L. & Mount, M. K.	Business and Industry (multinational corporation)	middle-level manager	260-T 142-C	rater training (performance appraisal)	computer-assisted-instruction (CAI); CAI plus behavior modeling;	knowledge learning test; attitude learning assessment; rating on appraisal form effectiveness; subordinates' satisfaction on appraisal system and discussion	SL(2), OL(2), SB(2), OR(2)
1984	Russell, J. S., Wexley, K. N., & Hunter, J. E.	Business and Industry (plant)	supervisor	22-T 22-C	human relations	behavior modeling taught by Professor trainers, or by managers	reaction; learning test on managerial incidents; self and manager rated behavior change, peers' rated job performance	OL(2), SB(2), SB(2)
1985	Birkenbach, X. C., Kamfer, L., & ter Morshuizen, J. D.	Business and industry (manufacturing)	first-line supervisors	50-T 50-C	supervisory skills training	both cognitive methods and behavior-modeling	learning (knowledge and behavior); behaviorally based performance; organization results (disciplinary cases, absenteeism rates, grievance reports)	SB (1)
1985	Matton, R. J.	Business and industry (news paper)	managers	18-T 18-C	management training and development	traditional, films, T-group, assessment and feedback, action plan and homework	job performance appraisal rated by supervisor	SB (1)
1985	Urban, T. F. et al.	Business and industry (oil company)	supervisors	533-T 105-T 105-C	comprehensive supervisory training	not specify	turnover (attrition) rate, career progression (position change and pay grade increase)	OR (2)
1986	Briddell, W. B.	Education	Director, Dean, and above	12-T 12-C	time management workshop	traditional (lecture, discussion)	degree of occupational stress, type A behavior	SB (1)
1986	Frost, D. E.	Public sector (fire department)	supervisors	60-T 25-C	leadership development program	Leader Match and tradition method as alternative	leadership performance rating by trainees' supervisor	SB (2)

Table 4.1 (Continued)

Year	Author(s)	Settings	Trainee	Sample Size	Training Contents	Training Methods	Training Outcomes	Criteria & Indep. Data Point
1986	May, J. M., Keys, C. B.	Public sector (helping profession)	middle managers	100-T 87-C	general management program	combine didactic with experiential, small group activities	knowledge test, telephone interview on self-reported behavior change	OL(1), SB(1)
1987	Earley, P.C.	Business and Industry (manufacturer of electronic products)	low-level management	60-T 20 C	intercultural training	documentary, interpersonal, and combination	self-reported and supervisor-rated performance, self-assessed intensity of adjustment to new culture	SB(3), SB(3)
1987	Mathieu, J. E., & Leonard, R. L. Jr.	Business and industry (banking)	supervisors	65-T 65-C	supervisory skills training	tradition plus behavior modeling	performance (behavior) rating by trainees' supervisor; utility analysis	SB (1)
1988	Esposito, C.	Education	superintendent	17-T 23-C	National Superintendents' Academy--a general executive development program	multiple methods	leadership behavior and management practice	SB (1)
1988	Schwier, R. A., & Misanchuk, E. R.	Business and industry (food supermarket)	assistant managers and department managers	24-T 12-C	technical training (financial management)	computer-supported-training	learning test, time to complete instruction	OL (2)
1989	Feldhusen, J. F., Haeger, W. W., & Pellegrino, A. S.	Education	school administrators	45-T 51-C	technical skills training (program planning and development for gifted)	tradition, exercises, field visit	knowledge test on gifted education; attitudes change toward gifted education	SL (1), OL (1)
1989	Gist, M. E.	Government (a federal agency for R&D)	managers	29-T 30-C	training on innovative problem solving	traditional and cognitive modeling	self-efficacy for the idea generation tasks; idea generation performance task scores	SL (1), OL (1)

Table 4.1 (Continued)

Year	Author(s)	Settings	Trainee	Sample Size	Training Contents	Training Methods	Training Outcomes	Criteria & Indep. Data Point
1989	Gist, M. E., Schwoerer, C. & Rosen, B.	Education (university)	managers and administrators	54-T 54-C	technical training (use of computer software)	tutorial (cognitive), and behavior modeling	software self-efficacy (self reported software learning ability developed through training), task performance, working style (an index of constructive approaches to work)	SL(1), SL(1), OL(1)
1990	Nelson, E. P.	Education	beginning principals	30-T 14-C	general administrative and leadership skill training	role-playing, behavior modeling, simulation, and mentoring	measured two dimensions of leadership behavior, rated by self and subordinates	SB (1)
1990	Whitford, E. V.	Education	school administrators	13-T 13-C	technical skills training (writing of supervisory reports)	cooperative learning	holistic scores on participants' "supervisory observation report" written on the job	OR (1)
1991	Niska, J. M.	Education	principals	13_T 13-C	cooperative learning supervision training	traditional, simulation, modeling & demonstration, feedback, coaching for application	skills of providing feedback, knowledge and use of cooperative learning concepts, level of confidence	SL (1), SB (1)
1992	Edwards, W. S.	Education	secondary principals and assistant principals	29_T 39-C	training seminar on safe schools (crisis management)	traditional, case study (scenario)	crisis management ability tested by the CMSI instrument	OL (1)
1992	Harrison, J. K.	Military	civilian managers, including management trainees	33-T 32 C	cross-cultural management training	cultural assimilator, behavior modeling, combination	reaction, learning test, assessment of role playing	OL(3), OL(3)
1992	Smith, R. M., Montello, P. A., & White, P. E.	Education	administrators	10-T 12-C	interpersonal skills as leader	tradition, and role playing	interpersonal skills demonstrated in a video-taped interview role playing	SL (1)

Table 4.1 (Continued)

Year	Author(s)	Settings	Trainee	Sample Size	Training Contents	Training Methods	Training Outcomes	Criteria & Indep. Data Point
1992	Sniderman, R. L.	Business and industry (retail)	executives	49-T 20-C1 27-C2	general management development	traditional, case study, exercises, simulation, SYMLOG assessment, group dynamics	managers' interpersonal behavior measured by subordinates with SYMLOG rating form	SB (1)
1993	Bankston, J. R.	Education	principals	13-T 15-C	instructional leadership training	traditional, assessment and feedback, improvement plan and implementation strategies	instructional leadership behaviors rated by participants and their subordinates	SB (1)
1995	Engelbrecht, A. S., & Fischer, A. H.	Business and industry (assurance society)	supervisors	41-T 35-C	general management program (developmental assessment center)	developmental assessment center process (mainly behavioral methods)	managerial job performance	SB (1)
1995	Jones, R. G., & Whitmore, M. D.	Business and industry (insurance company)	managers	113-T 167-C	developmental assessment center (general management)	various behavioral methods	promotion to division level position; recent performance appraisal rating	SB (1)
1996	Barling, J., Weber, T. & Kelloway, E. K.	Business and Industry (banking)	branch managers	9-T 11-C	leadership training	traditional, behavior modeling, supervision and monitoring	subordinates' perceptions of behavior change, subordinates' organizational commitment, branch financial performance	SB(1), SR(1), OR(1)
1996	Young, D. P., & Dixon, N. M.	Business and industry, Government, and Education	mainly executives, and upper-middle managers	29-T 38-C	self-awareness programs	tradition, exercises, experiential and nontraditional learning activities	making positive behavior change to be more effective on the job	SB (1)

Training Content and Training Method

There are a total of six management training content areas covered by research studies in this study population. Human relations/leadership training is the biggest focus of training evaluation studies. Thirteen studies evaluated training programs with the objective of enhancing interpersonal skills and/or leadership behavior of the managerial personnel. It counts for 45% of the study population. A total of 31 data points (51% of the data set) were generated from this category (Table 4.3). The second large content area is the general management training (6 studies with 7 data points) followed by technical training (5 studies with 10 data points). They account for 38% of the study population and 28% of the data set. The remaining 17% includes 2 studies of self-awareness program and 2 studies of problem-solving/decision-making program, and 1 study of rater training program. They contributed 13 data points (21% of the data set). Regarding the training method, all studies used more than one training technique. By classifying them into three categories, the biggest group is the "combination of cognitive and behavioral methods". Twenty-six effect size estimates (43%) were generated from it. More than one-third (22 of 61) effect size estimates come from the "cognitive methods mainly" category, and the remaining 13 effect size estimates were in the "behavioral methods mainly" category. In order to compare training effects of different methods, two studies used all three types of methods, and a few others employed an alternative teaching method. The distributions of training content and training method were summarized in Table 4.3 and Table 4.4.

Table 4.2

Distribution of Organization Setting and Trainee Position

Organization Setting	# of Study	% of Study Population	# of Data Point	% of Data Set	Trainee Position	# of Study	% of Study Population	# of Data Point	% of Data Set
Business & industry	14	48	33	54	supervisor	6	20.5	10	16.4
					manager	6	20.5	21	34.4
					executive	2	7	2	3.2
Education	10	35	14	23	manager	1	4	3	5
					administrator	9	31	11	18
Military, Government, Hospital	5	17	14	23	supervisor	1	3	2	3
					manager	4	14	12	20
Total	29	100	61	100		29	100	61	100

Table 4.3

Distribution of Training Content

Training Content	# of Study	% of Study Population	# of Data Point	% of Data Set
Human relations/Leadership	13	45	31	51
General management	6	21	7	12
Technical training	5	17	10	16
Self-awareness program	2	7	2	3
Problem-solving/decision making	2	7	3	5
Rater training	1	3	8	13
Total	29	100	61	100

Table 4.4

Distribution of Training Method

Training Method	# of Study	% of Study Population	# of Data Point	% of Data Set
Combination of cognitive and behavioral methods	17		26	43
Cognitive methods mainly	11		22	36
Behavioral methods mainly	6		13	21
Total	>29*		61	100

* Some studies employed more than one type of training methods.

The training method used in a program is highly correlated with the nature of the skill/task that is being taught. There is no psychomotor skill/task taught in managerial training. Ten programs taught cognitive task and generated 21 data points (34% of the data set). Eight programs focused on interpersonal skills and resulted in 8 data points (13%). The majority of the studies (16 studies with 32 data points) evaluated programs that aimed to train managerial personnel with multiple skills.

To train cognitive skills, 76% (16 of 21) of the programs were taught by cognitive methods, 14% (3 of 21) by behavioral methods, and 10% (2 of 21) by combined methods. To train interpersonal skills, 50% (4 of 8) of the programs were taught by behavioral methods, 25% (2 of 8) by cognitive methods, and the remaining 25% by combined methods. When the training skills/tasks were multiple, the dominant method, which was employed by 69% (22 out of 32) of the programs, was the

combination of cognitive and behavioral methods. Either cognitive methods or behavioral methods taught the other programs (12% and 19% respectively).

Training Outcomes Measurement Criteria

The data presented several facts. First, all studies used more than one level of criteria to evaluate training outcomes with only a few including reaction measurement. Second, 15 studies (52%) with 25 data points (41%) evaluated participants' learning. Only 7 studies (24%) were limited to internal criteria, i.e., not beyond the level of learning. Third, 17 studies (59%) with 24 data points (39%) evaluated trainees' job performance, of which many used multiple measurements. Last, 8 studies (28%) with 12 data points (20%) evaluated organization results. The data of measurement criteria were summarized and listed in Table 4.5.

These facts suggested that major efforts and progress have been made by the training field to demonstrate the effects of training. Learning has always been and still is an important outcome of training to be evaluated. However, the situation of "exclusive reliance of management development research on internal criteria" (Campbell et al., 1970; Clement, 1981) has been changed dramatically. Training professionals are skillful in developing knowledge tests, not only for cognitive learning skill/task but also for behavioral learning skill/task, to measure trainees' learning objectively. Depending mainly on performance appraisal instruments, about 59% of the primary studies produced multiple measures of job behavior (by trainee, their supervisor, subordinates, or expert) that were then converted into effect size for this meta-analysis. However, the measure of organization results was still in a pre-matured

Table 4.5

Distribution of Training Outcomes Measurement Criteria

Measurement Criteria	# of Study	% of Study Population	# of Data Point	% of Data Set
Subjective learning (SL)	5		7	11.5
Objective learning (OL)	10		18	29.5
Subjective behavioral (SB)	17		24	39
Subjective results (SR)	3		5	8
Objective results (OR)	5		7	12
Total	40*		61	100

stage. The indicators used in this category and the methods used to determine their magnitudes were highly diversified. About half of the studies that measured organization results were dropped by the present study due to inadequate data to calculate effect size.

Training Needs Assessment and Favorable Condition of Transfer of Training

There was little direct information on training needs assessment (TNA) and favorable condition of transfer of training (TOT) in training evaluation studies. By treating these two variables as study features rather than the manipulated independent variables, the present study created a way to code TNA and TOT information from the primary studies.

