4-1987

Effects of Training on Measures of Productivity: A Meta-Analysis of the Findings of Forty-Eight Experiments

Alice Susan Leddick
Western Michigan University

EFFECTS OF TRAINING ON MEASURES OF PRODUCTIVITY: A META-ANALYSIS OF THE FINDINGS OF FORTY-EIGHT EXPERIMENTS

by

Alice Susan Leddick

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Education
Department of Educational Leadership

Western Michigan University
Kalamazoo, Michigan
April 1987

EFFECTS OF TRAINING ON MEASURES OF PRODUCTIVITY: A META-ANALYSIS OF THE FINDINGS OF FORTY-EIGHT EXPERIMENTS

Alice Susan Leddick, Ed.D.

Western Michigan University, 1987

The objective of this study was to determine the effects of training on three employee productivity measures: output, withdrawal, and disruption.

A meta-analysis of forty-eight field experiments published in the United States between 1971 and 1981, inclusive, was conducted to explore differences in training effectiveness associated with eight substantive training variables (five design variables and three contextual variables) and two study design variables.

Across all studies, training programs contributed to overall improvement of experimental groups over control groups of 0.67 standard deviation. Training interventions were most successful applied to disruption; however, a shortage of data cast doubt on the reliability of this finding. Mean effect size for the output criterion was a near-match for the overall mean effect size; mean effect size for withdrawal was a tenth of a standard deviation less.

The training design parameters of information processing method, knowledge objective, external trainer,

executive training content, and line manager involvement in design were associated with strongest effect sizes in the subject studies. Contextual variables associated with strongest effect sizes included small (fewer than 100 employees) organizations, private for-profit organizations, and managers as the training recipients. In all cases, consistency of measures did not necessarily follow strength. Strongest mean effects were associated with the design variables of longer elapsed time between training end and criterion measurement and comparison groups receiving other training. Missing data in the original reports substantially reduced the number of effect sizes available for analysis.

Training was shown to have been a useful intervention in productivity improvement efforts, more successfully applied to certain criteria than others. Further research was recommended to give further guidance to training professionals.

8714649

Leddick, Alice Susan

EFFECTS OF TRAINING ON MEASURES OF PRODUCTIVITY: A META-ANALYSIS OF THE FINDINGS OF FORTY-EIGHT EXPERIMENTS

Western Michigan University ED.D. 1987

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark __√__.

1. Glossy photographs or pages _____
2. Colored illustrations, paper or print ______
3. Photographs with dark background ____
4. Illustrations are poor copy ______
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page ______
7. Indistinct, broken or small print on several pages __√__
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine ______
10. Computer printout pages with indistinct print ______
11. Page(s) __________ lacking when material received, and not available from school or author.
12. Page(s) __________ seem to be missing in numbering only as text follows.
13. Two pages numbered ______. Text follows.
14. Curling and wrinkled pages _____
15. Dissertation contains pages with print at a slant, filmed as received __√__
16. Other ____

University
Microfilms
International

TABLE OF CONTENTS

Table of Contents--Continued

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

According to many observers of working America, one of the most pressing problems of human performance in American business and industrial organizations is productivity. Productivity growth is important to profitability and, ultimately, survival of American companies as they compete in a world market (Kendrick, 1984). Consequently, improving productivity has become an issue of considerable interest among practicing managers and behavioral scientists. Some 1200 books and articles concerning productivity were published between 1976 and 1978, alone (Brand & Belitsky, 1980). Although diverse methods for improving productivity exist, one of the most common is training and instruction (Guzzo & Bondy, 1983; Katzell, Bienstock, & Faerstein, 1977). This dissertation is an exploration of the effects of training in 48 productivity improvement experiments published in America between 1971 and 1981, inclusive. Specifically, the study is a meta-analysis of effects of training variables and organization contingencies on three composite productivity criteria or outcomes.

Need for the Study

Field study findings are important sources of information on the effects of training and other interventions for improving productivity. Synthesis of these findings, however, is needed if patterns of meaning are to be derived to guide practice. Bass (1982), for example, called for further compilation and analysis of field studies in productivity improvement. He noted the persuasive power of case study findings accumulated over time.

Although the literature contains many articles superficially linking productivity and training, there exists no systematic and thorough analysis of training as planned intervention in productivity improvement programs. This gap is surprising given the results of a Conference Board study (cited in Delamontagne, 1984). The Conference Board reported that 83% of responding companies with a human productivity function assigned it to the training and development department. One would, therefore, presume that training and development professionals would be actively engaged in study of the relationship between training and human productivity. Instead, efforts seem to be piecemeal and unconnected, with little attempt to synthesize existing information. Despite the movement for more evaluation of demonstrable results of training in all types of organizations, collective findings across the spectrum of training programs, settings, and approaches are minimal. Goldstein and Buxton (1982) summarized the situation:

> Training programs are expected to benefit productivity directly as measured by a large number of criteria (e.g., increased work quantity and quality, and reduced turnover, absenteeism, and accidents). Whereas training programs are a potentially powerful technique that can achieve positive results, few programs are designed appropriately so that positive benefits are likely to accrue, and even fewer are evaluated to determine whether they achieved specified goals and objectives. Instead, the choice of procedures and design of programs are based upon which new technique happens to be the latest fad. (p. 173)

What, for instance, are the apparent effects of program content or training technique on productivity results in field experiments? Which groups or classes of workers have successfully improved performance after training? Are there common characteristics of "successful" training programs aimed at improving worker productivity? If so, how do they differ from characteristics of "unsuccessful" programs? Taken together, these sorts of questions ask, "What is the composite effect of training in productivity improvement programs?"

One might expect that answers to these questions and others would appear in the "Summary of Research" sections of the Katzell et al. (1977) and Guzzo and Bondy (1983) studies. Such is not the case. Both summaries are narrative in nature, with the following conclusion typical of the earlier study (numbers in the original text refer the reader to article abstracts in the Katzell publication):

> An additional recent area of activity has been the use of training to help disadvantaged workers be productive and successful. Several such experiments have yielded positive results (6, 70, 72, 82, 84, 90). (p. 16)

Guzzo and Bondy (1983) concluded:

> Gradually, organizational attention has been moving away from selection practices, while the emphasis on training continues to mount. Still, the emphasis has not yet assumed a defined form; organizations have institutionalized a wide variety of training procedures for reasons which are equally as diverse. In light of this, a summary statement bearing on the salient themes with respect to training procedures and outcome cannot be made at this time. (p. 9)

Perhaps the failure to identify those "salient themes" in this group of 48 studies spanning eleven years is one of method, not of fact.

Recognizing "possible effects of reviewer bias or insensitivity in those earlier analyses," Guzzo, Jette, and Katzell (1985) later performed a meta-analysis on the combined set of productivity experiments originally identified and studied by traditional methods. The average effect size reported for training and instruction interventions was 0.78--the highest or most effective of all eleven interventions studied. Such information is useful: it demonstrates the strength of training as a productivity improvement intervention. It fails, however, to answer the questions of training effectiveness based on design and other variables of interest posed above.

Problem Statement

This study is an investigation of the effects of training in 48 productivity improvement experiments identified in the Katzell et al. (1977) and Guzzo and Bondy (1983) studies. A meta-analysis of the findings of those experiments will be conducted, using categorical classifications of training variables and the three condensed productivity criteria or outcomes used in the Guzzo et al. (1985) meta-analysis.

Background on Productivity and the Target Experiments

Students of productivity are faced with problems of definition and measurement. Productivity may generally be considered "the amount of goods and/or services produced per hour of human labor" (Muckler, 1982, p. 13). Kendrick (1984) gives a more comprehensive definition: productivity is "the ratio of output to inputs of labor and other resources, in real terms" (p. 1). These and similar definitions reduce productivity to a single quantifiable index. Muckler (1982) identified from the literature seven different productivity indices, ranging from those using total system input measures to those using only labor inputs. These indices were used either as expressions of labor effectiveness or demonstrations of the economic value of labor. Muckler concluded that definitions of productivity are inadequate and inconsistent.

Although the single-figure index may be useful in a global way, many investigators have recognized the utility of multiple-criterion measurement. In a study of effect of feedback on performance, Pritchard, Montagno, and Moore (1978) split productivity into three criteria for measurement: quantity of product, quality of product, and job satisfaction. Andrews (1979) listed thirty dimensions of product and quality from a study of scientific productivity. The classification scheme devised by Katzell et al. (1977) used the following categories of productivity outcomes or criteria:

<u>Production</u>
1. Quantity or rate of production
2. Quality or accuracy of production
3. Financial costs
4. Miscellaneous

<u>Withdrawal of Personnel</u>
5. Turnover
6. Absenteeism
7. Tardiness
8. Miscellaneous

<u>Disruption</u>
9. Accidents
10. Strikes
11. Slowdowns
12. Grievances

13. Alcoholism and Drug Abuse
14. Miscellaneous
15. Attitudes (If measured in addition) (p. 10)

These categories are designed to encompass any aspect of worker output or any aspect of input or cost. They are, therefore, indicators of change at either the output or input side of the productivity ratio.

When Guzzo et al. (1985) conducted a meta-analysis of all psychologically based intervention programs represented in the Katzell et al. (1977) and Guzzo and Bondy (1983) studies, they adopted the three productivity measurements used by Guzzo and Bondy in the 1983 study: output, withdrawal, and disruption. (Note the similarity to the headings in the list presented above and taken from the 1977 study.) The multiple-criteria operational definition used in the Guzzo et al. meta-analysis is accepted for use in this dissertation for three reasons. First, since the dissertation involves meta-analysis of the experiments identified by these authors, using their definition is implicit. Second, using multiple outcome variables allows for a comprehensive analysis of data. Third, the relatively limited number of experiments available for meta-analysis dictates lumping data into larger categories for the sake of reliability.

Productivity definition and measurement are not ends in themselves. "It is very clear that the main stated purpose of productivity measurement and assessment is improving worker productivity and not measurement of productivity per se" (Muckler, 1982, p. 18). Adam, Herschauer, and Ruch (1981) title their book Productivity and Quality: Measurement as a Basis for Improvement. Productivity improvement, not measurement or definition, is the real goal of productivity study.

Methods for improving productivity abound. As a general rule, methods are a manipulation of the human and technical elements of a work system. System output (productivity) is thought to derive ultimately from the interaction of those two elements (Guzzo & Bondy, 1983; Sutermeister, 1976). Improvements in technology, from the cotton gin to the computer, have contributed greatly to the productive capacity of the employee. But what does the employee contribute to the productivity equation? And how can that contribution be enlarged? These are questions for the behavioral scientist, and they are the questions that form the foundation of this study. Yet, as important as these questions are, a caution is in order: "in modern technology, output is rarely, in either goods or services, a simple and direct result of what people [italics added] do" (Muckler, 1982, p. 28). Interpreting the findings of behavioral science as they relate to productivity improvement must be done with awareness that technology may mediate results.

Katzell et al. (1977) identified fourteen categories of psychologically-based improvement programs described in published accounts of field experiments during the years 1971-1975. The categories follow: selection and placement, job design, job development and promotion, group design, training and instruction, supervisory methods, appraisal and feedback, organizational structure, management by objectives, physical working conditions, goal setting, work schedule, financial compensation, and socio-technical system. (Note that these categories ignore the technological aspects of productivity improvement and focus only on the employee.) Of the 103 experiements abstracted and indexed by these authors, 37 featured "training and instruction." Duplicating the procedures of Katzell's team, Guzzo and Bondy (1983) extended the study of field experiments to include those published between 1976 and 1981. The later study found 104 productivity improvement experiments, with 22 featuring "training and instruction." (Guzzo and Bondy defined training and instruction as "an organizational practice whereby employees are educated with regard to specific skills deemed central to their jobs," p. 9). Conclusions of both studies agreed: the most common psychologically-based intervention program on worker productivity is training and instruction.

Before continuing with other findings on how often training and instruction programs are used to improve productivity, a pause to define terms is in order. The Guzzo and Bondy definition (above) of training and instruction guided the classification of the field experiments. This definition combines elements of commonly accepted definitions developed by Nadler (1984). He defined "human resource development" as "organized learning experiences in a definite time period to increase possibility of improving job performance and growth" (p. 1.3). In Nadler's taxonomy, "training," "education," and "development" are human resources development activity areas, each with its own definition. "Training" refers to learning directed toward performance on the trainee's present job. "Education" refers to learning directed toward performance on some identified future job. "Development" refers to learning directed toward more ambiguous needs of the organization in the future. Many writers do not discriminate carefully among the terms "human resources development" (HRD), "training," "education," and "development." "HRD" is casually substituted for "training"; "training and development" can be a synonym for "HRD." Consequently, readers of this study should keep in mind that the Guzzo and Bondy definition of training may allow for inclusion of programs Nadler would classify more closely as education or development. The term "training" is used throughout this study instead of "human resources development" or "HRD" in keeping with the Guzzo and Bondy precedent.

At least one other study has reported the frequency of training and instruction programs as productivity improvement tactics. The New York Stock Exchange ("People and Productivity," 1983) sampled the 49,000 U.S. corporations with more than 100 employees. This grouping represents some 41 million people or 55% of all private nonagricultural employment. The purpose of the study was to determine the extent of quality of work life programs and industry's overall experience with these and other human resources programs. The study identified fifteen most common human resource activities. Findings on frequency of use of the fifteen programs are presented in Table 1.

Table 1

Percent of Corporations Reporting
Use of Human Resource Programs in 1983

Program/Activity	% Reporting Use
Formal training and instruction	76
Employee appraisal and feedback	72
Setting employee goals	64
Setting company objectives	55
Job design/redesign	46

Table 1--Continued

Program/Activity	% Reporting Use
Surveys of employee attitudes	45
Quality circles	44
Scheduling workflow	43
Organizational structure	40
Suggestion systems	38
Task forces	35
Structuring plant and office space	29
Personalized work hours	28
Profit sharing	25
Labor-management committees	25

Over three-fourths of all responding companies reported formal training and instruction programs. Furthermore, the Exchange found that two-thirds reported subsequent improvements in employee attitudes and morale; over half reported improvement in costs, worker productivity, and product quality. Between one-half and one-third reported improvements in service, quality of work life, safety, overtime, turnover, absenteeism, lateness, and formal grievances. Amount of improvement in any category was not attributed to individual programs, so it is not possible to determine from these findings how instrumental training was in improvement, compared to other programs or interventions. Nonetheless, formal training was ranked third in the ten activities identified as "very successful" in improving productivity.

Overview of the Dissertation

The following pages comprise the study report. Chapter II contains a review of selected literature related to productivity and training, with special emphasis on training. Chapter III is a description of study design and methods, with particular attention to meta-analysis. Chapter IV is a report of the meta-analysis findings and other data descriptive of the 48 subject studies. The fifth and final chapter is a discussion of conclusions, meaning, and recommendations drawn from the data.

CHAPTER II

REVIEW OF LITERATURE

Introduction

The need for a study such as this one goes beyond timeliness. Training is big business and training professionals face increasing pressures for accountability, proof that larger and larger investments in human resource development contribute tangible returns to the sponsoring organization. The productivity measures used in the Katzell et al. (1977) and Guzzo and Bondy (1983) studies are very similar to performance indicators in training impact evaluation, suggesting that further investigation of the subject studies can contribute to the field of training evaluation as well as training design. Evaluation of training attempts to determine training effectiveness. This section of the study will contain findings that relate to the training field at large, to training effectiveness, and to training impact evaluation and productivity measures. A basic ecology model of training effectiveness is developed and presented.

Need to Study Training

Estimates of the magnitude of American expenditures for training vary. Private employers could be spending as much as forty billion dollars a year, discounting additional costs of wages and benefits. Compiling all costs may push the figure closer to eighty or one hundred billion dollars (Eurich, 1985). This range of between forty and one hundred billion dollars was corroborated in a training industry report conducted by Training: The Magazine of Human Resources Development. The industry report estimated that "U.S. organizations will collectively deliver approximately 1.377 billion hours of formal training by the end of 1985" (Feuer, 1985, p. 45).

Formal training can be found in all kinds of organizations. Training surveyed the following industries: manufacturing, transportation/communications/utilities, wholesale/retail trade, finance/insurance/banking, business services, health services, educational services, and public administration. These categories include both public and private organizations.

Training touches all levels of employees. Training magazine's survey reported the following job classifications and mean number of hours of training received by individuals in each classification annually: executives (41.4), professionals (38.3), sales representatives

(37.7), first-line supervisors/foremen (36.4), middle managers (36.3), senior managers (35.7), customer service (33.9), production workers (33.4), administrative employees (19.2), and office/secretarial (18.8). These hours derive from formal training programs offered by in-house staff and outside vendors.

Formal training programs offer a variety of subjects or content. Training's Feuer (1985) reported:

> The three types of training most frequently offered by U.S. organizations with 50 or more employees fall into the areas of supervisory skills, management development and technical skills/knowledge updating, in that order.... Also provided by more than half of all organizations is training in communication skills, customer-relations/services skills, new methods or procedures, executive development, sales skills, clerical/secretarial skills and personal growth. The least common offerings...include remedial basic education, customer education and disease prevention/health promotion. (p. 48)

Content emphasis may shift according to the need of the moment or a particular trend. A Training editor, Feuer (1985) noticed, for example, that customer relations or services skills and sales skills have increased in training emphasis, perhaps as a result of more competitive environments or current management theories.

The pervasiveness of training programs in American organizations, the breadth of employees being trained, the expense involved, and the scope of training content illustrate that training activity is extensive. Yet training literature in general has been found lacking empirical or theoretical basis and "dominated by low utility anecdotal presentations" (Burke & Day, in press). This finding suggests that studies which attempt to extract theoretical principles from the body of extant training studies may be very useful to the profession. In particular, shortcomings exist in knowledge of design of effective training and evaluation of organizational impact.

Two design or methods studies can illustrate a typical approach found in training literature. Carroll, Paine, and Ivancevich (1972) attempted to study the relative effectiveness of training methods by analyzing research data. They found, however, that there were many limitations in the studies they identified and that the amount of research on particular training methods varied greatly. As an alternative, they surveyed two hundred training directors in firms with the largest number of employees. Subjects were asked to rate nine training methods for effectiveness in achieving six training objectives. Methods included programmed instruction, case study, lecture method (with questions), conference or discussion method, role playing, sensitivity training (T-group), TV-lecture (lecture given to large audience over TV), movie films, business gaming (using computer or hand calculator). Training objectives included acquisition of knowledge, change in attitudes, participant acceptance, retention of what is learned, development of interpersonal skills, development of problem solving skills. The investigators found in general that

> Training directors believed that about half of the training methods listed were effective and the other half not very effective for the training objective stated. Furthermore, the training methods considered effective for one objective were usually considered ineffective for another objective. This seems to indicate that the training directors are properly discriminating in their evaluations of the various alternative training methods. (p. 505)

In 1980, Newstrom replicated the earlier Carroll et al. study. Newstrom reported three conclusions. First, all but one training objective had a unique "best method" associated with it, suggesting that methods had value only when objectives were considered. Second, the televised lecture was rated as having little effectiveness for any objective; role playing was the favorite method. Third, at least two methods (and up to six) were perceived to be moderately or more effective for each of the six objectives, suggesting that choices in method are available without compromising achievement.

What is noteworthy in these studies completed almost a decade apart is not the findings, but the questions asked and methods used to study the questions. The training objectives are all oriented to the individual trainee's accomplishments; not one objective is oriented to the organization's performance. This fact lends credence to the observation that "Industrial and organizational psychology is obsessed with the evaluation of personnel functions at the individual level of performance" (Russell, Terborg, & Powers, 1985, p. 849). Adequate research findings were not available to answer the questions about effective training methods, so the investigators were forced to resort to expert opinion. This fact supports Burke and Day's (in press) contention that training literature lacks theoretical substance. Training methods were imposed on the training directors for rating. It is possible that an in-depth study of training in process or training evaluation reports would have provided quite a different list of methods. Finally, expert opinion may not hold up under more rigorous analysis of matching method to objective. Despite the obvious shortcomings of such "opinion polls," or popularity contests among training techniques or strategies, investigators have continued to use the method (e.g., Newstrom, 1975; "Trainers Rank," 1983).

This study will counter some of the shortcomings of traditional training study, then, by addressing organizational performance measures (productivity measures), by using findings of experimental studies to construct theory of effective training design, by identifying methods actually used in training situations, and by applying rigorous analytic methods (meta-analysis) to the findings of the studies of interest.

Training Effectiveness

Emphasis on Cost-Effectiveness

A common approach to training effectiveness in training literature is cost-justification of the training function. Articles with titles like "How to Calculate the Costs and Benefits of an HRD Program" (Spencer, 1984) complement evaluation book chapters with titles like "Measuring the Return on HRD" (Phillips, 1983). Phillips (1983) went so far as to introduce a special definition of ROI (return on investment) for use by training professionals. He spoke of "payback periods" (p. 193) to estimate the time required to evaluate a capital investment in training, paralleling the evaluation of a capital expenditure in equipment. He explained how "discounted cash flow...a method of evaluating investment opportunities" (p. 194) could be applied to training programs.

This propensity to reduce training effectiveness to dollars may be explained by Juran (1974). Although Juran was writing for quality professionals, not training professionals, his explanation is relevant. He began by noting that a company contains two languages:

> At the bottom of the company (i.e. nonsupervisors, foremen) the language is in terms of "things." Anyone who wishes to communicate with these levels must talk about tons, meters, kilowatts, man-hours, units of product. At the top of the company the common language is money, i.e., sales, profits, investment, return on investment, etc. Only through use of such a common language are the top managers able to make useful decisions about alternative courses of action. In such a state of affairs, middle managers and professionals <u>must be bilingual</u>. They must be able to talk to the people at the bottom in the language of things. They must be able to talk to the top management in the language of money. Failing this, they jeopardize their objectives in the competition against those middle managers who are bilingual. (p. 3-11)

If Juran is right, training professionals striving to convert training effectiveness measures to dollar figures are speaking the language of top management, the source of continued funds to keep the training department in business.

In such attempts at cost-justifying training, the entire training function seems to be the focus of evaluation. In other words, effectiveness is often directed to entire training units, not just individual training programs or certain training methods. Training units appear to be in jeopardy, in danger of being phased out in hard times (Phillips, 1983). It is no wonder, then, that such energy is expended to defend the contribution of training to the sponsoring organization. Other aspects of training unit effectiveness are presented in the next section.

Organizational Effectiveness Models Applied To Training Units

Attempts to define effectiveness in the various fields of behavioral science have been less than conclusive. The effective manager may be the one who knows the "right things to do" (Drucker, 1964, p. 5), or the one with "ability to choose appropriate objectives or the appropriate means for achieving a given objective" (Stoner, 1978, p. 13). Organizational effectiveness may be "determined by...goals and objectives, but ... effectiveness is a function of: (1) output variables (productivity/performance), (2) intervening variables (the condition of the human resources), (3) short-range goals, (4) long-range goals" (Hersey & Blanchard, 1982, p. 115). When one considers that a training unit is an organization within its sponsoring organization, application of organizational effectiveness literature can lead to the development of a model for training effectiveness.

Miles (1980) identified two traditional approaches to defining organizational effectiveness: the goal-attainment model and the systems model. He pointed out deficiencies in each, built on their strengths, and created a third model, the ecology model. Each of these models can be illustrated with training examples.

The goal-attainment model focuses on ends. In this model effectiveness is measured by "productivity (number of units produced or services delivered in a given period of time) or efficiency (number of output units produced or delivered for a given measure of input units)" (Miles, 1980, p. 362). A training unit whose effectiveness is measured under this model would report numbers of courses developed and delivered annually, number of trainees trained, and the like. This model also lends itself to quantifiable outcomes at different levels of training. For example, a reasonable goal of training is that learners master content; effective training would ensure that the goal was met. A goal of training could be that programs exhibit well-founded instructional principles; effective programs would contain elements based on the principles.

The goal-attainment model of effectiveness has inherent problems, however. According to Miles, public goals of an organization may not be its operative goals. In a training unit, a published goal may be to offer a broad range of training to the entire staff of the organization; yet one group may receive more attention because of special interests, problem analysis bias, or even training staff expertise. Different subunits of an organization may have different goals. In an industrial setting, the manufacturing and quality control departments have production and quality goals that sometimes conflict. Training units serving the individual goals of the different departments may be contributing to overall organizational dysfunction. Even profit maximization (the bottom line) is inadequate as an ultimate goal in Miles' view because it can be manipulated by management to reap short-term profits at the expense of long-term growth. In the training unit, costs can be manipulated by offering low-cost courses to employees at the bottom of the wage scale.

The systems model of organizational effectiveness focuses on means, not ends. In particular, this model emphasizes the organization's "ability to develop and manage the complex interdependencies among and between important internal and external groups" (Miles, 1980 , p. 368). Like the goal-attainment model, which relied on quantifiable data, the systems model has typical measurement strategies:

> Assessors have tended to concentrate on such factors as the health of the organization (as measured by the absence of employee strain, and so on), its ability to acquire resources needed to sustain operations, its flexibility in adapting to changes in environmental factors, its efforts in continuing employee self-development, and its adroitness in balancing resource allocations to internal maintenance, adaptive, and primary transformation processes. (p. 369)

The training unit measuring its effectiveness by the systems model would concentrate on building a bigger budget year by year, on building good human relations among staff, on quick response to requests for training programs, on staff development, and on budget efficiencies.

Miles saw problems with the systems model, as well. It can fail to question the appropriateness of an organization's direction, substituting the general goal of survival. A training department can grow bureaucratic and manage to subsist from budget to budget, but subsistence may not qualify as effectiveness. Contingencies in technologies, environment, workforce, organization life cycle stage, and cultural context dictate systems and processes in organizations. As these contingencies grow more and more complex, there is no blueprint for the "best" system. For training departments, as the sponsoring organization changes, as its strategic emphases shift as a result of threats and opportunities in its environment, as the workforce and skills mix change, the likelihood of one "best" structure, policy, and procedure decreases dramatically.

Given the deficiencies of the traditional goal-attainment and systems models of organizational effectiveness, Miles introduced the ecology model. This model assesses organizations in terms of their effectiveness in "minimally satisfying each of the goals imposed on them by the various constituencies upon which they depend" (p. 375). Measurement is made in a manner appropriate to the multiple goals the organization faces as it interacts with multiple elements in its environment.