Fourteen data points (23% of the data set) were produced from training programs without training needs assessment efforts. More than half (56%) reported some (score 1-2), and 21% reported adequate (score 3 and more) needs assessment activities. The programs with TNA accounted for 77% of the total data set. This was a remarkable effort and progress made by training professionals to enhance the quality and effect of training interventions.

It was interesting to see the relationship between TNA and TOT. Among programs without TNA, the majority (64%) had no favorable condition of transfer of training. Among programs with moderate TNA, most (85%) had moderate favorable condition of TOT. The programs with adequate TNA had the same phenomenon of matching (see Table 4.6). The positive correlation of these two variables was calculated and discussed more lately in the moderator analysis section.

Training Duration

Training duration was related to the length of training time that the trainee received. Although the present study did not include it as a potential moderator, it influenced the training result. The data in Table 4.7 gave information about the program profile from this perspective.

The largest category (41%) was the intensive program (one to several days). It fit the situation of busy working management personnel. Normally the short program (i.e., only a couple of hours) was not long enough for teaching behavioral or multiple skills. The coded data showed that the 6 studies (about 20% of the study population) all taught cognitive skill/task.

Table 4.6

Distribution of Training Needs Assessment and Favorable Condition of Transfer of Training

Training Needs Assessment (TNA)	# of Data Point	% of Data Set	Favorable Condition of Transfer of Training (TOT)	# of Data Point	% of Data Set
No TNA conducted or reported	14	23	No favorable condition of TOT	9	15
			Moderate favorable condition of TOT	5	8
Moderate TNA conducted	34	56	No favorable condition of TOT	1	1.5
			Moderate favorable condition of TOT	29	47.5
			Strong favorable condition of TOT	4	7
Adequate TNA conducted	13	21	Moderate favorable condition of TOT	2	3
			Strong favorable condition of TOT	11	18
Total	61	100		61	100

Attention should be paid to "intensive program plus distributed sessions" although it was the smallest category (17%). The distributed sessions were developmental process for trained managerial personnel in their job situations. The previously informal on-the-job experience, conditional transfer of training, and provisional mentoring became an integrated part, which was formal, well designed and well implemented, of the training program. The coded data showed that these programs all taught multiple skill/task, had at least moderate and most of adequate training needs

Table 4.7

Distribution of Training Program Duration

Training Duration	# of Study	% of Study Population	# of Data Point	% of Data Set
Short (< 1 day)	6	21	14	23
Intensive (one to several days)	12	41	25	41
Intensive program plus distributed sessions	5	17	7	11.5
Regular program lasting several weeks/months	7	24	15	24.5
Total	30*	>100	61	100

* One study compared two training programs of different duration.

assessment, and all had a strong favorable condition of transfer of training. This was a new model and direction to design training programs with the objective of improving trainees' job performance.

Research Design Characteristics

Table 4.8 summarized coded data of research design characteristics of primary studies. All 29 studies were conducted in work setting. To find out the true cause-effect and control group. As stated earlier in Chapter III, one group pretest and posttest design and static two group comparison (i.e., no random assignment of treatment to groups and no control on pre-existed difference) were excluded in this meta-analysis, because their results had little research value.

Table 4.8

Research Design Characteristics of Primary Studies

Research Design Characteristics	Classification	# of Study	% of Study Population	# of Data Point	% of Data Set
Research Design	quasi-experimental	13	45	15	25
	experimental	16	55	46	75
Random Sampling	not random sampling	22	76	42	69
	random sampling	7	24	19	31
Random Assignment to Treatment	no random assignment	14	48	16	26
	random assignment	15	52	45	74
N>=30	N<30	7	24	11	18
	N>=30	22	76	50	82
Control of pre-existed difference	pre-test	17	59	33	54
	matched control group	7	24	11	18
	others	5	17	17	28
Significance of result	not significant	N/A		26	43
	significant	N/A		35	57

When inference statistics are employed in the research, researchers will do their best to increase the sample size. Although it was difficult in conducting experimental research in the work setting with managerial personnel, the percentage of studies with a total sample size larger than 30 is 76%. Sample size is very important in meta-analysis as well. The small sample size causes bias in the estimated effect size. It was corrected in this study by applying Hedges et al. (1985, 1989) formula. When combining the effect sizes across studies, the individual effect size was weighted by its sample size.

This means that a study with a small sample size only makes a relatively small contribution to the combined result.

Regarding the significance of the research findings of primary studies, among the 61 independent data points, 26 (43%) were non-significant. Researchers do publish their non-significant results. Thus the worry about publication bias could be relieved. More importantly, a non-significant result does not always equal to no effect of treatment. If individual non-significant results are of positive value repeatedly, they can make contributions to a significant accumulated result when the meta-analysis approach is applied across studies.

Management Training Effectiveness

Research Question 2: What is the magnitude of training effect measured by each of the five criteria (SL, OL, SB, SR, OR) of management training programs? Under each criterion, what is the magnitude of training effect of management training programs that teach different content or employ different methods?

Theoretically, this meta-analysis might yield 45 distributions (representing 5 measurement criteria, 6 training content areas, and 3 training method categories). In fact, several combinations of the three factors were vacant or had only a small number of data points. Hence a full range of spectrum was not available.

Since different criteria measure different training outcomes, the whole data set was heterogeneous. Thus a meaningful overall effect size (the grand d_{++}) of management training was not given. However, through grouping and sub-grouping the

data set by using moderators, several groups and subgroups passed the homogeneity test and produced a meaningful average effect size (d_+) for its category.

Training Effect of the Whole Data-Set
and Training Effect by Criteria

First, the data set was grouped by the five pre-determined measurement criteria: subjective learning (SL), objective learning (OL), subjective behavior (SB), subjective results (SR) and objective results (OR). Table 4.9 is the result of the analogue weighted ANOVA for this data set.

As expected, the whole data set is highly heterogeneous. The Q_T value of 206.284 was significant at α=0.001 level. All five groups divided by measurement criteria were also heterogeneous except the subjective result (SR) group. The rest of the analysis was conducted on each sub-data-set of a specific measurement of criterion.

Training Effect of the Subjective
Learning Sub-Data-Set

As shown in Table 4.9 the subjective learning sub-data-set was heterogeneous (Q=35.491, significant at α=0.001 level). After several attempts to get a homogeneous subgroup, the researcher found that "what's measured" was the most influential factor of the effect size.

Table 4.9

Effect of Management Training by Criteria and Overall

Source		Q value	Degrees of freedom	Test of homogeneity at α=0.05 level	d_+	S_{d+}
Between groups	Q_B	31.874	4	heterogeneous		
Within groups	Q_W	174.41	56	heterogeneous		
Within group 1 (SL)	Q_{W1}	35.491	6	heterogeneous	0.468	0.071
Within group 2 (OL)	Q_{W2}	29.532	17	heterogeneous	0.801	0.063
Within group 3 (SB)	Q_{W3}	92.237	23	heterogeneous	0.503	0.055
Within group 4 (SR)	Q_{W4}	4.193*	4	passed the test	0.245*	0.089
Within group 5 (OR)	Q_{W5}	12.957	6	heterogeneous	0.487	0.077
Total	Q_T	206.284	60	heterogeneous	0.535	0.032

Note: Q value with * is smaller than the critical value of the chi-square distribution with the specific degrees of freedom at the α=0.05 level. This means that the group passes the homogeneous test at the α=0.05 level. Therefore a combined effect size d_+ for this group is calculated with meaning. Only d_+ with * is tested for its significance later.

Among the seven data points in this category, four measured self-efficacy and three measured attitude change of the participants. In one study, the "software self-efficacy" was defined as the software learning ability developed as a result of hands-on experience gained during training. One study measured the "sense of efficacy" defined as the participants' (school principals in that study) expectations of successfully helping teachers become more effective in using cooperative learning through their supervising. Another study measured the "self-efficacy" of both

magnitude and strength for the ability of idea generation. The three data points measured the attitude which resulted from two studies. One study measured the attitude toward the training content—gifted education—while the other study measured the attitude toward the evaluation system of the organization and the importance of reducing defensive argument during active listening.

These analyses suggest it is easier to raise self-efficacy than change attitude through training. Thus, the effect size (the difference between the trained group and the non-trained control group) of self-efficacy measure tends to be higher than those of attitude. The subjective learning sub-data-set was sub-grouped by "what's measured"—self-efficacy and attitude.

As expected, this factor explained 68% of the heterogeneity of the subjective learning category (see Table 4.10). The between-group heterogeneity (Q_B=24.3) is significant at α=0.001 level (i.e., the possibility that the two subgroups are homogeneous is less than 0.1%). At the same time, the heterogeneity within each subgroup became small, and the "self-efficacy" subgroup passed the homogeneity test at the α=0.05 level. Therefore a meaningful average effect size of the "subjective learning—self-efficacy" subgroup could be drawn. The magnitude of the average effect size is 0.959 and its sampling error is 0.122. Though it was close, the attitude subgroup (Q_W=6.164) did not pass the homogeneity test at the α=0.05 level, so there was no meaningful average effect size for this subgroup.

Table 4.10

Effect of Management Training by "what's measured" on Subjective Learning

Source		Q value	Degrees of freedom	Test of homogeneity at α=0.05 level	d_+	S_{d+}
Between groups	Q_B	24.3	1	heterogeneous		
Within groups	Q_W	11.191	5	heterogeneous		
Within group 1 (attitude)	Q_{W1}	6.164	2	heterogeneous	0.224	0.084
Within group 2 (self-efficacy)	Q_{W2}	5.027*	3	passed the test	0.959*	0.122
Total	Q_T	35.491	6	heterogeneous	0.468	0.071

Note: Q value with * is smaller than the critical value of the chi-square distribution with the specific degrees of freedom at the α=0.05 level. This means that the group passes the homogeneous test at the α=0.05 level. Therefore a combined effect size d_+ for this group is calculated with meaning. Only d_+ with * is tested for its significance later.

Sub-grouping by training content and training method explained much less heterogeneity than the "what's measured" factor. Also the number of total data points was small, so the subjective learning group was not further divided by content and method.

Training Effect in the Objective

Learning Sub-Data-Set

The objective learning sub-data-set was larger and less heterogeneous than the subject learning category. With the previous experience, it was sub-grouped by "what's measured" factor first.

The majority of studies (67%) measured objective learning outcome through some kind of knowledge test. This is the traditional means to evaluate what is learned at the end of the program especially for cognitive learning tasks. Several programs which focused on interpersonal skills developed a knowledge test on learning points or managerial incidents. Two studies measured participants' leaning by task performance (e.g., complete a task with the learned software, or generate innovative ideas to soive a given problem). One cross-cultural training program and one leadership training program measured participants' learning in interpersonal skills through assessing their audio/video-taped role playing at the end of the program.

It was expected that the training effect of the knowledge test would be higher than that of the task performance and the role playing. The objective learning sub-data-set was sub-grouped by "knowledge test", "performing task" and "role playing".

As shown in Table 4.11, both "knowledge test" and "role playing" subgroups passed the homogeneity test at the α=0.05 level. Thus, a conclusion of a meaningful average effect size of the "objective learning—knowledge test" subgroup and of the "objective learning—role playing" subgroup were given. The magnitudes were 0.819 and 0.534, with a sampling error of 0.071 and 0.224 respectively. The larger magnitude of effect size and the smaller standard error of the knowledge test subgroup compared to the role playing subgroup could be explained by the different degree of difficulty of learning cognitive and behavioral skills. The "performing task" subgroup was still too heterogeneous to calculate an average effect size. Also, it was too small for further division.

Table 4.11

Effect of Management Training by "what's measured" on Objective Learning

Source		Q value	Degrees of freedom	Test of homogeneity at α=0.05 level	d_+	S_{d+}
Between groups	Q_B	1.573	2	not significantly heterogeneous		
Within groups	Q_W	27.959	15	heterogeneous		
Within group 1 (knowledge test)	Q_{W1}	16.177*	11	passed the test	0.819*	0.071
Within group 2 (performing task)	Q_{W2}	7.009	1	heterogeneous	0.841	0.161
Within group 3 (role playing)	Q_{W3}	4.773*	3	passed the test	0.534*	0.224
Total	Q_T	29.532	17	heterogeneous	0.801	0.063

Note: Q value with * is smaller than the critical value of the chi-square distribution with the specific degrees of freedom at the α=0.05 level. This means that the group passes the homogeneous test at the α=0.05 level. Therefore a combined effect size d_+ for this group is calculated with meaning. Only d_+ with * is tested for its significance later.

Training Effect of the Subjective

Behavior Sub-Data-Set

The subjective behavior sub-data-set was the most important category for this study since it measured the participants' on-the-job behavior. The whole sub-data-set was extremely heterogeneous with a 92.237 Q value at 23 degrees of freedom (α=0.001).