An Ecology Model of Training Unit Effectiveness

Based on Miles' (1980) work, an ecology model of training unit effectiveness can be constructed. (For simplification, only two key constituencies are presented, although the model allows for more detail if it is found useful for analysis.) Training units depend both on the trainees, themselves, and the organization which sponsors the unit. Training units transform trainee needs, organizational resources, and performance problems by applying the training process--assessment, design, delivery, reinforcement, and evaluation--to satisfy both the goals of the trainees and the sponsoring organization. A graphic representation of the model follows as Figure 1.

The ecology model clarifies the emphasis of this study. Although effectiveness involves serving both the trainees and sponsoring organization, this study focuses on how training serves the organization. Other studies may attend to how training units best provide the useful learning trainees expect.

The model represents graphically the impulses that come from critics of limited training interventions. Dunn and Thomas (1985) deplored that "many companies spend hundreds of thousands of out-of-pocket dollars for a particular program, not to mention tens of thousands more in the lost productive time of trainees, without being reasonably sure the program will work" (p. 65). Robinson (1984) argued that a training department can have "accountability that focuses upon <u>results achieved</u>, both in terms of on-the-job behavior change and the organizational impact of the training" (p. 42). Burke (1972/1978) stated that for managerial training, "when there is evidence that training does not (1) respond to what managers report they want and need, and (2) lead to organizational change, then something is amiss" (p. 173).

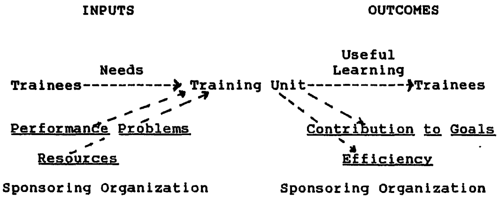

Figure 1. An Ecology Model of Training Unit Effectiveness

In terms of the present study, the ecology model illustrates the need for training method, training objective, training type, delivery system, and degree of management involvement to operate at optimum within the context of the organization to assure training effectiveness. The next section of this paper expands on how organization impact--goal contribution and efficient use of resources--can be evaluated.

Productivity Measures and Training Impact Evaluation

Evaluation of training becomes more common as training budgets grow, as technology for evaluation improves, and as emphasis on training results increases in all organizations (Phillips, 1983). As a basis for this study, a brief discussion of relevant evaluation issues is provided. Purposes and objects of evaluation are emphasized, and need for continued study is explained.

Purposes of Training Evaluation

The appropriate purposes of training evaluation are as numerous as the audiences who have a vested interest in training outcomes, as numerous as the types of training and trainees. In a word, evaluation is meant to be "useful" (Brinkerhoff, 1981). Evaluation should provide information and can be considered feedback (Friedman & Yarbrough, 1985). Those participating either as trainers, trainees, or designers can use such information to alter future activities based on response to current activities. Lists of training evaluation purposes have been constructed by several authors. The following table presents two examples for illustration.

Table 2

Purposes for Training Evaluation by Author

Author	Evaluation Purposes
Brinkerhoff (1981)	To determine the "payoff" of training, its impact on the organization
	To make the training more efficient
	To be more accountable for the training function
	To publicize the benefits of training
	To determine the cost/benefit of training to the organization
	To compare different training approaches
	To determine whether training delivery and effects are congruent with what was planned
	To identify problems with the training and ideas for their solution
	To determine needs for more, or different, training
	To gain support for particular training methods or sessions
	To determine whether a particular training approach is likely to work
Phillips (1983)	To determine whether a program is accomplishing its objectives
	To identify the strengths and weaknesses in the HRD process

Table 2--Continued

Author	Evaluation Purposes
Phillips	To determine the cost/benefit ratio of an HRD program
	To decide who should participate in future programs
	To identify which participants benefitted the most or the least from the program
	To reinforce major points made to the participant
	To gather data to assist in marketing future programs
	To determine if the program was appropriate
	To establish a data base which can assist management in making decisions

In many ways, this dissertation is an evaluation study. It is designed to extract useful information from reports of training interventions. Its purposes are to identify relationships among training variables, organization contingencies, and training outcomes (productivity measures) and to suggest ways in which the relationships discovered can guide effective training design and useful evaluation.

Objects of Training Evaluation

Just as there are multiple purposes for training evaluation, there are multiple objects. Brinkerhoff described a continuum of evaluation, "from the very immediate aspects of the training intervention itself toward the more removed benefits (payoffs) of the training" (1981, p. 35). He explained that certain points along the evaluation continuum have been intrinsically valued over others at certain times, creating opportunity for evaluators to evaluate the wrong things. In Brinkerhoff's view, "any focus for evaluation is potentially worthwhile, and none is intrinsically 'better' than another" (1981, p. 36). Other writers have presented taxonomies for evaluation. Laird (1982) recommended evaluating for training's contribution to organization goals, achievement of the learning objectives, and perceptions (of the trainees, generally).

Phillips (1983) reviewed four approaches: the Kirkpatrick approach, the Parker approach, the Bell System approach, and the CIRO approach. Kirkpatrick advocated four levels of evaluation: reaction, learning, behavior, and results. Parker divided evaluation studies into four groups: job performance, group performance, participant satisfaction, and participant knowledge gained. The Bell System was an evaluation of reaction outcomes, capability outcomes, application outcomes, and worth outcomes. In their CIRO approach, Warr, Bird, and Rackham categorized evaluation studies as context evaluation, input evaluation, reaction evaluation, and outcome evaluation. Phillips concluded that "the common thread among the evaluation experts...is the ultimate outcome which results from improved group performance" (1983, p. 41). He argued that "ultimate outcomes" can be measured and documented just as immediate and intermediate training outcomes (individual learning and behavior change) can be measured.

Phillips' description of "ultimate outcomes" of training as that which "results from improved group performance" suggests a link between the three criterion measures of productivity adopted for the present study (output, disruption, withdrawal) and training impact indicators. Again, lists compiled by evaluators may be instructive comparisons.

Table 3

Training Impact Indicators

Author	Indicators
Phillips (1983, pp. 182-188)	Value of increased output Value of cost savings

Table 3--Continued

Author	Indicators
Phillips	Value of improved quality (scrap/waste, rework, customer/client dissatisfaction, product liability, internal losses, employee morale)
	Value of time savings (wages/salaries, better service, penalty avoidance, opportunity for profit, training time)
Laird (1982, p.258)	Units of work per hour
	Units of work per worker
	Number of sales
	Dollar value per sale
	Ratio of sales to calls
	Percent of quota achieved
	Total dollar value of sales
	Number of grievances
	Percent of grievances decided
	Percent of grievance decisions sustained
	Percent of counseling problems solved
	Total minutes tardiness
	Total days absenteeism
	Number of absenteeism incidents
	Scrap
	Rejects
	Backorders filled

Table 3--Continued

Author	Indicators
Laird	Dollar value for backorders filled
	Tasks completed
	Percent of tasks completed properly
	Budgets submitted
	Budgets achieved within X percent of forecast
	Employee turnover
	Inventory turnover
	Machine downtime
	Number of disabling accidents
	Cost of accidents
	Letters and reports completed
	Percent of letters and reports which get the desired results
Donaldson and Scannell (1978, pp. 138-139)	Direct cost reductions
	Grievance reductions
	Productivity of trained versus untrained employees
	Productivity after versus before training
	Work quality
	Quantitative results
	Accident rates
	Absenteeism
	Employee suggestions

Table 3--Continued

Author	Indicators
Donaldson and Scannell	Sales volumes
	Turnover rates
	Supervisory ratings
	Profits
	Customer complaints
	Worker efficiency
	Training time required for proficiency
	Cost per untrained employee
	New product development
	New customers
	Public relations

These lists share many common elements, indicating agreement among evaluators of organizational impact. The lists could be reorganized into the three productivity criteria used in this proposed study: output (sales figures, for example), withdrawal (absenteeism), and disruption (grievances).

As noted in a previous section, some investigators have criticized industrial and organizational psychology for being "obsessed with the evaluation of personnel functions at the individual level of performance" (Russell et al., 1985, p. 849). If the lists presented above are taken as evidence, a growing emphasis on more organizational evaluation is afoot. The justification for this emphasis may be financial:

> Most training programs are directly or indirectly financed by the organizations that employ the people who participate in them. Hence, a fundamental issue in evaluation is assessing the degree to which the functioning of the sponsoring organization has been affected by that training. When we shift from looking at the behavior of specific individuals to looking for evidence that quantifies their impact on the organization in which they work, we face an increasingly demanding challenge. (Friedman & Yarbrough, 1985, pp. 234-235)

More information is needed about how training interventions impact the organization that supports them before that challenge can be effectively met.

Meta-Analysis Findings on Training Effectiveness

Meta-analysis is becoming a more common technique in behavioral science research (Kulik, 1984), but few meta-analyses have been applied to training research. A search of the literature identified two such meta-analyses. Findings of those studies are reported here.

The Guzzo, Jette, and Katzell Study

Guzzo, Jette, and Katzell (1985) completed a meta-analysis of all the 207 worker productivity experiments originally identified by Katzell, Bienstock, and Faerstein (1977) and by Guzzo and Bondy (1983). Guzzo and his associates compared the effectiveness of eleven intervention programs on three productivity measures, three organizational context variables, and two research design variables. The intervention programs (defined as "the introduction of an experimental treatment or change in one or more independent variables," (p. 277) included: recruitment and selection, training and instruction, appraisal and feedback, management by objectives, goal setting, financial compensation, work redesign, decision making techniques, supervisory methods, work rescheduling, and socio-technical interventions. The three productivity measures were output, withdrawal, and disruption. The organizational context variables were organization size, organization type, and type of worker. Research design variables were number of weeks between intervention start and measurement of a dependent variable and use of comparison groups.

The Guzzo team studied

> differences among intervention programs in the magnitude of their productivity effects, differences in effects according to nature of the productivity criterion, joint effects of multiple simultaneous interventions, and the relationship between methodological features of studies and the findings they report. (p. 279)

The investigators found that the experiments as a group represented an improvement in productivity of about one-half a standard deviation. By comparison, training programs increased productivity by an average of .78 standard deviation. Training was the strongest intervention among the eleven studied. By criterion, training had statistically significant positive effect on output measurements only (product quality, product quantity, and cost-effectiveness). Training did not demonstrate a statistically significant effect on withdrawal (turnover and absenteeism) or disruption (accidents, strikes, and other disturbances) even though the mean effect size for training was greater than 0.5 in each instance. No evidence was found to suggest that combining any two interventions resulted in a synergistic effect. (A shortage of reliable data may have influenced this result.)

Effect sizes were found to vary according to organizational context. Smaller organizations enjoyed more payback from productivity improvement interventions than did larger ones. (This finding supports a survey result reported by the New York Stock Exchange, 1983.) Impact was greater for governmental organizations than for either private for-profit or non-profit firms. Productivity effects were smaller for blue-collar and clerical workers than for sales and managerial or professional employees. Training was a frequent intervention for managerial or professional employees. (This finding is supported by the October, 1985, industry report of *Training* magazine, which reported mean number of hours for managers and professionals to be greater than for other levels of employees.)

Effect sizes also varied according to research design. Intervention effect was smaller in designs with randomized control groups than in those with other comparison groups. Weaker effects were registered as the length of the measurement interval increased.

The Guzzo, Jette, and Katzell study is important to the present study for several reasons. First, it clarifies that training was not only the most frequent productivity improvement intervention reported, but also the strongest in the published record spanning eleven years. Second, it establishes methodological precedent for application of meta-analysis to training literature. Third, it provides a design framework that can be augmented to allow investigation of a number of training variables.

The Burke and Day Study

Burke and Day (in press) conducted a meta-analysis of managerial training effects. They compared six training content areas, seven training methods, and four types of criteria. The training content areas included general management programs, human relations/leadership programs, self-awareness programs, problem solving and decision making programs, rater training programs, motivation/values training programs. Training methods included lecture, lecture/group discussion, leader-match, sensitivity training, behavioral modeling, lecture/group discussion with role playing or practice, and multiple techniques. Subjective and objective dimensions of learning, behavior, and results were measured as criteria; specifically, the four criteria employed were subjective learning, objective learning, subjective behavior, and objective results. Burke and Day attempted to answer these questions:

> (a) Across studies with respect to various criteria, how effective is managerial training? (b) for each type of criterion-measure, what is the relative effectiveness of different types of managerial training?...(c) For each type of criterion-measure, what is the relative effectiveness of different managerial training methods and combinations of methods? (p. 5)

Studies selected for analysis met three criteria: involvement of managerial or supervisory personnel, evaluation of effectiveness of one or more training programs, and inclusion of at least one control or comparison group. Total number of papers for the meta-analysis was 70.

Detailed reports were made of the meta-analysis procedures (calculation formulas) employed. These reports included all correction techniques for reducing error variance.

As predicted, results varied according to the variables under study. Burke and Day first reported findings related to managerial training content. Human relations training programs were found to be effective in improving subjective learning measured by self-report or observation, belief or judgment. It was not possible to identify variables contributing to variance of observed effect sizes in objective learning criteria (such as a test). Effectiveness of problem solving and decision making training programs was highly negative. (Few studies were available.) No definite conclusions were reported for effectiveness of general management functions training. Training for change in motivation and values was found to be positive. (Again, few studies were available.) Self-awareness training was effective in changing on-the-job behavior, as measured by self-report or observation. The investigators criticized the utility of such measures, suggesting job performance criteria be used instead. Human relations training showed improvement in managerial performance on objective results criteria such as reduced costs, improved quality or quantity, promotions and reduced errors in performance ratings; but once more, few studies were available to provide data. Overall, training content, at least as categorized by these authors, was not a powerful predictor of success on criterion measures. In fact, Burke and Day recommended that "trainers and organizational decision makers not rely heavily on training program content descriptions and labels when choosing and judging the probable utility of a managerial training program" (p. 29).