The majority (71%) of the data points measured participants' on-the-job performance by some kind of performance appraisal instruments. Most studies used multiple measures, i.e. measures by self, by superior, by subordinates, or by expert with the same instrument or a similar instrument. The variation of measurement has been significantly reduced by averaging the dependent effect sizes before entering them into the data set for meta-analysis. Instead of measuring job performance, three studies measured participants' behavior change after training. Two studies measured some kind of self-judgment about on-the-job behavior. For example, one stress management program measured the degree of participants' stress-reduction after returning to their job position. Another study measured the self-reported degree of difficulty of adjusting to a new culture.

Again, the "what's measured" factor was examined first. The sub-data-set was grouped into three subgroups—"behavior change", "performance appraisal" and "self- judgment". The results were shown in Table 4.12.

The "what's measured" factor was effective in explaining heterogeneity. The between-group heterogeneity was significant at the α=0.005 level (Q_B=12.577, *df*=2). However, the within-group heterogeneity of each subgroup remained high.

Since the "behavior change" and "self-judgment" subgroups were relatively small and were heterogeneous from the "performance appraisal" subgroup, they were put aside so the analysis could focus on the data points generated by job performance appraisal—the most commonly used method for evaluating job performance in work organizations.

Table 4.12

Effect of Management Training by "what's measured" on Subjective Behavior

Source		Q value	Degrees of freedom	Test of homogeneity at α=0.05 level	d_+	S_{d+}
Between groups	Q_B	12.577	2	heterogeneous		
Within groups	Q_W	79.66	21	heterogeneous		
within group 1 (behavior change)	Q_{W1}	10.338	2	heterogeneous	0.461	0.105
within group 2 (performance appraisal)	Q_{W2}	55.042	16	heterogeneous	0.439	0.063
within group 3 (self judgment)	Q_{W3}	14.28	3	heterogeneous	1.131	0.184
Total	Q_T	92.237	23	heterogeneous	0.503	0.055

Note: Q value with * is smaller than the critical value of the chi-square distribution with the specific degrees of freedom at the α=0.05 level. This means that the group passes the homogeneous test at the α=0.05 level. Therefore a combined effect size d_+ for this group is calculated with meaning. Only d_- with * is tested for its significance later.

Management is often defined as getting things done through people. It was not surprising that 45% of the study population (13 studies) focused on Human Relations/Leadership programs. The General Management Programs normally had an important component of interpersonal skills. The training objectives of these programs were enhancing participants' managerial ability by improving their interpersonal skills and/or leadership skills. The training effect on participants' job performance was measured by a performance appraisal instrument. Therefore this subgroup was the most representative subgroup of the management training outcome of job performance.

However, this representative subgroup was too heterogeneous to calculate an average effect size. When checking the Q_i values, data points from three studies were found to be extreme. After examining the primary studies, one highly effective program was found to have a rare situation: After being trained on a specific country's culture, all participants had a several months overseas mission to that country. This study was taken out of the data set. After this decision, 14 data points remained in the performance appraisal subgroup, which still failed to pass the homogeneity test (Q_T =25.806, *df*=13, significant at the α=0.025 level).

Further sub-grouping was performed. First, the representative sub-data-set was sub-grouped by training content. There were two content areas—general management programs and human relations/leadership programs. The latter was the larger one. The result was shown in Table 4.13.

The between-group heterogeneity was significant at the α=0.05 level (Q_B =4.351, *df*=1). It indicated that the two subgroups were heterogeneous (the dividing was meaningful and successful). The two subgroups—general management program and human relations/leadership program—both passed the homogeneity test at the α=0.05 level, and so gave meaningful average effect sizes. The "SB (performance appraisal)-general management program" subgroup had an average effect size of 0.167, with a standard error of 0.105. The "SB (performance appraisal)-human relations/leadership program" had an average effect size of 0.46, with a standard error of 0.095.

The representative sub-data-set was then sub-grouped by training method. There were three method categories—cognitive method mainly, behavioral method mainly,

Table 4.13

Effect of Management Training by Content on Subjective Behavior (performance appraisal)

Source		Q value	Degrees of freedom	Test of homogeneity at α=0.05 level	d_+	S_{d+}
Between groups	Q_B	4.351	1	heterogeneous		
Within groups	Q_W	21.455	12	heterogeneous		
within group 1 (general mgmt.)	Q_{W1}	4.024*	2	passed the test	0.167*	0.105
within group 2 (human relations/ leadership)	Q_{W2}	17.431*	10	passed the test	0.46*	0.095
Total	Q_T	25.806	13	heterogeneous	0.325	0.071

Note: Q value with * is smaller than the critical value of the chi-square distribution with the specific degrees of freedom at the α=0.05 level. This means that the group passes the homogeneous test at the α=0.05 level. Therefore a combined effect size d_+ for this group is calculated with meaning. Only d_+ with * is tested for its significance later.

and combination of cognitive and behavioral method. The last subgroup was the largest, which showed a trend of using both cognitive and behavioral methods in management training. The result of the analogue weighted ANOVA was shown in Table 4.14.

The between-group heterogeneity (Q_B=5.491, *df*=2) was significant at the α=0.1 level (i.e., the possibility of wrongly claiming the 3 subgroups are heterogeneous is less than 10%). After the sorting, all three subgroups passed the homogeneity test, so the average effect sizes were meaningful.

Table 4.14

Effect of Management Training by Method on Subjective Behavior (performance appraisal)

Source		Q value	Degrees of freedom	Test of homogeneity at α=0.05 level	d_+	S_{d+}
Between groups	Q_B	5.491	2	heterogeneous at α=0.10 level		
Within groups	Q_W	20.315	11	heterogeneous		
within group 1 (cognitive)	Q_{W1}	0.776*	1	passed the test	0.202*	0.214
within group 2 (behavioral)	Q_{W2}	5.694*	3	passed the test	0.182*	0.1
within group 3 (combined)	Q_{W3}	13.845*	7	passed the test	0.517*	0.11
Total	Q_T	25.806	13	heterogeneous	0.325	0.071

Note: Q value with * is smaller than the critical value of the chi-square distribution with the specific degrees of freedom at the α=0.05 level. This means that the group passes the homogeneous test at the α=0.05 level. Therefore a combined effect size d_+ for this group is calculated with meaning. Only d_+ with * is tested for its significance later.

For the "SB (performance appraisal)-cognitive method" subgroup, the average magnitude of training effect was 0.202 with a standard error of 0.214. The "SB (performance appraisal)-behavioral method" subgroup had an average effect size of 0.182, and its standard deviation was 0.1. The "SB (performance appraisal)-combined method" subgroup had the largest effect between the trained group and the control group. Its average effect size and standard error were 0.517, and 0.11 respectively.

Training Effect in the Subjective

Result Sub-Data-Set

This was the smallest sub-data-set by measurement criteria (5 data points). Among the three studies included, one measured employees' commitment to the organization as a result of manager's improved leadership skill; one study measured the employees' satisfaction with their performance appraisal discussion as a result of the managers' improved skill gained from the rater training program; and the third study measured the group effectiveness as a result of the managers' group leadership training.

As shown in Table 4.15, this small sub-data-set easily passed the homogeneity test at the α=0.05 level without further division (Q_T =4.193, df=4). So the average effect size of 0.245 for the "Subjective Result" category was meaningful. The standard error was 0.089.

Training Effect in the Objective

Result Sub-Data-Set

Training effect in the objective result category seemed the most appealing result to those who paid for training in work organizations. However, the sub-data-set was relatively small. Six studies that measured organizational results were collected and then dropped due to inadequate data to calculate effect size, or to non-managerial trainees. Among the five available studies, "what's measured" was very diversified: two measured job accuracy (appraisal form, negative incidents at work), one measured financial performance, and the rest measured productivity, turnover, and product quality. The sub-data-set was grouped into two subgroups—"accuracy" and "others".

Table 4.15

Effect of Management Training by "what's measured" on Subjective Result

Source		Q value	Degrees of freedom	Test of homogeneity at α=0.05 level	d_+	S_{d+}
Between groups	Q_B	0.765	1	not significantly heterogeneous		
Within groups	Q_W	3.428*	3	passed the test		
within group 1 (employee satisfaction or commitment)	Q_{W1}	3.355*	2	passed the test	0.215*	0.095
within group 2 (group effectiveness)	Q_{W2}	0.073*	1	passed the test	0.438*	0.237
Total	Q_T	4.193*	4	passed the test	0.245*	0.089

Note: Q value with * is smaller than the critical value of the chi-square distribution with the specific degrees of freedom at the α=0.05 level. This means that the group passes the homogeneous test at the α=0.05 level. Therefore a combined effect size d_+ for this group is calculated with meaning. Only d_+ with * is tested for its significance later.

As shown in Table 4.16, this category was heterogeneous as Q_T =12.957 and *df*=6 (significant at α=0.05 level). The dividing by "what's measured" was effective since the Q_B (10.496) explained 81% of the total heterogeneity. Two subgroups of "accuracy" and "others" both passed the homogeneity test. Due to the small size of data points and the diversified measurement, average effect sizes for these two subgroups must be viewed with caution: The "Objective Result (accuracy)" subgroup had an average effect size of 0.264 with a standard error of 0.105. The "Objective Result (others)" had an average effect size of 0.764 with a standard error of 0.114. Since the sample sizes of the two data points measured by financial indicators and by product

Table 4.16

Effect of Management Training by "what's measured" on Objective Result

Source		Q value	Degrees of freedom	Test of homogeneity at α=0.05 level	d_+	S_{d+}
Between groups	Q_B	10.496	1	heterogeneous		
Within groups	Q_W	2.461	5	passed the test		
within group 1 (accuracy)	Q_{W1}	1.349*	2	passed the test	0.264*	0.105
within group 2 (turnover & productivity)	Q_{W2}	1.112*	3	passed the test	0.764*	0.114
Total	Q_T	12.957	6	heterogeneous	0.487	0.077

Note: Q value with * is smaller than the critical value of the chi-square distribution with the specific degrees of freedom at the α=0.05 level. This means that the group passes the homogeneous test at the α=0.05 level. Therefore a combined effect size d_+ for this group is calculated with meaning. Only d_+ with * is tested for its significance later.

quality were several times smaller than that of the data points measured by turnover and productivity, the average effect size of the "others" subgroup was mainly determined by the data points of turnover and productivity. Thus, this subgroup was renamed as "OR (turnover & productivity).

Test of the Significance of the Average Effect Sizes

A total of 11meaningful effect sizes of groups or subgroups were obtained by rational sub-grouping and the analogue weighted ANOVA. They were all positive in magnitude (see Table 4.17). However, are they a real effect caused by management

Table 4.17

Test of Significance of the Average Effect Sizes

Category	Total No. of subjects	Average effect size d+	Sampling error Sd+	95% confidence interval for the population effect size δ	
				lower boundary	upper boundary
SL (self-efficacy)	301	0.959	0.122	0.72	1.198
OL (knowledge test)	970	0.819	0.071	0.68	0.958
OL (role playing)	85	0.534	0.224	0.095	0.973
SB (performance appraisal) General Management Program	396	0.167	0.105	-0.039	0.373
SB (performance appraisal) Human Relations/Leadership Program	480	0.46	0.095	0.274	0.646
SB (performance appraisal) Cognitive method mainly	95	0.202	0.214	-0.217	0.621
SB (performance appraisal) Behavioral method mainly	409	0.182	0.1	-0.014	0.378
SB (performance appraisal) Combination of cognitive and behavioral methods	372	0.517	0.11	0.301	0.733
SR	548	0.245	0.089	0.071	0.419
OR (accuracy)	392	0.264	0.105	0.058	0.47
OR (other)	324	0.764	0.114	0.541	0.987

training, or is the result just by chance (sampling error)? The test of the generally expressed null hypothesis of H_{01} was conducted to answer this question.

According to Hedges and Olkin (1985), the average effect size estimate is approximately a normal distribution. With the calculated average effect size $(d_+)_i$ and its standard error $(S_{d+})_i$, a 95% confidence interval for the true population effect size δ_i can be calculated. If the confidence interval does not cover zero, it means that the population training effect has a positive magnitude. The null hypothesis is rejected at the α=0.05 level. It is then concluded that the training makes a real difference on the specific measurement between the managerial personnel who receive training and those who do not received the training.

H_{01} : There is no difference in the training effect measured by X criterion of managerial personnel who receive management training that taught Y content (or that is taught by Z method) and those who do not receive the management training that taught Y content (or is taught by Z method).

The generally expressed null hypotheses H_{01} was specifically expressed for each of the 11 average effect sizes. The results of the testing of these 11 null hypotheses were summarized in Table 4.17, and described as follows:

H_{01} A: There is no difference in the training effect measured by self-efficacy (subjective learning criterion) of managerial personnel who receive management training and those who do not receive the management training.

Hypothesis H_{01} A is rejected at the significant level of α=0.001 level. It is concluded that there is significant difference in the training effect measured by

self-efficacy (subjective learning criterion) of managerial personnel who receive management training and those who do not receive the management training. The average effect is 0.959 with a standard error of 0.122. The total sample size to accumulate this result is 301, and the result is drawn across various training content areas and training methods.

H_{01} B: There is no difference in the training effect measured by knowledge test (objective learning criterion) of managerial personnel who receive management training and those who do not receive the management training.

Hypothesis H_{01} B is rejected at the significant level of α=0.001. It is concluded that there is significant difference in the training effect measured by knowledge test (objective learning criterion) of managerial personnel who receive management training and those who do not receive the management training. The average effect size is 0.819 with a standard error of 0.071. The total sample size to accumulate this result is 970, and the result is drawn across various training content areas and training methods.

H_{01} C: There is no difference in the training effect measured by role playing (objective learning criterion) of managerial personnel who receive management training and those who do not receive management training.

Hypothesis H_{01} C is rejected at the significant level of α=0.05. It is concluded that there is significant difference in the training effect measured by role playing (objective learning criterion) of managerial personnel who receive management training and those who do not receive management training. The average effect size is 0.534 with a standard error of 0.224. The total sample size to accumulate this result is 85, and

the result is drawn from human relations/leadership programs with various type of training methods.

H_{01} D: There is no difference in the training effect measured by performance appraisal (subjective behavior criterion) of managerial personnel who receive General Management training and those who do not receive General Management training.

It failed to reject hypothesis H_{01} D at the significant level of α=0.05. It is concluded that the difference (expressed by the average effect size 0.167) in the training effect measured by performance appraisal (subjective behavior criterion) of managerial personnel who receive General Management training and those who do not receive General Management training is very likely sampling error and other artifactual sources.

H_{01} E: There is no difference in the training effect measured by performance appraisal (subjective behavior criterion) of managerial personnel who receive Human Relations/Leadership training and those who do not receive Human Relations/Leadership training.

Hypothesis H_{01} E is rejected at the significant level of α=0.001. It is concluded that there is significant difference in the training effect measured by performance appraisal (subjective behavior criterion) of managerial personnel who receive Human Relations/Leadership training and those who do not receive Human Relations/Leadership training. The average effect size is 0.46 with a standard error of 0.095. The total sample size to accumulate this result is 480. Crossing various training methods, the result is drawn from the most popular content area of the management training which is aimed to improve interpersonal and/or leadership skills.

H_{01} F: There is no difference in the training effect measured by performance appraisal (subjective behavior criterion) of managerial personnel who receive management training that is taught mainly by cognitive methods and those who do not receive the management training that is taught mainly by cognitive methods.

It failed to reject H_{01} F at the significant level of α=0.05 level. The standard error is bigger than the average effect size. It is obviously that the difference in the training effect measured by performance appraisal (subjective behavior criterion) of managerial personnel who receive management training that is taught mainly by cognitive methods and those who do not receive the management training that is taught mainly by cognitive methods comes from sampling error and other artifactual sources. The total sample size to accumulate this result is 95. The result seems to support the common belief that cognitive methods are limited in training behavior.

H_{01} G: There is no difference in the training effect measured by performance appraisal (subjective behavior criterion) of managerial personnel who receive management training that is taught mainly by behavioral methods and those who do not receive the management training that is taught mainly by behavioral methods.

It failed to reject H_{01} G at the significant level of α=0.05. The existing evidence is not sufficient to prove that there is a real difference in the training effect measured by performance appraisal (subjective behavior criterion) of managerial personnel who receive management training that is taught mainly by behavioral methods and those who do not receive the management training that is taught mainly by behavioral methods. The average effect size is 0.182 with a standard error of 0.1. The total sample size to accumulate this result is 409.

H_{01} H: There is no difference in the training effect measured by performance appraisal (subjective behavior criterion) of managerial personnel who receive management training that is taught by combination of cognitive and behavioral methods and those who do not receive the management training that is taught by combination of cognitive and behavioral methods.

Hypothesis H_{01} H is rejected at the significant level of α=0.01. It is concluded that there is significant difference in the training effect measured by performance appraisal (subjective behavior criterion) of managerial personnel who receive management training that is taught by combination of cognitive and behavioral methods and those who do not receive the management training that is taught by combination of cognitive and behavioral methods. The average effect size is 0.517 with a standard error of 0.11. The total sample size to accumulate this result is 372. The result is supportive to the increasing trend of using combination of cognitive and behavioral methods in management training.

H_{01} I: There is no difference in the training effect measured by subjective result criterion of managerial personnel who receive management training and those who do not receive the management training.

Hypothesis H_{01} I is rejected at the significant level of α=0.002. It is concluded that there is significant difference in the training effect measured by subjective result criterion of managerial personnel who receive management training and those who do not receive the management training. The average effect size is 0.245 with a standard error of 0.089. The total sample size to accumulate the result is 548. Since the number

of studies (3) is small, and the diversity of content area and method is moderate, It should be cautious to make generalization from this result.

H_{01} J: There is no difference in the training effect measured by accuracy (objective result criterion) of managerial personnel who receive management training and those who do not receive the management training.

The hypothesis H_{01} J is rejected at the significant level of α=0.05. It is concluded that there is significant difference in the training effect measured by accuracy (objective result criterion) of managerial personnel who receive management training and those who do not receive the management training. The average effect size is 0.264 with a standard error of 0.105. The total sample size to accumulate this result is 392. This measure is usually used for technical training or rater training.

H_{01} K: There is no difference in the training effect measured by turnover and productivity (objective result criterion) of organizations where managerial personnel receive management training and those where managerial personnel do not receive the management training.

H_{01} K is rejected at the significant level of α=0.001. It is concluded that there is significant difference in the training effect measured by turnover and productivity (objective result criterion) of organizations where managerial personnel receive management training and those where managerial personnel do not receive the management training. The average effect size is 0.764 with a standard error of 0.114. The total sample size to accumulate the result is 324. The result is drawn mainly from supervisory training in business and industry.

Moderator Analysis

Research Question 3: Are the selected study characteristics—measurement criteria, training content, training method, training needs assessment, and favorable condition of transfer of training—moderator variables that mediate the magnitude of training effect of management training programs?

The grouping and sub-grouping process and analogue weighted ANOVA showed that evaluation criteria was an effective factor to explain heterogeneity. When the whole data set was divided by the five measurement criteria (see Table 4.9, the between-group heterogeneity Q_B had a value of 31.874 (df=4), which was significant at the level of α=0.001.

In each criterion group, "what's measured" factor further described the evaluation criterion and was a very effective way to divide data points into less heterogeneous sub-groups. Other factors, such as training content, training method and training needs assessment, could explain some heterogeneity within a specific evaluation criterion. The moderator analysis to detect the influence of them was conducted in two representative categories: Objective Learning (knowledge test), and Subjective Behavior (performance appraisal).

The tentative moderator variable of favorable condition of transfer of training (TOT) was found to be highly correlated with the moderator variable of training needs assessment (TNA). The total score of TNA of the data set was 61, and the total score of TOT of the data set was 67. The correlation coefficient of these two variables was 0.819. This was consistent with the coding process of these two variables in which scores of TOT often depended on scores of TNA. For example, a training program with

sound organization analysis for training needs was very likely to provide participants' opportunity to use what they learned from the training, and to have supervisor's support. A training program with individual analysis for training needs always increased the individual's motivation to apply the knowledge and skills they learned from training. This analysis pointed out that the influence of favorable condition of transfer of training to the training effect was not independent from the influence of the training needs assessment. Therefore, the following moderator analysis did not treat these two moderator variables separately.

Moderator Analysis on Objective Learning (Knowledge Test)

In the objective learning criterion, knowledge tests were the most commonly used evaluation means since they covered almost every training content area. The division by training method explained very little heterogeneity: Q_B=0.778, less than 5% of the Q_T. (see Table 4.18). The division by training needs assessment (TNA) did not explain much heterogeneity either (Q_B=2.165, about 13% of the Q_T) (see Table 4.19). Both of them failed to reject the homogeneous assumption across subgroups.

This result had a two-fold meaning. First, training method and training needs assessment were not moderator variables for training result from knowledge tests of managerial training. Second, the significant training effect (average effect size of 0.819) measured by knowledge test (objective learning criterion) of managerial personnel was gained from various training content areas and training methods. It confirmed the conclusion drawn from the test of hypothesis H_{01} B.

Table 4.18

Training Method as Moderator to Effect of Management Training (OL: Knowledge Test)

Source		Q value	Degrees of freedom	Test of homogeneity at α=0.05 level	d_-	S_{d+}
Between groups	Q_B	0.778	2	not significantly heterogeneous		
Within groups	Q_W	15.399*	9	passed the test		
within group 1 (cognitive)	Q_{W1}	9.842*	5	passed the test	0.848*	0.1
within group 2 (behavioral)	Q_{W2}	1.134*	2	passed the test	0.625*	0.235
within group 3 (combined)	Q_{W3}	4.423*	2	passed the test	0.826*	0.105
Total	Q_T	16.177*	11	passed the test	0.819*	0.071

Note: Q value with * is smaller than the critical value of the chi-square distribution with the specific degrees of freedom at the α=0.05 level. This means that the group passes the homogeneous test at the α=0.05 level. Therefore a combined effect size d_- for this group is calculated with meaning. Only d_+ with * is tested for its significance later.

Moderator Analysis on Subjective Behavior

(Performance Appraisal)

In the subjective behavior criterion, performance appraisal was the most commonly used means to evaluate participants' on-the-job improvement. The division by training content area explained 17% of the total heterogeneity. Q_B value was 4.351 (df=1) and significant at the α=0.05 level. It means that the two content subgroups were heterogeneous (see Table 4.13). The division by training method explains 21% of the heterogeneity. Q_B value was 5.491 (df=2) and significant at the α=0.1 level. The three

Table 4.19

TNA as Moderator to Effect of Management Training (OL: Knowledge Test)

Source		Q value	Degrees of freedom	Test of homogeneity at α=0.05 level	d_+	S_{d+}
Between groups	Q_B	2.165	1	not significantly heterogeneous		
Within groups	Q_W	14.012*	10	passed the test		
within group 1 (no TNA)	Q_{W1}	7.525	2	heterogeneous	1.058	0.176
within group 2 (moderate TNA)	Q_{W2}	6.487*	8	passed the test	0.776*	0.077
Total	Q_T	16.177*	11	passed the test	0.819*	0.071

Note: Q value with * is smaller than the critical value of the chi-square distribution with the specific degrees of freedom at the α=0.05 level. This means that the group passes the homogeneous test at the α=0.05 level. Therefore a combined effect size d_+ for this group is calculated with meaning. Only d_+ with * is tested for its significance later.

subgroups were heterogeneous (see Table 4.14). The division by training needs assessment (TNA) explained 22% of the heterogeneity. Q_B value was 5.757 (df=2) and significant at α=0.1 level (see Table 4.20).

Table 4.20

TNA as Moderator to Effect of Management Training (SB: Performance Appraisal)

Source		Q value	Degrees of freedom	Test of homogeneity at α=0.05 level	d_+	S_{d+}
Between groups	Q_B	5.757	2	heterogeneous at α=0.10 level		
Within groups	Q_W	20.049	11	heterogeneous		
Within group 1 (no TNA)	Q_{W1}	2.322*	3	passed the test	0.27*	0.173
Within group 2 (moderate TNA)	Q_{W2}	10.098*	5	passed the test	0.232*	0.089
Within group 3 (adequate TNA)	Q_{W3}	7.629*	3	passed the test	0.652*	0.155
Total	Q_T	25.806	13	heterogeneous	0.325	0.071

Note: Q value with * is smaller than the critical value of the chi-square distribution with the specific degrees of freedom at the α=0.05 level. It means that the group passes the homogeneous test at the α=0.05 level. Therefore a combined effect size d_+ for this group is calculated with meaning. Only d_+ with * is tested for its significance later.

Comparisons Among Average Effect Sizes

Across Groups/Subgroups

Each division, which was based on one moderator variable, resulted in a set of two or three average effect sizes. Each average effect size represented the training effect at a level of the moderator variable. The homogeneity tests have showed that the effect sizes are not homogeneous across subgroups. The generally expressed null hypothesis H_{02} is designed to test whether the difference between average effect sizes

$(d_{+})_i$ is significant. It was tested by means of contrasts of the average effect sizes for the subgroups (Hedges & Olkin, 1985, p. 159).

Hypothesis H_{02} : There is no difference in the magnitude of training effect among management training programs that are measured by X criteria (and that teach Y content, or that are taught by Z methods, or that have conducted training needs assessment, or that have favorable condition of transfer of training).