Training methods were found to account for more observed effect size variance, in general, than did training content. Behavioral modeling, sensitivity training, lecture with discussion and either role playing or practice, and multiple techniques were all reported as effective across subjective learning criteria; behavioral modeling was particularly strong. All combinations of lecture were effective for objective learning criteria. The lecture method was also effective against subjective behavior criteria, despite its relatively low popularity among training directors (cf. Carroll et al., 1972; Newstrom, 1980). The leader-match and behavior modeling training methods were also effective for subjective behavior criteria. Too few studies were available for reliable findings of the effectiveness of training methods against objective results criteria.

Burke and Day concluded: (1) researchers should improve reports to include more specific statistical data, sample characteristics, and methodology descriptions; (2) training program content descriptions may not be as good a predictor of usefulness of training programs for managers as training method; (3) more research is needed on the experience of the trainer and training effectiveness; (4) different managerial training methods "do not necessarily lead to increased knowledge and improved job performance" (p. 30), although even small effect sizes of less than one-half a standard deviation can mean sizable economic gain to an organization. Their study is important to the present one because it contains detailed computations, it is a study of some of the same variables, and it provides findings for comparison.

CHAPTER III

DESIGN AND METHODOLOGY

Design

Integrative Research Reviews

This study is an integrative research review of 48 training experiments identified by Katzell, Bienstock, and Faerstein (1977) and Guzzo and Bondy (1983). Such a review is defined as "the synthesis of separate empirical findings into a coherent whole" (Cooper, 1982, p. 291). Four purposes for conducting such research reviews have been identified in behavioral science literature: to identify and assess new developments in a field; to use empirical evidence to highlight, illustrate, or assess a particular theory or to propose new theories; to organize knowledge from divergent lines of research; and to integrate knowledge from convergent research efforts (Bangert-Drowns, 1984). Because it is designed to uncover previously undetected patterns of training effects on productivity outcome measures, the proposed study falls into Bangert-Drowns' third category, an organization of knowledge from divergent research.

Two distinct approaches to data analysis exist in integrative research reviews: traditional narrative methods and quantitative or statistical methods. Experimental evidence shows that extreme differences arise among conclusions of reviewers using the two methods (Cooper & Rosenthal, 1980, as cited in Glass, McGaw, & Smith, 1981). The approaches may be opposite ends of a continuum of increasingly stringent methods (Bangert-Drowns, 1984). On the one hand, the traditional narrative reviewer "collects some samples of studies and uses his or her expertise to offer speculative explanations for the discrepancy in study outcomes" (Bangert-Drowns, p. 3). As the number of separate studies grows, the traditional narrative reviewer has three alternatives for managing the quantity of information collected: list all studies collected, highlighting certain independent variables and their outcomes; restrict the number of studies, incurring likely selection bias; and include all relevant studies, integrate them, and draw conclusions (Hunter, Schmidt, & Jackson, 1983, as cited in Bangert-Drowns, 1984). On the other hand, the quantitative reviewer has review procedures explicitly defined. Methods for representing individual studies fairly in the review are increased. Statistical methods are used to interpret findings (Bangert-Drowns, 1984).

Experimental evidence demonstrating that reviewers come to different conclusions when using the two approaches was the impetus for Cooper's (1982) scientific guidelines for preserving internal and external validity in integrative research reviews. Cooper defined five stages in the integrative research process: problem formulation, data collection, evaluation of data points, data analysis and interpretation, and presentation of results. For each stage Cooper described potential threats to validity and suggested steps to protect validity. In the problem formulation stage, reviewers should use broad conceptual definitions rather than narrow ones. They should, however, examine many operational study details rather than broad classifications. In the data collection phase, the reviewer should access as many information sources as possible to ensure that selected studies adequately represent all studies on the topic and that sample units represent all units of interest. Sample characteristics in the compiled studies should be reported. In the evaluation of data phase, evaluative criteria used to weight the findings of separate studies should be substantive. Although the reviewer cannot change the situation, omitted data in the original studies may pose problems in evaluation. In the analysis and data interpretation stage, the reviewer must report all rules of inference used to draw meaning from the studies being analyzed. In the public presentation stage, completeness is essential: details about how the review was conducted and details about units and relations must be reported in full. Cooper's guidelines suggest that anyone proposing an integrative review should address the issues he has raised. See Appendix A for application to this dissertation.

Meta-Analysis

As Guzzo and Bondy's (1983) conclusion indicates, traditional narrative methods of integrating the findings relating training and productivity have been non-productive. Even the more recent meta-analysis of all 207 studies in the Katzell et al. (1977) and Guzzo and Bondy (1983) work has failed to distinguish particular training variables associated with successful training interventions for productivity improvement (Guzzo et al., 1985). In this dissertation, the statistical methods of meta-analysis will be applied to the 48 subject studies in an attempt to reduce separate findings to some unified whole.

Meta-analysis grew out of a need to reduce large bodies of research evidence to reliable statements that can guide policy decisions (Glass, McGaw, & Smith, 1981). Meta-analysis was introduced in 1976 by Gene Glass and has been used subsequently in approximately 1000 studies (Kulik, 1984). Meta-analysis is "the statistical analysis of the summary findings of many empirical studies" (Glass et al., 1981, p. 21). Meta-analysis makes use of transformations of commonly reported statistics in original studies. These transformations, or effect sizes, then become a new data set to which new statistical analyses can be applied. Meta-analysis attempts to identify central tendency in outcomes of studies and analyze variation without using original data (Bangert-Drowns, 1984). Bangert-Drowns (1984) offered the following sketch of meta-analysis methodology:

> A reviewer gathers all of the studies bearing on a certain topic. A reviewer may exclude studies on any grounds that are consistent with her or his purposes and explicitly stated. The outcomes of the studies are represented by a measure that is independent of the different dependent criteria. These outcomes are averaged to measure central tendency. Typically, the meta-analyst will go further and look for relationships between the study outcomes and study features as an explanation for study outcome variability. (p. 49)

Three strengths of meta-analysis bear mentioning. First, meta-analysis is quantitative. Large quantities of data can scarcely be comprehended in any other systematic fashion. Second, meta-analysis does not prejudge research findings in terms of research quality. Instead of using quality of research methods in advance as a criterion to exclude certain studies and their findings, meta-analysis uses study quality afterward in discussions of method and findings. Third, meta-analysis seeks general conclusions. Meta-analysis is capable of extracting information reliably from different studies with different approaches, different statistical analyses, and even different variables being measured. (This paragraph is a summary of a lengthy section to be found in Glass et al., 1981, pp. 21-24.)

Other investigators have criticized meta-analysis. Kulik (1984) argued that including studies with multiple outcome measures causes bias in the data set. (Glass advocated incorporating every finding or change into the new data set of transformed statistics. For example, given ten studies, one of which reports five findings, the researcher would find one-third of the new data points deriving from a single study.) Kulik explained that the multiple-representation problem is especially troublesome in regression analysis. Kulik also expressed confusion about generalizability of the effect size, claiming that there can be conceptual confusion about what the effect size represents. Hedges (1982) has criticized the categorization process in meta-analysis. He suggested testing for "fit" to determine adequacy of the categories used to classify elements in meta-analysis. Yet another criticism--this one general and not associated with a particular critic--is inclusion of all studies on a topic, regardless of methodological quality judged a priori.

Bangert-Drowns (1984) has provided a comprehensive study of the brief history of meta-analysis. He reported that consensus in the field has agreed that although meta-analysis does not remove all subjectivity from integrative reviews, it does give greater precision than traditional narrative methods. Bangert-Drowns has identified five meta-analysis methods differentiated on four factors. Methods one and two represent one of two basic differences in meta-analysis: use of meta-analytic methods to clarify literature reviews. Methods three, four, and five tend to use meta-analysis as data pooling to determine true population parameters. A summary of the five methods appears as Appendix D. Bangert-Drowns (1984) found that the methods of Glass and Kulik are most commonly used, that the Rosenthal and Hedges methods are rarely used, and that the Hunter and Schmidt method has promise for wider use.

Appropriateness of Meta-Analysis to the Present Study

As previously noted, traditional narrative analysis methods have yielded little when applied to the productivity improvement experiments of interest. Meta-analysis is applicable because it is a logical next choice, a known method not yet tried to answer the specific questions raised about training variables and productivity outcomes.

The subject studies are a jumble of methods, statistics, and findings. Fourteen dependent variables are involved. Control groups may or may not exist. Meta-analysis is applicable because its techniques can adapt to this diversity in the primary data.

The research questions probe the ways properties of training programs, study design, and organizational contingencies affect findings. Meta-analysis is applicable because quantification of the properties of interest allows statistical description of type and strength of effect.

One selection criterion for the subject experiments was that results be reported in quantitative form. The fact that quantitative data exist suggests quantitative analysis.

Method

Research Design and Procedures

The following steps are necessary to operationalize this study:

1. Use all 48 of the experiments of interest.

2. Develop a coding system. Coding will include both methodological (experimental design-specific) and substantive (training variable-specific) categories (Glass et al., 1981). Glass offered suggestions for categories, but emphasized that there is no canned method for identifying the properties to be coded. Bangert-

Drowns (1984) reported that variables are preselected by the reviewer as potential determinants of variability in study findings. They represent differences in treatment application, methodological differences and controls for validity, differences in publication history, and ecological differences, i.e., differences in setting, subjects, researcher, and so on (pp. 26-27).

3. Read the original experiment reports.

4. Code the studies.

5. Conduct statistical analysis of the studies' findings. Several substeps will be required. The meta-analysis techniques of Glass et al. (1981) will be used. Data will be transformed into effect sizes to prepare for appropriate statistical analysis. Extract effect size subgroups as they correspond to the coding variables. Prepare summary statistics for each subgroup of effect sizes.

6. Present findings: tables, charts, and descriptive statistics.

7. Present conclusions and recommendations: narrative explanation of findings and recommendations for future research or application of findings.

Selection of Studies

The 48 reports of interest in this dissertation were first identified and classified by Katzell et al. (1977)

and Guzzo and Bondy (1983). The reports share the following characteristics:

> (1) The reports were of experiments, defined here as a planned change of some feature or practice in a workplace with systematic observation and measurement of the consequences of the change. (2) The results of the study pertained to an objectively measured aspect of productivity. (3) The experiment took place in enduring organizations with missions to provide goods or services and in which workers were gainfully employed; thus, reports of simulations, laboratory studies, and research in other than field settings were excluded. (4) Each was a publicly available report of a productivity experiment carried out in the United States. (5) the date of publication was between 1971 and 1981, inclusive. (Guzzo et al., 1985, p. 280)

In addition, each report features training and instruction used as an intervention to improve worker productivity. It is anticipated that some of the reports upon closer examination, may lack information needed for calculating an effect size. In that case, such reports will be coded for variables but will provide no effect sizes. (Guzzo et al., 1985, found only 98 usable reports in an original group of 207.)

Productivity Measures

Outcomes will be reported as one of three types or aspects of productivity: output, withdrawal, and disruption. These measures, used by Guzzo et al. (1985), encompass the multiple criteria first established by Katzell et al. (1977). Output includes product quantity, product quality, and cost effectiveness. Withdrawal includes turnover and absenteeism. Disruption includes accidents, grievances, disciplinary action, and a miscellaneous category. Although Katzell identified three report findings on "attitudes" in the earlier study, Guzzo and Bondy dropped the category in their extension. Because of the low representation of this category, it is omitted here, as well.

Substantive Training Variables for Coding

Training Method

Three methods will comprise the classification scheme: information processing, simulation, and on-the-job activities. As defined by Holviak (1982), these categories describe groups of more specific methods. Information processing includes lecture, conference, T-group, programmed instruction, closed-circuit TV, laboratory training, and films. Simulation includes case study, role playing, and in-basket exercises. On-the-job activities include job rotation, on-the-job coaching, performance feedback, and apprenticeships. A slight revision of Holviak's definitions is appropriate to clarify coding decision rules in the present study. Modeling is coded as a simulation method.

Training Objective

Training objectives will be coded as knowledge, skill, or attitude. Derived from Bloom's (1956) cognitive, psychomotor, and affective domains, these categories have been given operational definitions by Davies (1981). Knowledge comprises what, when, who, how, why, and where. Skills are a matter of what to do, how to do it, and when to do it. Attitudes include what to like and dislike, how much to like and dislike, and when to like and dislike.

Training Type

A nine-element typology suggested by Tracey (1971) is useful. (1) Orientation of new employees is "designed to provide the new employee with as much information about the enterprise in as short a time as possible so that he can readily adjust to the environment and become productive" (p. 26). (2) In safety training programs "designed for either operative or supervisory personnel, correct safety procedures are demonstrated or discussed" (p. 27). (3) Trade and semiskills training is "given to clerks, typists, and other people from high schools, trade, and technical schools" (p. 27). Shopworkers learning to use new equipment or materials also belong to this category. (4) Technical training "often involves short courses in particular technical areas: ... production control, work simplification ... computer programming, [and] operations research" (p. 28). (5) Sales and dealer training includes product knowledge, sales skills, and sales promotion. (6) Human relations training attempts to "change the behavior of employees in their day-to-day relationships with both enterprise personnel and customers" (p. 28). (7) Presupervisory training is designed to develop supervisory, human relations, and leadership skills in those about to be promoted to supervisory positions. (8) Middle-management development focuses on management theory to prepare technicians for management roles. (9) Executive development continues to add skills to middle managers in preparation for top management. Since "supervisory methods" is classified as a separate productivity improvement intervention in the body of reports at hand, "presupervisory training" will be dropped as a category for training type in the proposed study, leaving eight of Tracey's nine.