The generally expressed hypothesis was specifically expressed for each of the three moderator variables, and tested as follows:

H_{02} A: There is no difference in the magnitude of training effect measured by performance appraisal (Subjective Behavior criterion) among management training programs that teach General Management and those that teach Human Relation/Leadership. (see Table 4.21). The d_{+i} is approximately normally distributed with mean δ_i and variance $\sigma_{\infty}^{2}(d_{+i})$ (Hedges & Olkin, 1985, p.159). The 95% confidence interval of the contrast $(\delta_2 - \delta_1)$ is $0.005 < \delta_2 - \delta_1 < 0.561$

Hypothesis H_{02} A is rejected at the significant level of α=0.05. It is concluded that there is significant difference in the magnitude of training effect measured by performance appraisal (Subjective Behavior criterion) between management training programs that teach General Management and those that teach Human Relation/Leadership. The training content mediates the magnitude of effect size in this situation of management training programs' outcome is measured by performance appraisal.

Table 4.21

Training Content as Moderator to Effect of Management Training (SB: Performance Appraisal)

Moderator Variable	Training Content	Sample Size N	d_+	S_{d+}
Level 1	General Management Training	396	0.167	0.105
Level 2	Human Relations/Leadership	480	0.46	0.095

H_{02} B: There is no difference in the magnitude of training effect measured by performance appraisal (Subjective Behavior criterion) among management training programs that is taught mainly by cognitive methods, by behavioral methods, and taught by combination of cognitive and behavioral methods. (see Table 4.22).

Table 4.22

Training Method as Moderator to Effect of Management Training Measured by Performance Appraisal (SB)

Moderator Variable	Training Method	Sample Size N	d_+	S_{d+}
Level 1	Cognitive methods mainly	95	0.202	0.214
Level 2	Behavioral methods mainly	409	0.182	0.1
Level 3	Combination of cognitive and behavioral methods	372	0.517	0.11

The 95% confidence interval of the contrast $(\delta_3 - \delta_1)$ is $-0.157 < \delta_3 - \delta_1 < 0.787$. The 95% confidence interval of the contrast $(\delta_3 - \delta_2)$ is $0.043 < \delta_3 - \delta_1 < 0.627$ Hypothesis H_{02} B for the contrast of $(\delta_3 - \delta_2)$ is rejected at the significant level of α=0.05, while it failed to reject Hypothesis H_{02} B for the contrast of $(\delta_3 - \delta_1)$ at the significant level of α=0.05. It is obvious that there is no significant difference between δ_1 and δ_2.

It is concluded that there is significant difference in the magnitude of training effect measured by performance appraisal (Subjective Behavior criterion) between management training programs taught mainly by behavioral methods and those taught by combination of cognitive and behavioral methods. The difference in the magnitude of training effect between programs taught mainly by cognitive methods and those taught by combination of cognitive and behavioral methods are not systematic but random. The results are mixed.

H_{02} C: There is no difference in the magnitude of training effect measured by performance appraisal (Subjective Behavior criterion) among management training programs that do not conduct training needs assessment, those that conduct moderate training needs assessment, and those that conduct adequate training needs assessment. (see Table 4.23).

Table 4.23

TNA as Moderator to Effect of Management Training Measured by Performance Appraisal (SB)

Moderator Variable	Training Needs Assessment	Sample Size N	d_+	S_{d+}
Level 1	No TNA conducted or reported	148	0.27	0.173
Level 2	Moderate TNA conducted	538	0.232	0.089
Level 3	Adequate TNA conducted	190	0.652	0.155

The 95% confidence interval of the contrast $(\delta_3 - \delta_2)$ is $0.069 < \delta_3 - \delta_2 < 0.771$. The 95% confidence interval of the contrast $(\delta_3 - \delta_1)$ is $-0.073 < \delta_3 - \delta_1 < 0.837$. The results are mixed again. Hypothesis H_{02} C for the contrast of $(\delta_3 - \delta_2)$ is rejected at the significant level of α=0.05, while it failed to reject Hypothesis H_{02} C for the contrast of $(\delta_3 - \delta_1)$ at the same significant level.

Training needs assessment is important for training effect according to theory. If we fail to reject a false null hypothesis at the α=0.05 level, the possibility of making a type II error is high. According to Gay (1996, p. 474), when reducing type II error (i.e., to conclude that the TNA makes no difference while it really does) is more important than type I error such as in an exploratory study, the α level could be set up at 0.1. Then the 90% confidence interval of the contrast $(\delta_3 - \delta_1)$ becomes $0.0004 < \delta_3 - \delta_1 < 0.764$. The hypothesis H_{02} C for the contrast of $(\delta_3 - \delta_1)$ then is rejected at the significant level of α=0.1. It is concluded that training needs assessment does make difference on

the magnitude of training effect of management training programs measured by performance appraisal.

Summary

After rejection of outliers, meta-analysis procedures were applied to a data set of 61 independent data points obtained from 29 primary studies. An analogue weighted ANOVA developed and illustrated by Hedges and Olkin (1985, Chapter 7) was conducted. Data points were grouped by study characteristics, and the total heterogeneity Q_T, within-group heterogeneity Q_W and between-group heterogeneity Q_B were calculated similar to a normal ANOVA.

The homogeneity test was then conducted by compare the Q value with the critical value of the chi-square distribution at the given significant level (α=0.05 for this study). The Q_B value (with a degree of freedom equals to the number of groups minus 1) was expected to be large enough to reject the null hypothesis that all studies share a common effect size (the true effect size of the population). If so, the dividing of group or subgroup by a specific potential moderator variable was effective. The Q_W value (with a degree of freedom equaled to the number of data points in the group minus 1) was expected to be small enough to fail to reject the homogeneity null hypothesis. A meaningful average effect size only obtained from those groups or subgroups that passed the homogeneity test.

In order to get homogeneous groups and subgroups, the present study followed Burke and Day's (1986) right direction of grouping data points by measurement criteria, but went one step further. Within a group of a specific measurement criterion, the factor

of "what is measured" explains the heterogeneity much more effectively than training content and training method factors. As a result, 11 group and subgroups passed homogeneity test at the α=0.05 level. Therefore, a total of 11 average effect sizes and their standard errors were obtained. All average effect sizes are positive in magnitude.

Test of the null hypothesis H_{01} examined whether the average effect size d_{+i} is a real difference from training treatment or just a sampling error. Since d_{+i} is approximately normally distributed with mean δ_i and variance $\sigma_\infty^2(d_{+i})$, a 95% confidence interval of δ_i was constructed to perform this examination. If the interval does not cover zero, the average effect size is a real difference caused by the management training.

Test of the null hypothesis H_{02} examined whether the difference between the average effect sizes d_{+i} calculated from various levels of the moderator variable is a real difference or just a sampling error. A 95% confidence interval of $(\delta_1 - \delta_2)$ and other possible pairs of contrast was constructed to perform this examination. If the interval does not cover zero, the dividing factor is a significant moderator variable to the magnitude of training effect.

CHAPTER V

FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

Summary of Findings

Status of Empirical Evaluation Studies of Management Training (1983-1997)

When looking at the study population and the whole data set, rather than individual studies and data points, some patterns emerged. Business and industry conducted more management training programs than any other organization setting. Human Relations/Leadership is the largest focus content area in management training since interpersonal and leadership skills are most critical in enhancing managerial capacity. There was a trend to use multiple training techniques, especially to combine the cognitive methods and behavioral methods together, as most programs were designed to train managers in multiple skills/tasks.

Training professionals have made great efforts to enhance training effect by conducting training needs assessment and creating favorable conditions of transfer of training. At the same time, they have conducted training evaluation beyond the reaction and learning levels to demonstrate that training does have positive effects on job performance and organization results.

Training evaluation researchers have made progress as well during the past two decades. The evaluation studies were conducted in the work setting with real managerial personnel, rather than in college classrooms with students. Researchers exert control on variables to design and conduct true or quasi-experimental research in order to find cause-effect relationships between training intervention and trainees' learning, job performance and organization results.

In regard to the critique of Campbell et al. (1970) about depending "exclusively on the statistical significance as an indicator of judging success or failure of training", there is little response from the researchers of primary studies. Only two studies calculate effect size as an addition to normal statistical procedures. This problem is solved when the meta-analysis approach is applied to the data set, which is made up of primary studies.

Summary of Findings of Training Effects

There are 12 findings regarding the magnitude of training effect in terms of average effect size (an estimate of the true effect size of the population), and five findings regarding moderator variables drawn from this meta-analysis:

1. The effects of management training which are measured by various criteria are heterogeneous. They do not share a common true population effect size. So there is no overall effect size for management training from this meta-analysis. After grouping the data set into five measurement criteria, four of the groups (SL, OL, SB, OR) are still heterogeneous. Thus only the subjective result (SR) group passes the homogeneity test

and gets a group average effect size. All other average effect sizes obtained from this meta-analysis are calculated from homogeneous sub-groups.

2. There is significant (α=0.002) difference in the training effect measured by self-efficacy (subjective learning criterion) of managerial personnel who receive management training and those who do not receive management training. The average effect size is 0.959 with a sampling error of 0.122. This means that the training outcome measured by self-efficacy of the trained group is higher than the non-trained control group by about one standard deviation. In other words, the self-efficacy of a 50% percentile participant of the experimental group is equivalent to the self-efficacy of an 84% percentile person of the control group. The total sample size to accumulate this result is 301, and the result is drawn regardless of training content taught and training methods employed.

3. There is significant (α=0.001) difference in the training effect measured by knowledge test (objective learning criterion) of managerial personnel who receive management training and those who do not receive management training. The average effect size is 0.819 with a sampling error of 0.071. This means that the training outcome measured by knowledge test of the trained group is higher than the non-trained control group by more than 0.8 standard deviation. In other words, the knowledge test score of a 50% percentile participant of the experimental group is equivalent to the knowledge test score of an 80% percentile person of the control group. The total sample size to accumulate this result is 970, and the result is drawn regardless of training content taught and training methods employed.

4. There is significant (α=0.05) difference in the training effect measured by role playing (objective learning criterion) of managerial personnel who receive Human Relations/Leadership training and those who do not receive Human Relations/ Leadership training. The average effect size is 0.534 with a sampling error of 0.224. The total sample size to accumulate this result is 85. The result is drawn regardless of what type of training method is employed.

5. The average effect size for the training effect measured by performance appraisal of General Management Programs is 0.167 with a standard error of 0.105. However, it is not significant at the α=0.05 level.

6. The average effect size for the training effect measured by performance appraisal of Human Relations/Leadership Programs is 0.46 with a standard error of 0.095. It is significant at the α=0.001 level. The total sample size to accumulate this result is 480. This result is drawn from studies that employ all three types of training methods, and conduct adequate, moderate, or even no training needs assessment.

7. The average effect size for the training effects measured by performance appraisal of management training programs that are taught mainly by cognitive methods is 0.202 with a standard error of 0.214. It is not significant, but mainly due to sampling error.

8. The average effect size for the training effect measured by performance appraisal of management training programs that are taught mainly by behavioral methods is 0.182 with a standard error of 0.1. It is not significant at the α=0.05 level, but happens most likely by chance.

9. The average effect size for the training effect measured by performance appraisal of management training programs that are taught by the combination of cognitive and behavioral methods is 0.517 with a standard error of 0.11. It is significant at the α=0.001 level. The total sample size to accumulate this result is 372.

10. Management training programs are effective when measured by subjective result criterion. The average effect size is 0.245 with a standard error of 0.089. It is significant at the α=0.002 level. The total sample size to accumulate the result is 548. Since the number of studies (three) is small, and the diversity of content area and method is inadequate, more data are needed in the future to support more strongly the generalization of this result.

11. There is significant (α=0.05) difference in the training effect measured by accuracy (objective result criterion) of managerial personnel who receive management training and those who do not receive management training. The average effect size is 0.264 with a standard error of 0.105. The total sample size to accumulate this result is 392. This measure is usually used for technical training and rater training that are taught by either cognitive or combined methods.

12. There is significant (α=0.001) difference in the training effect measured by turnover and productivity (objective result criterion) of organizations where managerial personnel receive management training and those where managerial personnel do not receive management training. The average effect size is 0.764 with a standard error of 0.114. The total sample size to accumulate this result is 324. The result is drawn mainly from supervisory training in business and industry.

The five findings drawn from the moderator analysis are:

1. Training method and training needs assessment are not moderator variables for training result measured by knowledge tests of managerial training.

2. There is significant (α=0.05) difference in the magnitude of training effect measured by performance appraisal (subjective behavior criterion) between management training programs that teach General Management and those that teach Human Relations/Leadership. The training content mediates the magnitude of effect size in the situation in which management training programs' outcomes are measured by performance appraisal.

3. There is significant (α=0.05) difference in the magnitude of training effect measured by performance appraisal (subjective behavior criterion) between management training programs that are taught mainly by behavioral methods and those that are taught by a combination of cognitive and behavioral methods. The difference in the magnitude of training effect between programs taught mainly by cognitive methods and those taught by a combination of cognitive and behavioral methods is not systematic but random.