Trainer

Trainers will be classified as internal (an employee of the organization) or external (a vendor to the organization). A subclassification of "internal" will be made to include line manager, training specialist, or other (Tracey, 1971).

Management Involvement

A critical element of maintaining behavior acquired during training is management involvement (Michalak, 1981). Three forms of involvement, plus a "none" category, will be coded: (1) involvement in training design, including providing actual work problems and needs analysis data (Michalak, 1981); (2) serving as a trainer (Gallucci, 1983; Schonberger, 1986; Tracey, 1971); (3) providing feedback and follow-up, including conducting follow-up meetings, making systematic changes after training, assigning task forces to respond to training issues, and the like (Michalak, 1981).

Organizational Context Variables

Some elements of the organization may affect the impact of training and instruction on productivity. This study will use the variables employed in the Guzzo et al. (1985) meta-analysis in order to expand the findings reported for the training intervention in general. Those variables are organization size, organization type, and type of worker. Size will be recorded as small (100 or fewer employees), medium (101-1000 employees) or large (over 1000 employees). Organization type will be categorized as private (for profit), government (including military) or other (non-profit, education, health service). Type of worker will be recorded as managerial or professional, blue-collar labor, sales, or clerical and office.

Research Design Variables

Again, to expand the work of Guzzo et al. (1985), the study will use these design variables: (1) number of weeks between end of training and measurement of dependent variable and (2) use of comparison groups. Comparison groups will be categorized as true control groups, nonequivalent comparison groups, groups subject to training other than that of the experimental group, or self-controls with a time-series design. (Note that Guzzo et al. recorded number of weeks between beginning of intervention and measurement. Recording time elapsed between end of training and measurement will allow the present study to generate data on the enduring effects of training.)

Significance of the Study

If patterns of training effects on productivity criteria are found, this study can make important contributions to the field of human resource development in at least four ways:

1. The study can guide HRD professionals who must

create or help create the training components of productivity improvement programs in their own organizations.

2. The study will clarify the current role of training in productivity improvement programs, allowing the HRD field to assess where it stands, to demonstrate bottom-line results, and to plan for continued improvement.

3. The study will provide a model for aggregating training evaluation data which can serve both training practitioners and the academic community.

4. The study can demonstrate the efficacy of meta-analysis in HRD research.

If no patterns are found, the study will report support for Guzzo and Bondy's (1983) conclusion that no discernible patterns of meaning can presently be extracted from the diversity of training programs aimed at improving productivity. In any event, the proposed study demonstrates an important principle of research efficiency: it makes use of existing research data, extending present knowledge by alternative analysis approaches.

Limitations of the Study

The proper use of meta-analytic methods can strengthen the internal validity of integrative reviews (Cooper, 1982). Several threats to external validity or generalizability should be mentioned, however.

1. Selection bias in the procedures used for compiling the subject studies. Of primary concern is the criterion that the studies be published. Just how representative these studies are of the population (all reports of productivity improvement experiments featuring training and instruction) is questionable.

2. Possible bias toward publishing positive findings and away from publishing negative findings. (See Katzell et al., 1977, for a discussion of this point.) Despite this potential bias, it is still possible for "reviews that claim only to characterize the available findings [to] fulfill their purposes whether or not publication histories differentiate their findings" (Bangert-Drowns, 1984, p. 53).

3. Context of the original experiments may vary widely. Reports come from public, private, goods-producing, and service-producing organizations.

4. Cooper (1982) noted that the secondary researcher is helpless against missing data in the reports he or she studies. Perhaps categories and coding will need to be adjusted to accommodate only the most general characteristics of training program or experimental design. The trade-off for such a move is the possibility that nuances of meaning will be missed. (Bangert-Drowns, 1984, cites some critics of meta-analysis who call this a methodological fault.)

5. Variance in coding decisions between the earlier Katzell and later Guzzo teams. This concern may be minimized by the clarity and completeness with which the Katzell group defined its decision rules, but deserves mention, nonetheless.

These and other limitations are to be considered in study design and in analysis and reporting of results.

CHAPTER IV

RESULTS

Effect Sizes Overall

Effect sizes for each dependent variable in all subject studies were calculated according to the guidelines and models provided by Glass, McGaw, and Smith (1981). Effect sizes are a common metric that allow researchers to combine statistical findings of diverse studies. Numerically, they are a ratio of the difference in mean performance of experimental and control groups to the standard deviation of the control. Positive effect sizes indicate stronger performance by experimental groups; negative effect sizes indicate stronger performance by control groups.

The 155 effect sizes in this study were analyzed first as a total group, then in subgroups as they corresponded to the variables of interest. Because of unequal cell sizes analysis of variance was not performed. Significance testing, on the whole, is tenuous practice when analyzing effect sizes because sampling distribution theory is insufficient (Guzzo et al., 1985). Only descriptive statistics will be presented here. This more conservative method of analysis is advocated by Landman and Dawes (1982). Table 4 contains summary statistics for the total group of effect sizes. Subgroup statistics will be reported in succeeding tables.

Table 4

Summary Statistics for All Effect Sizes

Frequency	Mean	*SD*	Range
155	0.67	1.37	15.19

Combined, the effect sizes reflect the strength of training interventions in productivity improvement attempts. Across all subject studies experimental groups performed an average of .67 standard deviation better than their corresponding control groups. In other words, if mean performance of control groups on any productivity criterion was 100 with standard deviation of 10, experimental groups would tend to score 106.7. To the organization sponsoring training and expecting performance return on its investment, an improvement of this magnitude could translate to considerable economic gain.

Except for a single high value of 14.75 which was almost three times the next highest of 5.16, the effect sizes tended to cluster between zero and one. A frequency distribution for all effect sizes is presented as Figure 2. This distribution again supports the strength and positive direction of training outcomes as reported in the subject studies. Some researchers reported cases in which training did <u>not</u> improve performance, but most reported positive findings.

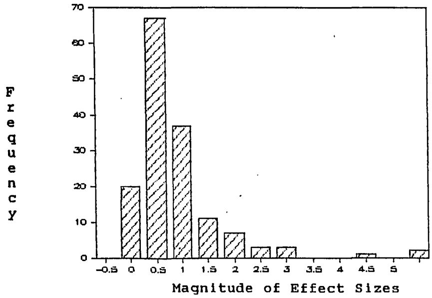

Figure 2. Frequency Distribution for All Effect Sizes

Productivity Criteria

All studies were coded for productivity criterion regardless of whether their findings produced effect sizes. (Some studies lacked adequate statistical information to allow effect sizes to be calculated or derived according to the Glass procedures employed.)

Three productivity criteria were used: output, including product quantity, product quality, and cost effectiveness; withdrawal, including turnover and absenteeism; and disruption, including accidents, grievances, and disciplinary action. These criteria are typical human resources performance measures. Nearly four-fifths of the measures presented in the subject studies concerned output; nearly one-fifth concerned withdrawal measures; fewer than one-tenth concerned disruption measures. If these studies were representative of all training interventions, a graduate entering the training and development field could anticipate most professional activity to revolve around trainee or organizational output. Table 5 is a display of coding frequency (number of studies with findings reported for the criterion) and summary statistics for all three criteria.

Table 5

Statistical Summary of Effect Sizes For Productivity Criteria

Criterion	Frequency of Coding	Mean	SD	Range
Output	40	.68	1.49	15.14
Withdrawal	10	.57	.63	2.48

Table 5--Continued

Criterion	Frequency of Coding	Mean	SD	Range
Disruption	4	.85	.44	2.83

Note: Frequency totals reflect studies with findings in more than one criterion category.

Strength of training interventions varied among criteria. Strongest effects (highest mean effect sizes) occurred when training was applied to combat disruptions in the workplace. The small data base (only eight effect sizes) makes this finding questionable, however. Strength of effect on output measures was a near-match for the overall mean effect size, possibly because of the concentration of data in this category. The weakest improvement appeared in interventions aimed at reducing absenteeism and turnover, yet experimental groups in this category still gained over half a standard deviation more than controls.

Substantive Training Variables

Subject studies were coded for eight substantive training variables. These included five variables usually under control of training developers or designers and three variables of a contextual or given nature. All studies were coded, regardless of whether they produced effect sizes. In the event a study author had failed to report on the dimension or variable in question, an "unreported" designation was given. Design variables included training method, training objective, trainer type, training content type, and management involvement level. Contextual variables included organization size and type and worker type. Effect sizes were extracted in subgroups corresponding to the coding classification within each variable. Results will be reported in this section on a variable-by-variable basis.

Unreported information was a continuing problem in this part of the meta-analysis. As is illustrated in Table 6, some variables were repeatedly omitted from study reports. Nearly one author in three, for instance, failed to identify the trainer in the training program as internal or external to the organization. About two in every five neglected to record the size of the organization. The impact of these omissions is multiplied by the subsequent inability to classify and analyze the effect sizes for the category. In other words, a study might contribute ten effect sizes to the total number in the meta-analysis. Those ten are lost each time a variable is unreported. In the case of the "Trainer Type" variable, for example, only 114 of the total 155 effect sizes were available for analysis--a twenty-seven percent loss in efficiency. The missing data problem is discussed in more detail in Chapter V.

Table 6

Coding Frequency for All Substantive Training Variables

Variable and Category	Coding Frequency
Training Method	
Information Processing	24
Simulation	20
On-the-job	8
Unreported	2
Training Objective	
Knowledge	15
Skill	34
Attitude	10
Unreported	1
Trainer	
Internal - Training Specialist	7
Internal - Line Manager	8
Internal - Other	3
External	17
Unreported	15
Management Involvement	
Design	8
Follow-up or Feedback	14

Table 6--Continued

Variable and Category	Coding Frequency
Management Involvement	
Trainer	3
Combined Design and Follow-up	3
Unreported	28
Training Content Type	
Technical	11
Executive (Management)	15
Trade and Semi-skills	3
Human Relations	9
Sales	2
Safety	3
Orientation	4
Other	1
Organization size	
Small	1
Medium	10
Large	17
Unreported	20
Organization Type	
Private for-profit	26
Government	14
Private non-profit	6

Table 6--Continued

Variable and Category	Coding Frequency
Organization Type	
Unreported	2
Worker Type	
Professional	7
Manager	21
Laborer	16
Clerical	2
Sales	3
Other	2

Note: Duplication exists in counts when variables were combined in a subject study.

Design Variables

Training Method

Training methods were classified as information processing, simulation, or on-the-job. Training programs relying on the more traditional lecture, conference, T-group, programmed instruction, closed-circuit TV, laboratory training, or films were coded as information processing. Programs emphasizing case studies, role plays, and in-basket exercises were coded as simulation. Those programs based on job rotation, on-the-job coaching, performance feedback, and apprenticeships were coded as on-the-job. In cases where a combination of methods was used in a single training program, the combination was coded. For example, a study including information processing and simulation would have been coded IP, SIM. Summary statistics contain duplication within categories as a result of combined methods.

As exhibited in Table 7, methods had varying effectiveness. The more traditional methods of information processing produced a mean effect size of 0.97, almost twice that of simulation methods (0.48) and nearly three and one-half times that of on-the-job training programs. This substantial difference may refuel arguments about the effectiveness of traditional, classroom-based instruction. (One may note that the dispersion of the information processing category includes the 14.75 value. This value does inflate the mean effect size by 0.23. Omitting the outlying value yields a corrected mean effect size for information processing of 0.74, still a considerable increase over the other two methods.)

Table 7

Statistical Summary of Effect Sizes for Training Method

Method	Mean	SD	Range	Freq.
Information Processing	0.97	2.11	15.07	55
Simulation	0.48	0.64	3.29	83
On-the-job	0.28	0.40	1.49	33

Note: Includes duplication resulting from combined methods.

Although the simulation and on-the-job methods did not produce such strong mean effects, the magnitude of their effects was less variable, as indicated by their respective standard deviations of 0.64 and 0.28. Standard deviation for information processing was 2.11 (1.00, omitting the single outlier for comparison). In other words, a trainer relying on these methods could expect less fluctuation in the performance outcomes than when relying on information processing methods.

Training Objective

Primary training objectives were coded as knowledge, skill, or attitude. As in the training method category, objectives were sometimes combined in the programs described in the subject studies. Duplication exists in the statistical summaries as a result.

It may have been predictable that the highest coding frequency for training objectives occurred in the skills category (57%; see Table 6). Improving productivity on-the-job might imply skills improvement. Yet the most frequent objective was not the strongest performer.

Summary statistics for training objective are presented in Table 8. The knowledge category had a mean effect size of 1.34--a comparative advantage for experimental groups of one and one-third standard deviations over controls. By contrast, skills objectives produced a mean effect size of 0.48--nearly one-half a standard deviation. Attitude objectives were less effective, producing mean effect size of 0.28--just over one-fourth a standard deviation. (When the extreme value is disregarded for knowledge measures, the mean effect size remains nearly double that of skill measures--0.97 compared to 0.48.)

Table 8

Statistical Summary of Effect Sizes for Training Objective

Objective	Mean	<u>SD</u>	Range	Freq.
Knowledge	1.34	2.50	14.93	36

Table 8--Continued

Objective	Mean	SD	Range	Freq.
Skill	0.48	0.60	3.29	119
Attitude	0.28	0.55	1.96	36

Note: Includes duplication because of combinations of objectives in a single training program.

As in the training methods category, more variation occurred in the knowledge subgroup of effect sizes: a standard deviation of 2.5 and a range of 14.93, compared to standard deviations of 0.60 and .55 and ranges of 3.29 and 1.96 for skills and attitudes, respectively. (Standard deviation for knowledge measures reduces to 1.14 with the extreme value omitted.)

Trainer Type

Training developers or managers can usually choose the trainer for a program. In fact, for the subject studies in which trainer type was reported, half were internal and half external to the organization (see Table 6). The question of interest is whether it makes a difference in the training's effectiveness if the trainer is an employee (internal) or not (external). Furthermore, if the trainer is internal, is there a difference in effectiveness among training specialists, line managers, or other internal trainers (special project directors in the cases represented here)? Table 9 displays the statistical summary of effect sizes for this variable.

Table 9

Statistical Summary of Effect Sizes for Trainer Type

Trainer Type	Mean	SD	Range	Freq.
Internal				
Tr. Specialist	0.43	1.50	1.73	12
Line Manager	0.30	0.60	2.69	33
Other	0.85	0.66	2.24	7
All Internal	0.38	0.61	2.79	52
External	0.77	1.88	15.07	62

Training programs taught by external trainers provided an advantage of just over three-fourths of a standard deviation (0.77) for experimental training groups. This mean effect size was almost exactly twice the average for all internal trainer groups (0.38). (When the extreme value is omitted, the mean effect size for external trainers drops to 0.54--still a 42% increase over the mean for internal trainers.)

Within the internal trainer classification, low frequency of effect sizes for the training specialist and "other" classifications make findings questionable. Nonetheless, it should be noted that the strongest performance came not from training specialists and line managers, but from "others," special project managers: mean effect size of 0.85 compared to 0.43 and 0.30, respectively.

More variation appears in the external trainer measures--1.88 versus 0.61. (When the extreme value is disregarded, however, the standard deviation for external trainers drops to 0.58, almost identical to that for internal trainers.)

Training Content

A wide variety of training content is available to the training developer attempting to improve productivity in the workplace. The programs described in the subject studies were coded into the following categories: technical, human relations, orientation, safety, executive, trade and semi-skills, and sales. As Table 10 shows, four of these seven categories were almost unrepresented by effect sizes for analysis. (Table 6 shows the distribution of subject studies, not just effect size cell frequencies.) The disproportionately large number of effect sizes deriving from mid-management training programs (over half the total available for analysis) and the subsequent few effect sizes in other categories make these findings questionable.

Table 10

Statistical Summary of Effect Sizes for Training Content

Training Content	Mean	*SD*	Range	Freq.
Executive	1.00	1.96	14.98	68
Technical	0.36	0.45	1.52	21
Human Relations	0.88	0.58	2.24	16
Orientation	** single value of 0.58 **			
Safety	0.59	0.22	0.63	6
Trade & Semi-skills	0.60	0.27	0.79	5
Sales	0.62	0.05	0.16	8

Strongest training effects were found for executive content. Over all results in the subject studies, managers trained in experimental groups out-performed control groups by a full standard deviation. Strong effects were also found for human relations training content--experimental group performance exceeded control group performance by 0.88 standard deviation. Technical training, on the other hand, produced comparatively weak effects--a mean effect size of 0.36. Strong effects, although questionable for lack of data, were also found for the other four content areas: orientation, safety, trades and semi-skills, and sales. Mean effect sizes tended to cluster at about three-fifths of one standard deviation, a little below the overall mean effect size reported earlier for the composite.

When the extreme value is omitted, the mean effect size for executive training content drops to 0.79. This corrected mean is slightly less than the mean for human relations training, suggesting that the strength of the two types of content may be quite similar.

Although executive training content may produce strong effects, these effects tend to be more erratic than effects from human relations, technical, or any of the other content areas examined. Even when the extreme value is disregarded, the standard deviation of the executive effect sizes remains at 1.01 (1.96 including the extreme value.) All other content areas produce tightly clustered effect size distributions with standard deviations of 0.58 or below. This analysis suggests that trainers could expect predictably strong overall improvement in executive training, but that specific programs would be likely to vary in strength of success.

Management Involvement

Training designers and developers can incorporate line managers into training programs in several ways.

The training programs described in the subject studies were coded for type of involvement of direct supervisors: serving as trainer, providing scheduled feedback and follow-up to trainees, providing information in the design phase, and combined design and follow-up assistance. Because study authors failed to report type of management involvement in many cases, about one-fourth of the effect sizes were lost to this analysis. Table 11 displays summary statistics for the management involvement variable.

Table 11

Statistical Summary of Effect Sizes For Type of Management Involvement

Involvement Type	Mean	*SD*	Range	Freq.
Feedback or Follow-up	0.38	0.46	1.86	69
Design	0.54	0.62	2.63	27
Trainer	0.86	0.35	0.89	4
Combined Design and Follow-up	0.42	0.40	1.73	14

Few dramatic differences in either strength or consistency of training effects are noticeable among the types of management involvement reported in the subject studies. The single exception is the high mean effect size for managers serving as trainers--0.86. This mean is based on only four values, however, so it must be viewed with caution. Involving managers in the design phase of training appears to have been more potent than involving them either after the training or at both points. Post-training involvement only yielded the lowest mean effects. Slightly more variation in outcome occurred with design involvement.

Contextual Variables

Following the precedent set by Guzzo et al. (1985), subject studies were coded for three contextual variables: organization size, organization type, and worker type. These variables are not manipulable by a training designer but may contribute to the success of a training intervention for improving productivity. As with design variables, summary data are presented variable by variable.

Organization Size

Organization size was classified as small (fewer than 100 employees), medium (100 to 1000 employees), or large (more than 1000 employees). Because of missing information in the subject studies, only 96 effect sizes were available for this analysis--a loss in efficiency of 38%. Table 12 presents statistical results for available effect sizes extracted for each classification.

Table 12

Statistical Summary of Effect Sizes
For Organization Size

Classification	Mean	SD	Range	Freq.
Small	2.19	3.89	14.75	12
Medium	0.68	0.57	2.79	44
Large	0.74	1.10	5.34	40

Although few effect sizes, all of which came from one study, were available for small organizations and results must, therefore, be questioned, training and instruction interventions appear to have been very successful in small organizations. With all data included, experimental training groups in small organizations out-performed controls by over two standard deviations. Removing the extreme value drops the corrected mean effect size for small organizations to 1.05, still a third of a standard deviation above the mean effect size for large organizations (0.74). More variation was also present in effect sizes from small organizations; however, excluding the extreme value lowers the standard deviation from 3.89 to 0.97. This corrected figure is slightly less than the standard deviation for large organizations (1.10).

In other words, strongest improvement effects were found in small organizations. Large organizations came next, and medium-sized organizations were third. Place is deceptive, however, for even in the setting with weakest effects average performance of experimental groups was two-thirds a standard deviation better than controls. Consistency was greatest among effect sizes from medium-sized organizations.

Organization Type

Organizations were coded as private for profit, governmental, and other. The "other" classification most frequently included hospitals. In only two cases did study authors fail to identify the type of organization (see Table 6). Summary statistics for organization type appear in Table 13.

Table 13

Statistical Summary of Effect Sizes
For Organization Type

Organization Type	Mean	SD	Range	Freq.
Private	1.11	1.94	14.83	66
Governmental	0.38	0.52	2.24	52

Table 13--Continued

Organization Type	Mean	SD	Range	Freq.
Other (e.g., hospitals)	0.31	0.32	1.32	21

Private for profit organizations produced the strongest training effects. Experimental groups exceeded control groups by an average of 1.11 standard deviations (0.90, corrected by removing the extreme value). These values represent roughly a three-fold (two-fold, when corrected) improvement over performance in government or other settings. Variation was also highest in the private setting (standard deviation of 1.94, corrected to 0.97). Relatively large and similar frequencies for private and governmental classifications lend credence to these findings.

Worker Type

Trainees were classified by type according to their job titles: professional, manager, labor, clerical, or sales. As would be expected, given the prevalence of mid-management content in subject training programs noted above, almost half the effect sizes available for this analysis refer to managers. Very little information was available for clerical or sales trainees. Table 14 contains a statistical summary of effect sizes for this variable.

Table 14

Statistical Summary of Effect Sizes
For Worker Type

Worker Type	Mean	SD	Range	Freq.
Manager	0.98	1.86	14.98	76
Professional	0.07	0.25	1.32	33
Labor	0.67	0.49	2.24	30
Sales	0.62	0.05	1.30	9
Clerical	** single value 0.70 **			

When experimental groups of managers were trained, they posted productivity improvements almost a full standard deviation better than those of control groups. (Correcting by excluding the extreme value, the mean effect size for managers falls to 0.80, still the strongest effect for all worker types.) Strong effects were also observed for labor--hourly employees--whose performance in experimental groups outstripped controls by two-thirds a standard deviation. Professionals in the subject studies were most often aides or others in the health care field (see the "other" classification of organization type in Table 13). Training of these employees was singularly ineffective as a productivity improvement tactic. Of all subgroups for all variables, this one demonstrated the weakest performance--a minimally positive improvement of experimental groups over controls of only 0.07 standard deviation. Lack of data for the clerical and sales groups makes results questionable for those classifications, despite the apparent strong showing for sales trainees.

Consistency of results also varied among trainee types. Training for managers was most variable (standard deviation of 1.86; 0.96 corrected for extreme value). Training for sales and professionals was least variable (standard deviations of 0.05 and 0.25, respectively). That is to say for professionals, at least, that results tended to be not just low, but consistently low.

Study Design Variables

Subject studies were coded for two design variables: (1) number of weeks between training completion and measurement of dependent variable and (2) type of control group employed. All studies were coded, regardless of whether they contained sufficient statistical information to produce effect sizes. It is important to note again that the generalizability of this meta-analysis is to published studies. Methodological characteristics of studies may contribute to differences in effect sizes.

Elapsed time between training and measurement was coded in six-month or twenty-six week blocks spanning two years. No study reported a measurement beyond this point. Comparison groups were coded as true control with random selection; other training, in which experimental groups were compared with controls receiving another sort of training or method of training; nonequivalent, where matching or convenience produced control groups; self-comparison, in which trainees were compared against their own performance in a time series design; and none, in which the researcher employed no comparison group. (There were no effect sizes for the three studies with no comparison group.) Summary statistics for the two design variables appear in Table 15.

Table 15

Statistical Summary of Effect Sizes
For Study Design Variables

Variable and Category	Coding Frequency	Mean	SD	Range	Freq.
Elapsed Time in Weeks					
0-26	23	0.52	0.67	3.29	69
27-52	10	0.61	1.08	5.45	40

Table 15--Continued

Variable and Category	Coding Frequency	Mean	SD	Range	Freq.
Elapsed Time in Weeks					
53-78	3	2.14	3.74	14.75	13
79-104	2	0.63	(two values only)		
Unreported	10				
Comparison Group					
True Control	8	0.36	0.34	1.74	35
Other Training	6	1.27	1.45	5.13	14
Nonequivalent	17	0.28	0.41	1.89	61
Self-Comparison	14	0.96	0.81	2.94	43
None	3	(No effect sizes)			

These data suggest that training effectiveness actually increased over time. Mean effect size for the most frequently-used immediate measurement period, up to six months after the training was completed, was 0.52. This was a consistent measurement, indicated by the relatively small standard deviation of 0.67 over 69 effect sizes. The forty measurements taken between six months and a year after training produced a mean effect size of 0.61 with standard deviation of 1.08. After a year, mean effect size increased to 2.14 (1.09, when the extreme value is omitted) with standard deviation of 3.74 (0.90, corrected for the extreme value). Only three studies incorporated measurements in the one year to eighteen months time block, so the thirteen effect sizes analyzed represented a small portion of the total number of studies. Two studies included measurements taken after eighteen months, but they produced only two effect sizes--too little data for analysis.

Methodological differences among the studies reflected differences in effect sizes. Nonequivalent comparison groups and true control groups under-performed experimental groups by just over one-fourth and one-third standard deviation, respectively. Differences were greater for experimental groups compared against their own performance: mean effect size for self-controls was almost a full standard deviation. Strongest differences appeared between experimental and control groups receiving two types or methods of training--1.27 standard deviations mean effect size. (This mean was derived from only fourteen effect sizes produced by six studies.) Measurements were more consistent for the more rigorous designs, as indicated by the smaller standard deviations.

CHAPTER V

SUMMARY AND DISCUSSION

Summary

This study has been a meta-analysis of forty-eight published reports of field experiments in which training and instruction programs were used as interventions to improve worker productivity. Three productivity criteria were the basis of measurement: output, withdrawal, and disruption. Three research questions were explored: (1) what were the effects of training interventions on each of the three productivity criteria; (2) how did effect sizes differ according to categories of eight substantive training variables; (3) how did effect sizes vary according to two study design variables?