4. Training needs assessment does make a difference on the magnitude of training effect of management training programs measured by performance appraisal (at the α=0.1 level). The management training programs which conduct adequate training needs assessment have a much higher training effect than those that conduct moderate training needs assessment or no training needs assessment. The average effect size of the adequate TNA subgroup is 0.652 with a standard error of 0.155. It is significant at the α=0.001 level.

5. The tentative moderator variable of "favorable condition of transfer of training" is highly correlated with the moderator variable of training needs assessment (r=0.819). It indicates that the influence of favorable condition of transfer of training to the training effect is associated with the moderator variable of training needs assessment. They work together to ensure the significant training effect measured by performance appraisal of management training programs. The total score of TOT is seven more than TNA. It suggests that there is still room to improve the conditions of promoting transfer of training after the training needs assessment has been done, but not much.

Through synthesis, nine significant results of effects of management training on trainees' learning, job performance and organization results are obtained and listed in Table 5.1.

Conclusions from the Findings

Several conclusions regarding effects of management training on trainees' learning, job performance and organization results can be drawn from the above findings. They suggest what works and what does not work in management training.

Effects of Management Training on Trainees' Learning

Management training makes a real difference in trainees' learning in various outcomes measurements, such as self-reported self-efficacy and objectively evaluated knowledge tests and role-playing. Average effect sizes of 0.959 and 0.819 indicate that the differences of self-efficacy and knowledge test scores between managerial personnel

Table 5.1

Effects of Management Training on Trainees' Learning, Job Performance and Organization Results

Measurement Criteria	Measured By	Moderator Variable	Total Sample Size	Average Effect Size	Standard Error	Probability of a Type I Error
Trainees' learning						
Subjective	Self-efficacy	N/A	301	0.959	0.122	< 0.002
Objective	Knowledge Tests	N/A	970	0.819	0.071	< 0.001
	Role Playing	N/A	85	0.534	0.224	< 0.05
Trainees' job performance						
Subjective	Performance Appraisal	Content (Human Relations/ Leadership)	480	0.46	0.095	< 0.001
Subjective	Performance Appraisal	Methods (Combined)	372	0.517	0.11	< 0.001
Subjective	Performance Appraisal	TNA* (adequate)	190	0.652	0.155	< 0.001
Organization results						
Subjective	Employees' commitment, satisfaction	N/A	548	0.245	0.089	< 0.002
Objective	Accuracy	N/A	392	0.264	0.105	< 0.05
Objective	Turnover & Productivity	N/A	324	0.764	0.114	< 0.001

* Favorable condition of transfer of training (TOT) is correlated with Training needs assessment (TNA). The correlation coefficient is r=0.819

who have had and those who have not had the training program are clear. These meta-analysis findings provide empirical evidence for the old belief in training in disseminating knowledge, facts, and enhancing people's confidence and learning ability. However, the learning of behavior is more difficult than the learning of knowledge, so the difference of behaviors that are learned at the end of the training program between trained and non-trained managers is relatively small. These conclusions of trainees' learning are applicable to various content areas and training methods.

Effects of Management Training on Trainees' Job Performance

Management training can make a real difference in trainees' job performance in certain conditions. Regarding the content of training program, a Human Relations/Leadership program, which is focused on human relation problems of leadership, supervision, attitudes toward employees and communication, makes a real difference in trainees' job performance when their on-the-job behavior is measured by performance appraisal. A General Management program, which is the broadest type of development effort and includes managerial facts, concepts and skills, has a positive but small (average effect size of 0.167) influences on trainees' job performance. The evidence is not strong enough to allow us to say that the difference in trainees' job performance is a real one. Regarding training method, a management training program which is taught mainly by cognitive methods, or mainly by behavioral methods has positive but small influences on trainees' job performance. However, either type of method alone is insufficient for a real difference in trainees' job performance. When the

cognitive methods and behavioral methods are combined together to train interpersonal and leadership skills, managerial training makes a real difference in trainees' job performance. The instruments to measure the job performance are performance appraisal, which are widely used in many work organizations. Multiple measures from self, subordinates, supervisor and/or experts are averaged to reduce subjectivity and variance.

Effects of Management Training
on Organization Results

This meta-analysis draws several conclusions about training effect on organization results with caution due to the small number of primary studies available in this category. Employees will have a greater commitment to the organization and better job satisfaction if their supervisor improves interpersonal and leadership skills through management training. The difference is not large, but is significant. When the results are measured by objective standards, job accuracy increased and negative incidents decreased, turnover rate decreased and productivity increased.

Influence of Training Needs Assessment
on Training Effects

As a moderator variable, training needs assessment activities influence the magnitude of training effects. However, its influence on different training outcomes is not equal. Training needs assessment of a management training program has little influence on trainees' learning when the outcomes are measured by knowledge tests.

When the training objective is mainly to enrich participant's learning, moderate training needs assessment activities, which identify generic needs of participants as a specific group of managerial personnel, and ensure the match between the training contents and the audience, are sufficient. It is not necessary to conduct an in-depth individual analysis for a specific management-training program if its main objective is to disseminate facts and knowledge. The interaction part of the program, such as the question-and-answer session will meet some special individual needs. However, when the training objective is to improve managers' interpersonal and leadership skills, training needs assessment becomes a significant moderator variable to the magnitude of training effect on trainees' job performance. Identification of only generic needs of the organization and/or the process/task is not enough. To change a specific person's on-the-job behavior, his or her individual performance improvement needs must be addressed during the training. Adequate training needs assessment activities will increase the training effects on job performance by about 42% (the average effect size increases from 0.46 to 0.652). In practice, there are many creative ways to conduct training needs assessment activities. Some management training programs give managerial or leadership behavior assessment to participants before the training, and provide a lot of specific feedback during the training. Some programs build in a developmental process, such as on-the-job supervision or mentoring, action plan formulation and review. These are very effective ways to meet an individual's performance improvement needs through training.

Influence of Favorable Condition of Transfer of Training on Training Effects

The moderator variable of favorable condition of transfer of training is found to be highly correlated with training needs assessment activity. When performance improvement and training needs are properly identified, and management and participants' buy-in of the training is obtained, most of the favorable conditions of transfer of training are in place. Those programs that use managers as trainers and those programs that combine intensive training with on-the-job developmental process are especially effective in addressing the priority needs and in providing positive support to apply new knowledge and skills. The favorable condition of transfer of training is associated with the moderator variable of training needs assessment to influence the training effect in a positive way. These two factors work together to ensure the significant training effect measured by performance appraisal of management training programs.

Recommendations to Training Practitioners

Management training is effective. It can make a real difference in trainees' learning, job performance and organization results. The empirical evidence shows that the efforts of work organizations to enhance their management through training are important and meaningful. The conclusions from this meta-analysis, both significant and non-significant, give good advice to training professionals about "what works" and "what does not work" when they consider management training programs for their organization.

1. Management training is highly effective to enhance participant's self-efficacy, regardless of training contents and methods. Therefore training is a proper solution when the objective is to enhance trainees' confidence and ability to learn certain skills (e.g., computer software) or some new topics.

2. Management training is highly effective to increase participant's knowledge, not only for cognitive learning tasks but also for behavioral learning tasks, which can be tested by learning points or management incidents. Thus when the objective is to disseminate or teach people facts, new concepts and knowledge, or even certain behaviors, training is a good solution.

3. When the training objective focuses on improving managerial personnel's on-the-job performance, training professionals should select a Human Relations/Leadership program rather than a General Management program, since the former is designed to enhance interpersonal skills while the latter only devotes one component to it. To ensure the training effect on job performance, training needs assessment of the Human Relations/Leadership program, especially the individual analysis, should be conducted well, and the program should be taught by both cognitive and behavioral methods. Each of these efforts has been proven to have a significant influence on trainees' job performance. It would be powerful to combine them together. When situations and resources permit, a design of "intensive program plus distributed sessions" should be considered to integrate the training with trainees' on-the-job experience and to ensure the transfer of training.

4. Management training is effective when the outcomes are measured by organization result criterion although training is not the only contributor. When the

employees' morale is low and the turnover rate is high, the supervisor's behavior in dealing with human relations should be checked. If there are problems, a management training program aimed to improve superviors' interpersonal and leadership skills will help improve the situation.

5. Training professionals should pay more attention to training needs assessment (TNA) for three reasons: First, TNA itself mediates the training effect on job performance in a positive way. Second, the favorable condition of transfer of training is highly correlated with TNA. When a systematic TNA is conducted, many favorable conditions of transfer of training will take place as consequences. This relationship enlarges the importance of TNA. Third, TNA activities occur before and during the training program, and the transfer of training occurs after the program. Without TNA to identify the right place (where in the organization), right knowledge and skills (what content), right individual (who) of the program, and to gain management and participants' buy-in of the program, it is hard to resolve the situation later and to obtain enough favorable conditions for transfer of training. Although training needs assessment has a strong influence on the magnitude of training effects on job performance, it has little influence on the magnitude of training effects on trainees' learning. Therefore, the budget for training needs assessment should not be allocated equally to every training program.

Recommendations to Training Researchers

Great effort has been made in training evaluation research studies, and much progress has been achieved as summarized in Chapter IV. However, there are still many areas and aspects of evaluating the training effect which need more work.

1. The situation of depending exclusively on the statistical significance as an indicator of judging success or failure of training should and could be changed in a primary study. When experimental and quasi-experimental studies are conducted, means of experimental group and control group and their respective standard deviation are available to calculate the effect size of the training treatment. The significance of the magnitude of the effect size could also be tested. In addition to the mean difference, correlation coefficient r could serve as effect size measurement as well. This gives the primary study a second chance to show whether the individual program is a success or a failure since some non-significant t tests or F tests may turn out to be a relatively small but significant effect size.

2. The reporting of data of the primary study needs to be improved. Many journal articles do not publish enough raw data, especially when the result is non-significant. In fact, it is much easier to calculate the effect size by author(s) of the primary study than by the meta-analyst. If for any reason the primary study does not give the effect size, the editor of the journal should at least publish the group means and their standard deviation.

3. Compared to the measurement of learning outcomes and job performance outcomes, the area of organization results outcomes needs much work and research. The indicators and methods to determine their numerical value are highly diversified and

some are not sensible, especially those that try to show training contributions to the bottom-line. Before a meta-analysis can draw any reliable conclusions, some basic work needs to be done at the primary study level, e.g., to identify and prioritize the organization indicators which are most relevant to training, and to develop scientific methods to determine the magnitude of these indicators.

4. In this meta-analysis, only two representative subgroups (i.e., knowledge tests and performance appraisal) have a relatively large number of data points and provide the basis to conduct moderator analysis within a homogeneous sub-data-set. Some categories are out of the analytical process at an early stage due to their small number of data points. Some of these areas are important, such as trainees' attitude change (as a subjective learning measurement), and trainees' behavior change (as a subjective behavior measurement) which is different from job performance. More primary evaluation studies should be conducted in these areas to provide research findings for future accumulation.

5. The data set with a lot of coded study features contains rich information. It points out several interesting research areas that are worthwhile for further investigation. For example, what is the relationship between the training skills/tasks to the training method employed? What is a good match between these two features? How does the combination influence the magnitude of training effect?

6. Using meta-results as guidance, in-depth primary studies can be designed and conducted to investigate why such a phenomenon exists. Several research topics seem very appealing. First, the relationship of training needs assessment and the transfer of training is worthwhile for further investigation at the primary study level. What is the

mechanism and the process in which these two factors work together to influence the training effects? In practice what is the most efficient way to conduct training needs assessment and achieve favorable conditions of transfer of training? Second, training practitioners are using many creative and practical ways to identify training needs of various levels. Training needs assessment activities can be formal and/or informal, centralized and/or decentralized, and be conducted before and/or during the training. Based on the rich practice, training researchers need to expand their existing theories and models. Third, the function of the combined cognitive and behavioral methods in training managerial skills is an interesting research area. Why does the combination of cognitive methods and behavioral methods produce an effect on job performance stronger than the sum of the two types of methods alone? Last of this incomplete list is about the training design and implementation of the integration of intensive training and on-the-job developmental experience. Several primary studies of this meta-analysis employed an approach of "intensive program plus distributed sessions". This is a new paradigm in training design and implementation which needs researchers' attention.

Recommendations to a New Meta-Analyst

1. The present study followed Hedges et al. (1985, 1989) meta-analysis procedures. The 1989 book was written for social science researchers who have no strong statistic background. It is straightforward and easily understood. The first-time meta-analyst can start from this book. The 1985 book is a more theoretical and systematic explanation about their meta-analytical procedures.

2. The homogeneity test is critical for calculating meaningful average effect size for groups or subgroups since training outcomes in different measures are very different in nature. Even in the same measurement criterion, e.g., subjective learning, the magnitude of the measure of attitude and the measure of self-efficacy are significantly different. The attitude subgroup has an average effect size of 0.224 because attitude change is difficult. The self-efficacy subgroup has an average effect size of 0.958 because confidence and ability of learning the trained content can always be gained through training. It is obvious that the attitude subgroup and the self-efficacy subgroup do not share a common population effect size. The large Q_b value (24.3 at degrees of freedom of 1) confirmed the between-group heterogeneity. Without checking the homogeneity of the total subjective learning (SL) group, an average effect size of 0.468 for the SL category would be misleading. Unfortunately, not every meta-analyst agrees with and follows this sensible analysis. It is confusing to a beginner.

3. The test of null hypothesis after calculating a meaningful average effect size is indispensable. A meaningful average effect size is not necessarily significant in its magnitude. Several heterogeneous groups or subgroups divided by a moderator variable do not necessarily have average effect sizes significantly different from each other. Therefore, a conclusion about the training effect and the moderator variable can be drawn only after the null hypothesis is tested. Both significant results and non-significant results have practical implications.

REFERENCES

Abrami, P. C., Cohen, P. A., & d'Apollonia, S. (1988). Implementation problems in meta-analysis. Review of Educational Research, 58(2), 151-179.

Baldwin, T. T., & Padgett, M. Y. (1994). Management development: A review and commentary. In C. L. Cooper & I. T. Robertson (Eds.), Key reviews in managerial psychology: Concepts and research for practice (pp. 270-320). New York: Wiley.

Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. Personnel Psychology, 4, 63-105.

Bassi, L. J., & Van Buren, M. E. (1998). The 1998 state of the industry report. Training & Development, 22-43.

Bassi, L. J., Benson, G., & Cheney, S. (1996). The top ten trends. Training & Development, 50(11), 28-42.

Bates, R. A. (1997). The impact of training content validity, organizational commitment, learning, performance utility, and transfer climate on transfer of training in an industrial setting (Doctoral dissertation, Louisiana State University, 1997). Dissertation Abstracts International.

Benjamin, S. (1989). A closer look at needs analysis and needs assessment: Whatever happened to the systems approach? Performance & Instruction, 28(9), 12-16.

Bennett, W. R. Jr. (1995). A meta-analytic review of factors that influence the effectiveness of training in organizations (Doctoral dissertation, Texas A&M University, 1995). Dissertation Abstracts International.

Benson, G. (1996). How much do employers spend on training? Training & Development, 50 (10), 56-58.

Blakely, G. L., Martinec, C. L., & Lane, M. S. (1994). Management development programs: the effects of management level and corporate strategy. Human Resource Development Quarterly. 5, (1), 5-25.

Burke, M. J., & Day, R. R. (1986). A cumulative study of the effectiveness of managerial training. Journal of Applied Psychology, 71(2), 232-245.

Campbell, J. P. (1971). Personnel training and development. Annual Review of Psychology, 22, 565-598.

Campbell, J. P., Dunnette, M. D., Lawler, E. E. III, & Weick, K. E. Jr. (1970). Managerial Behavior, Performance, and Effectiveness. New York: McGraw-Hill.

Chen, L. (1994). Cross-cultural training effectiveness: A meta-analysis (Master's theses, Central Michigan University, 1994). Masters Abstract International.

Clement, R. W. (1981). Evaluating the effectiveness of management training: Progress during the 1970's and prospects for the 1980's. Human Resource Management, 8-13.

Curry, L. (1992). Management training and development. In C. A. Marvin (Ed. in chief) Encyclopedia of Educational Research (6th ed.) (Vol. 3). New York: Macmillan.

Dixon, N. M. (1990). The relationship between trainee responses on participant reaction forms and posttest scores. Human Resource Development Quarterly, 1(2), 129-137.

Dodge, R. E. (1987). Getting started with a needs assessment: A nuts and bolts approach for putting one together. Performance & Instruction, 26(6), 16-20.

Faerman, S. R., & Ban, C. (1993). Trainee satisfaction and training impact: Issues in training evaluation. Public Productivity and Management Review, 16(3), 299-314.

Gay, L. R. (1996). Educational research: Competencies for analysis and application (5th ed.) Upper Saddle River, NJ: Prentice-Hall.

Gegné, R. M., & Medsker, K. L. (1996). The conditions of learning: Training Applications. Fort Worth, TX: Harcourt Brace College Publishers.

Gist, M. E., Bavetta, A. G., & Stevens, C. K. (1990). Transfer training method: Its influence on skill generalization, skill repetition, and performance level. Personnel Psychology, 43, 501-523.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. Educational Researcher, 5(10), 3-8.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). Meta-analysis in social research. Beverly Hills, CA: SAGE.

Goldstein, I. L. (1980). Training in work organizations. Annual Review of Psychology, 31, 229-272.

Goldstein, I. L. (1991). Training in work organizations. In M. D. Dunnette, & L. M. Hough (Eds.), Handbook of industrial and organizational psychology (2nd ed.) (Vol. 2, pp. 507-619). Palo Alto, CA: Consulting Psychologists Press, Inc.

Gordon, E. E., Morgan, R. R., & Ponticell, J. A. (1994). FutureWork: The revolution reshaping American business. Westport, CT: Praeger.

Guba, E. G., & Lincoln, Y. S. (1981). Effective evaluation. San Francisco: Jossey-Bass.

Harris, D. M., & DeSimone, R. L. (1994). Human resource development. Orlando: Dryden Press.

Harrison, J. K. (1992). Individual and combined effects of behavior modeling and the cultural assimilator in cross-cultural management training. Journal of Applied Psychology, 77(6), 952-962.

Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis. Orlando, FL: Academic Press.

Hedges, L. V., Shymansky, J. A., & Woodworth, G. (1989). A practical guide to modern methods of meta-analysis. Washington, DC: National Science Teachers Association.

Hilbert, J., Preskill, H., & Russ-Eft, D. (1997). Evaluating training. In D Russ-Eft, & L. J. Bassi (Eds.), What works: assessment, development, and measurement. (pp. 109-150). Alexandria, VA: American Society of Training and Development.

Holton, E. F., III. (1995). A snapshot of needs assessment. In J. J. Phillips, & E. F. Holton, III (Eds.), In Action: Conducting Needs Assessment (pp. 1-12). Alexandria: American Society for Training and Development.

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

Hunt, M. (1997). How science takes stock: The story of meta-analysis. New York: Russell Sage Foundation.

Hunter, J. E., Schmidt, F. L. & Jachson, G. B. (1982). Meta-analysis: Cumulation research findings across studies. Beverly Hills, CA: SAGE.

Hunter, J. E., & Schmidt, F. L. (1990). Methods of meta-analysis: Correcting error and bias in research findings. Newbury Park, CA: SAGE.

Katz, R. L. (1974). Skills of an effective administrator. Harvard Business Review, 52(1), 90-102.

Katzell, R. A. (1989). Foreword. In I. L. Goldstein, Training and development in organizations. San Francisco: Jossey-Bass.

Keys, B., & Wolfe, J. (1988). Management education and development: current issues and emerging trends. Journal of Management, 14(2), 205-229.

Kirkpatrick, D. L. (1959). Techniques for evaluating programs. Journal of the American Society of Training Directors, 13(11), 3-9.

Kirkpatrick, D. L. (1994). Evaluating training programs: the four levels. San Francisco: Berrett-Koehler.

Lai, L. C. (1996). A meta-analysis of research on the effectiveness of leadership training programs. (Doctoral dissertation, University of Wisconsin-Madison, 1996). Dissertation Abstract International.

Latham, G. P. (1988). Human resource training and development. Annual Review of Psychology, 39, 545-582.

Lewis, T., & Bjorkquist, D. C. (1992). Needs assessment—A critical reappraisal. Performance Improvement Quarterly, 5(4), 33-54.

McGehee, W. and Thayer, P. W. (1961). Training in business and industry. New York: John Wiley & Sons, Inc.

McLagan, P.A., & Bedrick, D. (1983). Models for excellence: The results of the ASTD Training and Development Competency Study. Training and Development Journal, 37(6), 49-59.

McLagan, P. A., & Suhadolnik, D. (1989). Models for HRD practice: The research report. Alexandria: American Society for Training and Development.

Miller, V. A. (1996). The history of training. In R. L. Craig (Ed.). The ASTD training and development handbook (4th ed.). New York: McGraw-Hill.

Moore, M. L., & Dutton, P. (1978). Training needs analysis: Review and critique. Academy of Management Review, 3(3), 532-545.

Morrow, C. C., Jarrett, M. Q., & Rupinski, M. T. (1997). An investigation of the effect and economic utility of corporate-wide training. Personnel Psychology, 50, 91-119.

Mullen & Rosenthal. (1985). Basic Meta-analysis: procedures and programs. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Nadler, L. (1990). Human Resource Development. In L. Nadler and Z. Nadler (Eds.), Handbook of human resource development (2nd ed.), (pp. 1.1-1.47). New York: John Wiley & Sons.

Nelson, D. L., & Quick, J. C. (1996). Organizational behavior: The essentials. St. Paul: West Publishing Company.

Noe, R. A., & Schmitt, N. (1986). The influence of trainee attitudes on training effectiveness: Test of a model. Personnel Psychology, 39, 497-523.

Phillips, A. D. (1990). Taking a good look at development. Issues and Observations, 10(3), 1-5.

Phillips, J. J., & Rothwell, W. J. (1997). In action: Linking HRD programs with organizational strategy. Alexandria, VA: American Society for Training & Development.

Phillips, J. J., & Holton, E. F., III. (1995). In action: Conducting needs assessment. Alexandria: American Society for Training and Development.

Rosenthal, R. (1991). Meta-analytic procedures for social research (# 6). In Series: Applied Social Research Methods. Newbury Park, CA: SAGE.

Rothwell, W. J. (1984). Strategic curriculum design for management training. Journal of Management Development, 3(3), 39-52.

Rouiller, J. Z. , Goldstein, I. L. (1991). Determinants of the climate for transfer of training. Presented at the Meeting of Social and Industry Organization Psychology.

Rouiller, J. Z., & Goldstein, I. L. (1993). The relationship between organizational transfer climate and positive transfer of training. Human Resource Development Quarterly, 4(4), 377-390.

Rummler, G. A., & Brache, A. P. (1990). Improving performance: How to manage the white space on the organization chart. San Francisco: Jossey-Bass.

Rummler, G. A., & Brache, A. P. (1995). Improving performance: How to manage the white space on the organization chart (2nd ed.). San Francisco: Jossey-Bass.

Russ-Eft, D. (1997). Behavior modeling. In L. J. Bassi, & D. Russ-Eft (Eds.). What works: Training and development practices. Alexandria, VA: ASTD.

Saari, L. M., Johnson, T. R., Mclaughlin, S. D., & Zimmerle, D. M. (1988). A survey of management training and education practices in U.S. companies. Personnel Psychology, 41, 731-743.

Schmidt, W. H., Lippitt, G. L, & Johnson, R. B. (1971). Management training in industry. In L. C. Deighton (Ed.). The Encyclopedia of Education (Vol. 6). The Macmillan Company & The Free Press.

Sleezer, C. M. (1991). Developing and validating the performance analysis for training model. Human Resource Development Quarterly, 2(4), 355-372.

Sleezer, C. M. (1992). Needs assessment: Perspectives from the literature. Performance Improvement Quarterly, 3(2), 34-46.

Sleezer, C. M. (1993). Training needs assessment at work: A dynamic process. Human Resource Development Quarterly, 4(3), 247-264.

Swanson, R. A., & Torraco, R. J. (1995). The history of technical training. In L. Kelly (Ed.), Technical and skills training handbook of the American Society for Training and Development (pp. 1-47). New York: McGraw-Hill.

Swanson, R. A. (1994). Analysis for improving performance: Tools for diagnosing organizations & documenting workplace expertise. San Francisco: Berrett-Koehler.

Tannenbaun, S. I., & Yukl, G. (1992). Training and development in work organizations. Annual Review of Psychology, 43, 399-441.

Tziner, A., Haccoun, R. R., & Kadish, A. (1991). Personal and situational characteristics influencing the effectiveness of transfer of training improvement strategies. Journal of Occupational Psychology, 64, 167-177.

Wexley, K. N. (1984). Personnel training. Annual Review of Psychology, (35), 519-551.

Wexley, K. N., & Baldwin, T. T. (1986). Management Development. 1986 yearly review of management of the Journal of Management, 12(2), 277-294.

Wexley, K. N., & Baldwin, T. T. (1986). Posttraining strategies for facilitating positive transfer: An empirical exploration. Academy of Management Journal, 29(3), 503-520.

Wolf, F. M. (1986). Meta-analysis: Quantitative methods for research synthesis (# 59). In Series: Quantitative applications in the social sciences. Newbury Park, CA: SAGE.

Zemke, R. E. (1994). Training needs assessment: The broadening focus of a simple concept. In A. Howard (Ed.), Diagnosis for organizational change: Method and models (pp. 139-151). New York: Guilford.