Summary statistics calculated for all subgroups of effect sizes associated with variable categories indicated that training programs contributed to an overall improvement of experimental group performance over control group performance of 0.67 standard deviation. Training interventions to improve disruption measurements were most successful; however, a shortage of data cast doubt on the reliability of this finding. Mean effect size for the output criterion was a near-match for the overall mean effect size; mean effect size for withdrawal was a tenth of a standard deviation less.

The training design parameters of information processing method, knowledge objective, external trainer, executive training content, and line manager involvement in design, were associated with strongest effect sizes in the subject studies. Contextual variables associated with strongest effect sizes included small (fewer than 100 employees) organizations, private for-profit organizations, and managers as the training recipients. In all cases, consistency of measures did not necessarily follow strength.

Strongest mean effects were associated with the design variables of longer elapsed time between training end and criterion measurement and comparison groups receiving other training. (The "other training" category lacked data and was followed in strength by self-comparison groups.) More rigorous comparison group designs were associated with considerably lower but more consistent measurements.

Missing data in the original reports substantially reduced the number of effect sizes available for analysis in the cases of several variables.

Conclusions

Conclusions will be drawn for each of the three research questions. In each instance, conclusions will be considered in the context of previous relevant integrative research.

<u>Strength of Training as a Productivity Improvement Intervention</u>

Training appears to be a useful intervention in productivity improvement efforts. This study found an overall mean effect size of 0.67 in the 48 subject studies. In other words, over all training programs reported in the studies, regardless of productivity criterion used, workers who were trained improved performance about two-thirds a standard deviation more than their comparison groups.

Although the specific numbers and analytic techniques differ between the present study and the Guzzo et al. (1985) study, findings are similar enough to offer strong evidence that training can, in fact, make a substantial contribution to improved worker productivity. The Guzzo group used slightly different meta-analytic techniques to compare training interventions to ten other psychologically-based interventions in 98 usable studies (including those analyzed herein). They reported an overall average effect size of 0.78 for training.

Furthermore, training is likely to be more successful when applied against certain productivity criteria. Findings for the present study differed slightly from those reported by the Guzzo team. Guzzo and his associates found strongest training effects against output measures; in this study, strongest effects were found against disruption measures. Lack of data in both studies make conclusions tenuous for both disruption and withdrawal, however. There is, therefore, some support in the results of this study for Guzzo's findings.

Curiously, the order of strength of the training intervention applied against the three productivity criteria found in this study matches exactly the order of strength of all eleven psychological interventions in the Guzzo study. The contradiction posed by disagreement on training's impact and agreement on overall intervention impact may indicate need for further investigation to clarify whether differences in findings are methodological or substantive in nature.

Substantive Training Variables

This study departed from the Guzzo meta-analysis at the point of exploding the intervention category "training" into component variables. The intent was to determine whether patterns of meaning were discoverable in the subject studies as regards training program parameters. It is possible that more questions were raised than answered by this procedure.

Training variables were subdivided into a group of five generally under the control of a training developer or designer and three of a contextual nature. It is tempting to take a cut-and-paste or formula approach to findings. That is, given that strongest training effects were found for information processing method, knowledge objective, external trainer, executive training content, and line manager involvement in design, one interpretation might be that this set of parameters is optimal in all cases. Similarly, small private for-profit organizations training their managers would appear to be an optimal context.

No interpretation could be more dangerous or more disregarding of the research design. It is important to remember that variables were examined separately, not simultaneously. Any conclusions must, therefore, exclude joining variables unless there is a logical and supportable connection between them. Findings of this study would <u>not</u> direct a training manager to attack all productivity improvement projects with training solutions. Nor would they direct that same manager to apply only knowledge objectives, to train managers, to dissolve the internal training department, and so on. Much more must be learned about the complex interactions of variables, criteria, and contexts before such directives would be valid.

Design Variables

Many of the findings of the present study support those of Burke and Day (in press). Their meta-analysis of 70 reports of managerial training addressed questions of content, method, and overall effectiveness of managerial training.

First, Burke and Day found that despite its relatively low popularity among training directors (cf. Carroll et al., 1972; Newstrom, 1980), the lecture method was effective against both subjective behavior (self-reports of on-the-job-behavior after training) and objective learning criteria. Insufficient data were available to allow Burke and Day to draw conclusions on the effectiveness of the lecture method on objective on-the-job behavior. By comparison, the information processing method, which capitalizes on lecture, was found to correspond to the strongest worker performance effects in the present study. Information processing methods were linked with higher performance improvements than either simulation or on-the-job methods. Since the effects recorded are predominantly objective behavior measures, these findings supplement those of Burke and Day and provide support for the effectiveness of lecture and similar information processing methods of training.

It is important to note, however, that the population in the Burke and Day study and the plurality in the present study were a special group of workers--managers, with presumably different experiences and educational levels from many other American workers. Even if information processing methods are most potent with this group, they may not be most potent for all groups. Further, the measures of improvement for managers may be substantially different from those for other worker groups. For instance, the most common measure for improvement among managerial or supervisory ranks found in the 48 subject studies was supervisor ratings of performance. If ratings improved after training, study authors concluded that a criterion for improved productivity was satisfied. Such ratings inject a substantial subjective note. Just how comparable improved supervisor ratings are to more objective criteria of productivity improvement such as increased production in a manufacturing plant is debatable. This potential weakness in meta-analytic techniques and implications for the present study are addressed again in a later section of this chapter.

Second, Burke and Day found that among the content areas of managerial training studied, human relations training was effective in improving objective results on-the-job. Although executive (or management) training content and human relations training content were coded separately in the present study, both content areas were associated with strong positive effects--mean effect sizes of a full standard deviation for executive or management training overall and 0.88 standard deviation for human relations training content. Just as the present study has exploded "training" into components for study, the Burke and Day study exploded "management training." As with training method, the findings of both studies are consistent with respect to training content, although the previous study provided a finer level of analysis.

Results of the study at hand suggest that human relations training may lead to strong results across many worker types in addition to managers. Perhaps job analysis research would contribute to this issue by ascertaining the degrees of human relations required by different jobs. It may be that human relations are so pervasive in job skills that there is always opportunity to improve, or it may be that certain jobs require very little in the way of human relations. Workers in such jobs would not be predicted to make great improvements in their job performance after human relations training. Offering human relations training to such workers would be ineffective.

Finally, Burke and Day concluded that more research is needed on the experience of the trainer and training effectiveness. Findings from the present meta-analysis suggest that trainers external to the sponsoring organization are associated with stronger training effects than those who are internal. Consultants and vendors may find marketing fuel in those results! The one exception from this study appears to be when direct supervisors were used as trainers. This result is questionable, however, because of insufficient data. Line managers serving as trainers of workers not necessarily reporting to them, on the whole, had lackluster training effectiveness.

Many questions for further research are raised: what accounts for the difference between effectiveness of external and internal trainers; what are the observable differences between programs presented by external and internal trainers; who in the sponsoring organization sanctions, requests, or supports training programs presented by each type of trainer? It is possible that external trainers are better trainers, but not likely, assuming similar backgrounds and experience. Programs may be different, with those of vendors more polished and tried. Training requested or sanctioned at higher authority levels in the organization may make a difference, assuming external trainers may often be imported by top management. These and other possible explanations can only be validated by research.

Burke and Day did not address the issues of direct supervisory involvement in training design, delivery, and follow-up or training objective. Results of the present study suggest that training designers who involve managers in the design phase of training can expect better training results than by involving them in either follow-up activities or combined design and follow-up. Perhaps training content is more accurately adapted or customized to fit the particular work situation when the manager contributes to design. Perhaps post-training behaviors have been approved and anticipated by managers during design, leading to more tacit reinforcement or prompting for the newly learned knowledge, skill, or attitude. These conjectures need research attention.

A rather surprising conclusion of this study is that knowledge objectives in training seem to be associated with stronger productivity improvements than are either skill or attitude objectives. This conclusion provides a convenient point for drawing together several of the consistencies between the Burke and Day study and the present study.

The number of subject studies reporting on management training programs, the effectiveness of traditional, information processing methods such as lecture, and the effectiveness of knowledge objectives seem to fit a pattern. It is possible that findings were mediated by learning preferences, past experiences and other characteristics of the trainees. If most managers in the subject studies are accustomed to learning in abstract terms and base their actions to some degree on information and knowledge, they may simply have been responding to methods familiar to them. Three times as many effect sizes were found for executive or management content as for technical content; over twice as many managers were identified in the worker type variable as either professionals or laborers. The sheer volume of management training represented in the studies makes this explanation plausible.

As mentioned previously, subjective supervisor ratings of performance taken as indicators of productivity improvement may also have contributed to this finding. That is, knowledge of a management theory might influence supervisor ratings for managers. The real question is whether knowledge of the theory of ignition systems would influence production line output in an automobile factory. Further investigation of the design variables across varied worker populations is recommended.

Contextual Variables

Conclusions for the contextual variables, organization size and type and worker type, relate to the Guzzo et al. (1985) meta-analysis described above. That study and the present one share both consistencies and inconsistencies with respect to organization context most receptive to successful productivity improvement interventions.

The Guzzo team reported findings on contextual variables for all eleven interventions combined; this study reported findings for training only. Findings were similar for organization size, different for organization type, and mixed for type of worker. Results of both studies indicated that small organizations could expect most success from attempts to improve productivity, regardless of the intervention technique. The findings of the present study were contradictory to Guzzo's for large and medium-sized organizations. The conclusion drawn in the previous meta-analysis, that intervention effectiveness decreases as organization size increases, must be questioned for training interventions. Guzzo's group found governmental organizations more responsive to interventions than either private for-profit or private non-profit organizations. This conclusion, too, must be questioned for training interventions, based on the findings of the present study. These findings indicated that business and industry got a substantial return for the training investment--about three times that of governmental or non-profit organizations. Both studies agreed that managerial workers tended to produce greatest rewards from productivity improvement interventions, training or otherwise. Sales personnel ranked high in both studies. But the studies disagreed on effectiveness of interventions for laborers: training interventions appeared to be very effective for this group, contrary to Guzzo's findings for all interventions combined.

The apparent discrepancies between findings of the Guzzo study and this one may not be discrepancies at all. It is possible that training has peculiarities of effectiveness based on organization context, peculiarities not shared by other productivity improvement interventions. Research questions are thus raised for behavioral scientists in many related fields.

Study Design Variables

Conclusions for study design variables in the training meta-analysis are almost identical to those in the Guzzo meta-analysis for eleven interventions, including training. In both instances, effects were smaller when true controls or nonequivalent comparison groups were used. Interestingly, strongest effects for all interventions and for training, alone, occurred when contrasting interventions or different training approaches were being compared. Self-comparisons were associated with the next highest effects.

Differences appeared in the strength of effects measured over time. Guzzo's team reported a general weakening of effects for all interventions as time increased between beginning of intervention and measurement of effect. This finding was reversed in the training study. Less data was available as training measurement intervals increased, but only at the 18-24 month interval was there insufficient data for analysis.

If it is valid, the trend for stronger effects over prolonged periods of time could have important implications for HRD professionals projecting the pay-back or economic gain of a training program. If a program can be expected to pay dividends for up to two years, it may leap in value to the organization. This apparent trend should be investigated further through longitudinal or time-series analysis.

Limitations of the Study

Several limitations of this study should be mentioned. Some limitations are specific to the study, while others are more general and apply to meta-analysis as a research methodology.

First, some studies which meet all the selection criteria for inclusion into a meta-analysis (including this one) are not represented in the final statistical analysis because they lack sufficient statistical information to allow effect sizes to be calculated or derived: means are omitted; standard deviations are not reported; exact values of test statistics are left out (see Appendix C). Similarly, insufficient reporting of program design or study parameters mean loss to analysis of any effect sizes associated with the missing information. It is impossible, for example, to quantify effects of internal or external trainers if the author of the subject study has omitted a description of the trainer. Loss of representation from either source may affect findings and conclusions. An obvious loss of research efficiency is incurred by meta-analysts when they face such gaps in primary information.

Second, variable procedures for calculating effect sizes makes direct numerical comparisons between meta-analyses difficult. Comparisons may be reduced to trends and generalities. This limitation was encountered when comparing findings of the Burke and Day study, the Guzzo meta-analysis, and the present study. General findings could be discussed rationally, but differences in mean effect sizes were generally presumed to be computation artifacts.

Third, this meta-analysis carries the threat of a selection bias. Nearly all effect sizes were positive, indicating superior performance of experimental groups over controls. Rarely were anything but success stories published. This potential bias casts doubts about generalizability of findings to field experiments and actual HRD interventions. A useful rule of thumb is that findings are generalizable only in so far that the 48 subject studies are representative of all actual programs. Industry reports indicate that the proportion of management training in the studies is not unusual. One would, however, expect to see more sales and clerical training reports than actually appeared.

Fourth, multiple findings from single studies were treated as individual values. Consequently, a single study could easily be disproportionately represented in the effect sizes data. In larger meta-analyses, the proportions of multiple findings decrease; in relatively small studies such as this one, they may be substantial.

Fifth, the number of variables and categories of variables used in this study sometimes fractured the data into unreliably small cell sizes. This condition cast doubt on some findings and precluded certain questions from being answered. Analysis of variance could not be performed among variables or the categories of a variable, for example, to test for significant differences because of radically unequal cell sizes. Data were insufficient to allow examination of distributions of a variable or a category of a variable over, for instance, the three productivity criteria. No interactions of criteria, design variables, and context could reliably be studied. Lack of data may be a plausible explanation for some findings, notably the indication that strength of training increased with time. This finding is certainly counter-intuitive for a number of reasons. It is well-established that recall diminishes with time. People forget about things and they forget how to do things if practice or update schedules are not in place. Certain knowledge or skills may cease to be important as jobs evolve and demands change. Given the weight of evidence to the contrary, the likelihood that training effects actually become stronger as time goes by is not great. The finding is probably an artifact of data distribution.