APPENDIXES

APPENDIX A

CODING SHEET

Coding Sheet of Research Studies of Management Training

Reference:

_____
_____
_____
_____

Coder Name _____ Study ID _____

Please circle one choice. Give description when necessary and possible, and try to be specific.

_____

A. Publication Information

a. Year of Publication _____

b. Type of Publication
- 1 = Journal articles
- 2 = Book/chapter of book
- 3 = Doctoral dissertation
- 4 = Masters theses
- 5 = Paper presented at conference
- 6 = Government document
- 7 = Technical report
- 8 = Unpublished manuscript

B. Subjects (sample) Characteristics

a. Type of Organization
- 1 = Business and Industry
- 2 = Military
- 3 = Government
- 4 = Public Sector
- 5 = Education
- 6 = Others

b. Job Position Classification
- 1 = Supervisor
- 2 = Manager
- 3 = Executive

c. Sample size in posttest

experimental group	group1( )	group2 ( )	group3 ( )
control group	group1( )	group2 ( )	group3 ( )
total sample size	experiment ( )	control ( )	total ( )

d. Trainee Characteristics (if reported in the study, please list below)

Age
Gender
Education
Work experience ( ) Management experience ( )
Race

Pretest/posttest

Personality

Learning style

Motivation

Self-efficacy

Cognitive ability

Psychomotor ability

General aptitude

Locus-of-control

Other (specify)

C. Training Program (Treatment) Characteristics

a. Treatment Description

_____
_____
_____
_____
_____
_____

b. Training Contents

1 = general management programs
2 = human relations/leadership programs
3 = self-awareness programs
4 = problem-solving/decision-making programs
5 = rater training programs
6 = motivation/values training programs
7 = technical skills training programs
8 = entrepreneurial skill
9 = ethical decision making
10 = others

c. Training Skill/Task Category

1 = Cognitive
2 = Interpersonal
3 = Psychomotor
4 = Multiple

d. Training Method (could check more than one method)

Tradition (classroom lecture, discussion)	( )
Case study	( )
Role playing	( )
Exercises	( )
Behavior-modeling	( )
Computer-supported-training	( )
Sensitivity training	( )
Leader match	( )
Outdoor and Field Study	( )
Others	( )

Convert to one of the three categories of the variable

1 = cognitive methods
2 = behavioral methods (experiential)
3 = combine cognitive and behavioral methods

e. Training Duration and Total Training Time

1 = short (< 1 day)
2 = intensive (one to several days)
3 = intensive program plus distributed sessions
4 =regular program lasting several weeks/months

f. Course Development

1 = Internal developed
2 = External customized design
3 = External on-the-shelf course

g. Course Delivery

1 = Internal instructor or manager/expert
2 = External consultant/professor
3 = External researcher of the study

h. Training Needs Assessment

	YES(score 1)	NO (score 0)
Organizational analysis	( )	( )
Process/task analysis	( )	( )
Individual analysis	( )	( )
or		
Explicit training objectives related to job performance improvement	( )	( )
Link training to organizational strategy and goals	( )	( )
subtotal score	( )	

Convert to ordinal Measurement

0 = No TNA conducted or reported (score 0)
1 = Moderate TNA was conducted or reported (score 1-2)
2 = Adequate TNA was conducted or reported (score 3 and more)

(Explicit training objectives related to job performance improvement, or efforts to link training to organizational strategy and goals are counted as one level TNA).

i. Favorable condition of transfer of training

	YES (score 1)	NO (score 0)
Transfer motivation	( )	( )
Opportunity to use	( )	( )
Peer support	( )	( )
Supervisor support	( )	( )
Positive personal outcomes	( )	( )
Others (if negative, score –1)		
Subtotal score	( )	

Convert to ordinal measurement:

0 = No favorable condition of transfer of training (score 0)
1 = Moderate favorable condition of transfer of training (score 1 to 2)
2 = Strong favorable condition of transfer of training (score 3 and more)

D. Research Design Characteristics

a. Research Location
1 = Laboratory
2 = Work setting

b. Research Design
0 = quasi-experimental
1 = experimental

c. Relative comparison
0 = without alternative treatment
1 = with alternative treatment

d. Sampling method
0 = not random sampling
1 = random sampling

e. Sample size N>=30
0 = N<30
1 = N>= 30

f. Assignment of subjects to treatment
0 = not random assignment
1 = random assignment

g. Control group characteristics
0 = no training
1 = traditional method of training

h. Control of pre-existing differences
0 = no control
1 = pre-test
2 = matched groups
3 = other

i. Rigor of the study

E. Outcome Measurement Characteristics

a. Timing of measure

0 = immediate after training
1 = follow-up

b. Type of criterion (constructs) measured (please describe what specific indicator is measured by what instrument)

1 = subjective learning
2 = objective learning
3 = subjective behavior
4 = objective behavior
5 = subjective results
6 = objective results (please indicate level of results: team/subunit, or whole organization)

Detailed Information (what measured and instrument)

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

c. Source of data

1 = self-reported
2 = subjective measure by others (supervisor, peers evaluation)
3 = subjective perceived results
4 = measured by objective criteria

F. Statistics for computing effect size (see the printout of studies.xls)

Group mean and standard deviation, sample size, or other statistics

G. Calculation of Original Data (see the printout of studies.xls)

Effect size
Standard error of the effect size

APPENDIX B

STUDIES IN META-ANALYSIS

Research Studies Included in the Meta-analysis

Bankston, J. R. (1993). Instructional leadership behaviors of a selected group of principals in Northeast Texas (Doctoral dissertation, East Texas State University, 1993). Dissertation Abstract International.

Barling, J., Weber, T., & Kelloway, E. K. (1996). Effects of transformational leadership training on attitudinal and financial outcomes: A field experiment. Journal of Applied Psychology, 81(6), 827-832.

Birkenbach, X. C., Kamfer, L., & ter Morshuizen, J. D. (1985). The development and evaluation of a behavior-modeling training programme for supervisors. South African Journal of Psychology,15 (1),11-19.

Briddell, W. B. (1986). The effects of a time management training program upon occupational stress levels and the type a behavioral pattern in college administrators (Doctoral dissertation, The Florida State University, 1986). Dissertation Abstract International.

Davis, B. L., & Mount, M. K. (1984). Effectiveness of performance appraisal training using computer assisted instruction and behavior modeling. Personnel Psychology, 37(3), 439-452.

Earley, P. C. (1987). Intercultural training for managers: A comparison of documentary and interpersonal methods. Academy of Management Journal, 30, 685-698.

Edwards, W. S. (1992). Training secondary school administrators for crisis management (Doctoral dissertation, University of Georgia, 1992). Dissertation Abstract International.

Engelbrecht, A. S., & Ficher, A. H. (1995). The managerial performance implications of a developmental assessment center process. Human Relations, 48(4), 387-405.

Esposito, C. (1988). Effects of a national association academy on leadership behavior and management practices (Doctoral dissertation, The University of Arizona, 1988). Dissertation Abstract International.

Feldhusen, J. F., Haeger, W. W., & Pellegrino, A. S. (1989). A model training program in gifted education for school administrators. Roeper Review, 11(4), 209-214.

Frost, D. E. (1986). A test of situational engineering for training leaders. Psychological Reports, 59(2), 771-782.

Gist, M. E. (1989). The influence of training method on self-efficacy and idea generation among managers. Personnel Psychology, 42(4), 787-805.

Gist, M. E., Schwoerer, C, & Rosen, B. (1989). Effects of alternative training methods on self-efficacy and performance in computer software training. Journal of Applied Psychology, 74, 884-891.

Harrison, J. K. (1992). Individual and combined effects of behavior modeling and the cultural assimilator in cross-cultural management training. Journal of Applied Psychology, 77(6), 952-962.

Jones, R. G., & Whitmore, M. D. (1995). Evaluating developmental assessment centers as interventions. Personnel Psychology, 48(3), 377-388.

Larsen, J. E. (1983). An evaluative study of the effect of self-management training on a group leadership training program for hospital managers (Doctoral dissertation, The University of Nebraska - Lincoln, 1983). Dissertation Abstract International.

Mathieu, J. E., & Leonard, R. L. (1987). Applying utility concepts to a training program in supervisory skills: A time-based approach. Academy of Management Journal, 30(2), 316-335.

Matton, R. J. (1985). Self, subordinate, and supervisor reports on the long-term effects of a management training program (Doctoral dissertation, East Texas State University, 1985). Dissertation Abstract International.

May, J. M., & Keys, C. B. (1986). Outcomes of a management development seminar for managers in the human services. Paper presented at the Annual Meeting of the Midwestern Psychological Association. Chicago, IL.

Nelson, E. P. (1990). The impact of the Springfield development program on principals' administrative behavior (Doctoral dissertation, University of Minnesota, 1990). Dissertation Abstract International.

Niska, J. M. (1991). Examination of a cooperative learning supervision training and development model (Doctoral dissertation, Iowa State University, 1991). Dissertation Abstract International.

Russell, J. S., Wexley, K. N., & Hunter, J. E. (1984). Questioning the effectiveness of behavior modeling training in an industrial setting. Personnel Psychology, 37(3), 465-481.

Schwier, R. A., & Misanchuk, E. R. (1988). The effect of interaction and perceived need for training on learning and time spent learning from computer-based instruction. Canadian Journal of Educational Communication, 17(3), 147-158.

Smith, J. F., Boshoff, A. B.; & de V. Vissor, M. (1983). Evaluation of a training programme for production supervisors based on the critical incident method. Journal for Research in the Human Sciences, 9(2), 189-193.

Smith, R. M.; White, P. E., & P. A. (1992). Investigation of interpersonal management training for educational administrators. Journal of Educational Research, 85(4), 242-245.

Sniderman, R. L. (1992). The use of SYMLOG in the evaluation of the effectiveness of a management development program (Doctoral dissertation, California School of Professional Psychology -- Los Angeles, 1992). Dissertation Abstract International.

Urban, T. F., Ferris, G. R., Crowe, D. F., & Miller, R. L. (1985). Management training: Justify costs or say goodbye. Training and Development Journal, 39(3), 68-71.

Whitford, E. V. (1990). The effects of cooperative learning in a staff development program for school administrators. Paper presented to the Annual Meeting of the International Association for the Study of Cooperation in Education. Baltimore, MD, July 5-11, 1990.

Young, D. P., & Dixon, N. M. (1996). Helping leaders take effective action: A program evaluation. Greensboro, NC: Center for Creative Leadership.

APPENDIX C

INSTITUTIONAL REVIEW BOARD (IRB)

APPROVAL FORM

OKLAHOMA STATE UNIVERSITY
INSTITUTIONAL REVIEW BOARD

DATE: 12-09-98 **IRB #: ED-99-068**

Proposal Title: EFFECT OF MANAGEMENT TRAINING ON TRAINEE JOB PERFORMANCE AND ORGANIZATION RESULTS: A META-ANALYSIS OF EVALUATION STUDIES FROM 1983-1997

Principal Investigator(s): Martin Burlingame, Jiping Zhang

Reviewed and Processed as: Exempt

Approval Status Recommended by Reviewer(s): Approved

Signature:  Date: December 10, 1998

Carol Olson, Director of University Research Compliance
cc: Jiping Zhang

Approvals are valid for one calendar year, after which time a request for continuation must be submitted. Any modification to the research project approved by the IRB must be submitted for approval. Approved projects are subject to monitoring by the IRB. Expedited and exempt projects may be reviewed by the full Institutional Review Board.

VITA

Jiping Zhang

Doctor of Education

Thesis: EFFECTS OF MANAGEMENT TRAINING ON TRAINEES' LEARNING, JOB PERFORMANCE AND ORGANIZATION RESULTS: A META-ANALYSIS OF EVALUATION STUDIES FROM 1983-1997

Major Field: Occupational and Adult Education

Biographical:

Personal Data: Born on March 11, 1946 in Beijing, China, the daughter of Beichen Zhang and Zhihua Xia.

Education: Graduated from the First Middle School Associated with the Beijing Normal University in July 1964; received Bachelor of Science degree (equivalent) from the Architecture and Civil Engineering Department of Tsinghua University in March 1970; received Master of Science degree from the same department of Tsinghua University in January 1981. Completed the requirements for the Doctor of Education degree at Oklahoma State University in May, 1999.

Experience: Consultant for staff training and distance learning of the Rural Development Department at the World Bank (1997-present); Consultant for various executive and government official training programs jointly sponsored by Chinese Government, North American universities and international organizations since 1990. Major clients include the World Bank, United Nations Development Programme, University of Western Ontario (Canada), Oklahoma State University; Senior Economist and Operational Officer of China Investment Bank (1988-1990); Faculty member of the Economic Management School at Tsinghua University from 1981 to 1988 and visiting scholar to Richard Ivey School of Business at the University of Western Ontario in 1986-1987; Construction worker, estimator and teacher at the company training center, 1970-1988.

Professional Memberships: KAPPA DELTA PI, Academy of Human Resource Development (AHRD)