Sixth, a major weakness of meta-analysis and also of this study centers on coding the studies. Glass et al. (1981) noted that there is no cookbook approach to selecting substantive variables for coding in meta-analysis. In the present study, care was taken to extract from training literature operational definitions that would in their specificity of language guide coding decisions. (The reader is referred to Chapter III.) For instance, programs were coded as safety training if they were described as "designed for either operative or supervisory personnel, and correct safety procedures

[were] demonstrated or discussed" (Tracey, 1971, p. 27). Not all subclassifications of all variables were so clear-cut, however.

To discriminate between knowledge and skills objectives, for example, required locating the description provided by the study author and making a determination based on the language appearing there. When the author used words such as "to know" or "to examine a theory of ...," the program was coded for knowledge objective. On the other hand, such words as "to do" or "to perform" signalled skills objectives. During the coding process, the meta-analyst is working twice removed from the actual training program in question. One interpretation was made by the study author as he or she selected the words to describe what actually took place. The meta-analyst must interpret again to categorize attributes for further analysis. Error is almost surely introduced by this process. Controlling the error may be a matter of inter-coder reliability components of meta-analyses, clear and explicit operational definitions of variables, and adherence to decision rules implied by the definitions. A complete list of coding for the subject studies is included as Appendix E.

Finally, this study may suffer from too broad a definition of productivity. Practitioners in the field could be misled if they took findings out of context of this broad definition. As noted in Chapter I, debate over single and multiple indices of productivity continues. The problems of research in productivity hinge on lack of commonly accepted definitions and measurements of productivity (Guzzo & Bondy, 1983).

An illustration drawn from the subject studies is in order. Consider the following examples of actual productivity indices taken from the studies: quality of patient care in a nursing home measured by supervisor rating, resignation rates for the subordinates of trained versus untrained supervisors, production rates, length of stay for patients in a mental institution whose staff was trained, rate of errors in tax forms prepared by trained versus untrained IRS agents, earnings increases over time for managers in development programs, percent of observed acts performed safely in a manufacturing environment. The variety of methods of measurement and the various criteria involved in these studies have been lumped into a "productivity" construct which could be challenged as unusual, if not unique. Measurement in the studies is complicated by at least two typical but different approaches: sometimes those being trained are subsequently evaluated; other times the behavior of subordinates of trainees is examined and taken to be the indicator of "productivity." (Perhaps this mixed approach to measurement could be clarified by the ecology model of training effectiveness offered earlier in this paper.)

In a word, the given definition of productivity which underlies this meta-analysis may be too inclusive. A more commonly-accepted description of the goals of the training programs might simply be "performance improvement." Because the investigators who originally identified the subject studies (and the other studies that made up their investigations) chose to define "productivity" so broadly, the present study suffers from possible misconceptions about the application of findings to more narrow productivity improvement efforts.

Implications and Recommendations

The stated purpose of this study was to identify relationships among training variables, organization contingencies, and multiple productivity measures and to suggest ways in which the relationships discovered can guide effective training design and useful program evaluation. On a methodological level, the study was designed to test the efficacy of meta-analysis as a means of collating training results from multiple studies and ordering them in a comprehensible whole. The following implications and recommendations derive from the study findings and the purposes for the investigation.

Training has been shown to be a potent intervention for improving worker productivity across multiple organization contexts, worker types, and productivity criteria. Training professionals can expect that their programs can contribute to the goals of the sponsoring organization as well as to the recipients of the training. Conversely, sponsoring organizations not reaping gains from current training programs aimed at improving worker productivity are justified in asking why not.

Certain training parameters are associated with strong training effects. Although many questions remain about peculiarities of these parameters in various contexts and with various employees, there appears to be evidence that knowledge objectives and information processing methods are important components of successful training interventions for productivity improvement.
More research into the interplay of design and context variables, along with high-quality publication of future field experiments, can contribute to knowledge of the true relationships among these factors. With knowledge of relationships can come better design, delivery, and follow-up.

Meta-analysis is useful as a method for integrating HRD findings. Thoughtfully applied, it can facilitate inductive thinking about training issues. It can, therefore, be helpful in theory development which will, in turn, spawn further deductive investigations. It allows disparate findings to accumulate along a continuum until similarities heap, anomalies separate, and a pattern emerges.

Meta-analysis may also be useful as a yardstick for estimating return on a training investment. In the words of an HRD practitioner, "You take that figure that already exists at the beginning--at the starting point for estimating what the payoffs might be to the organization" (Bove, 1986, p. 36). For this sort of application of meta-analytic findings to become commonplace, many more such studies will be required.

Quality of study reports in the training literature is essential but currently inadequate for meta-analysis to be most effective. Detailed statistical information should be presented, including means and standard deviations of experimental and control groups, actual test statistic values (whether significant or not), number of subjects, and complete training program information. A taxonomy for training program description should be developed for use as a checklist for those preparing manuscripts for publication.

Quality of reporting is but a portion of the quality of research needed in the training field. Certainly quality of design and interpretation are necessary, too.

This study raised many potential research questions: testing effectiveness of external and internal trainers, job analysis approaches to testing effectiveness of human relations training, and longitudinal tests of the enduring effects of training on productivity improvements, to name a few. Continued investigation is essential to fuel improvements in practice.

Students of productivity must construct a commonly-accepted definition of the term. Otherwise potential confusion such as that surrounding the findings of this study will reign. A more narrow definition than the one underlying the present study is likely to emerge.

Finally, future productivity studies should focus on improved measurement techniques. Objective measures differ substantially from subjective measures, but both are presently mixed as measures of productivity. Measures of trainee behavior or improved performance and measures of subordinates' behavior or performance should be distinguished. The model of training effectiveness developed and presented in Chapter II of this study provides a framework for a measurement typology based on constituent goals.

APPENDICES

Appendix A

Application of Cooper's Guidelines

Application of Cooper's Guidelines

STAGE	RESPONSE
1. Problem Formulation	a. "Training and instruction" defined as "an organizational practice whereby employees are educated with regard to specific skills deemed central to their jobs."
	b. "Productivity" defined as one of three outcome measures: output, withdrawal, disruption (a condensation of the 14 outcome measures in the original studies; used by Guzzo and Bondy, 1983, and by Guzzo, Jette, and Katzell, 1985.)
	c. Study details to be examined: time lapse between training end and criterion measurement; use of comparison groups.
	d. Aspects of training to be examined: training method, training objective, training type, delivery system, trainer, management involvement.
	e. Organizational context variables to be examined: organization size, organization type, worker type.
2. Data Collection	This review uses studies selected by previous researchers. Sources will be reported.
3. Evaluation of Data	Studies will not be

		weighted. All findings will be reported for analysis. Missing data will be noted.
4.	Analysis; Data Interpretation	Meta-analysis procedures will be reported as used.
5.	Public Presentation	Separate analyses on variables of interest will be performed. Procedural detail will be reported to ensure replicability.

Appendix B

Bibliography of Subject Studies

Bibliography of Subject Studies

Alpert, J.L. (1975). Teacher behavior and pupil performance: Reconsideration of the mediation of Pygmalion effects. Journal of Educational Research, 69(2), 53-57.

Aronoff, J., & Litwin, G.H. (1971). Achievement motivation training and executive advancement. Journal of Applied Behavioral Science, 7, 215-229.

Beatty, R.W. (1975). A two-year study of hard-core-unemployed clerical workers: Effects of scholastic achievement, clerical skill, and self-esteem on job success. Personnel Psychology, 28, 165-174.

Brown, M.G., Malott, R.W., Dillon, M.J., & Keeps, E.J. (1980). Improving customer service in a large department store through the use of training and feedback. Journal of Organizational Behavior Management, 2, 251-265.

Byham, W.C., Adams, D., & Kiggins, A. (1976). Transfer of modeling training to the job. Personnel Psychology, 29, 345-349.

Cook, D.W., Kunce, J.T., & Sleater, S.M. (1974). Vicarious behavior induction and training psychiatric aides. Journal of Community Psychology, 2, 293-297.

Copenhaver, J.R. (1973). Training, job enrichment reduce costs. Hospitals, 47(3), 118, 122, 126.

Crawford, L.S., Thomas, E.D., & Fink, J.J. (1980). Pygmalion at sea: Improving the work effectiveness of low performers. Journal of Applied Behavioral Science, 16, 482-505.

Cullen, J.G., Sawzin, S., Sisson, G.R., & Swanson, R.A. (1976). Training: What's it worth? Training and Development Journal, 30, 12-20.

Driscoll, J.M., Meyer, R.G., & Schanie, C.F. (1973). Training police in family crisis intervention. Journal of Applied Behavioral Science, 9, 62-82.

Dyer, E.D., Monson, M.A., & Cope, M.J. (1975).

Increasing the quality of patient care through performance counseling and written goal setting. Nursing Research, 24, 138-144.

Ely, D.D., & Morse, J.T. (1974). TA and reinforcement theory. Personnel, 51(2), 38-41.

Fiedler, F.E., & Mahar, L. (1979). A field experiment validating contingency model leadership training. Journal of Applied Psychology, 64, 247-254.

Goldstein, A.P., & Sorcher, M. (1974). Changing supervisor behavior (pp. 70-73). New York: Pergamon Press.

Hahne, C.E. (1977). How to measure results of sales training. Training and Development Journal, 31, 3-7.

Hand, H.H., & Slocum, J.W., Jr. (1972). A longitudinal study of theeffects of a human relations training program on managerial effectiveness. Journal of Applied Psychology, 56, 412-417.

Hand, H., Richards, M., & Slocum, R. (1973). Organizational climate and the effectiveness of a human relations training program. Academy of Management Journal, 16, 185-195.

Holder, J., Jr., (1971). Evaluation of an in-company management training program. Training and Development Journal, 26(4), 24-27.

Horrigan, J.T. (1979). The effects of training on turnover: A cost justification model. Training and Development Journal, 33, 3-7.

Hultman, K.E., & Cunningham, G. (1978). Preparing employees for upward mobility. Training and Development Journal, 32, 10-14.

Ivancevich, J.M., (1976). Effects of goal setting on performance and job satisfaction. Journal of Applied Psychology, 61, 605-612.

Ivancevich, J.M., & McMahon, J.T. (1976). Group development, trainer style, and carry-over job satisfaction and performance. Academy of Management Journal, 19, 395-412.

Karasu, T.B., Stein, S.P., & Charles, E.S. (1974). A preliminary study of the elimination of the internship: A comparative study of performance of internship- and non-internship-trained residents in a psychiatric inpatient setting. Archives of General Psychiatry, 31, 269-272.

Komaki, J., Barwick, K.D., & Scott, L.R., (1978). A behavioral approach to occupational safety: Pinpointing and reinforcing self-performance in a food manufacturing plant. Journal of Applied Psychology, 63, 434-445.

Komaki, J., Heinzman, A.T., & Lawson, L. (1980). Effects of training and feedback: Component analysis of a behavioral safety program. Journal of Applied Psychology, 65, 261-270.

Krackhardt, D., McKenna, J., Porter, L.W., & Steers. R.M. (1981). Supervisory behavior and employee turnover: A field experiment. Academy of Management Journal, 24, 249-259.

Latham, G.P., & Saari, L.M. (1979). Application of social learning theory to training supervisors through behavioral modeling. Journal of Applied Psychology, 64, 239-246.

Luke, R., Block, P., Davey, J., & Averch, V. (1973). A structural approach to organizational change. Journal of Applied Behavioral Science, 9, 611-635.

McCormick, J.H. (1971). An old standby that still works. Training and Development Journal, 26(10), 3-7.

Mehr, J. (1971). Evaluating nontraditional training for psychiatric aides. Hospital and Community Psychiatry, 22, 315-318.

Mikesell, J.L., Wilson, J.A., & Lawther, W. (1975). Training program and evaluation model. Public Personnel Management, 4, 405-411.

Miron, D., & McClelland, D.C. (1979). The impact of achievement motivation training on small businesses. California Management Review, 21, 13-28.

Moses, J.L., & Ritchie, R.J. (1976). Supervisory relationships training: A behavioral evaluation of a behavior modeling program. Personnel Psychology, 29, 337-343.

Nemeroff, W.F., & Cosentine, J. (1979). Utilizing feedback and goal setting to increase performance appraisal interview skills of managers. Academy of Management Journal, 22, 566-576.

Palmer, C.F. (1977). The St. Louis United Way--a training success story. Training and Development Journal, 31, 46-49.

Paul, G.L., McInnis, T.L., & Mariotto, M.J. (1973). Objective performance outcomes associated with two approaches to training mental health technicians in milieu and social learning programs. Journal of Abnormal Psychology, 82, 523-532.

Porras, J.R., & Anderson, B. (1981). Improving managerial effectiveness through modeling-based training. Organizational Dynamics, 9, 60-77.

Quilitch, H.R. (1975). A comparison of three staff-management procedures. Journal of Applied Behavior Analysis, 8, 59-66.

Roberts, M. (1971). A cost-benefit report on training disadvantaged youths for apprenticeship. Training and Development Journal, 26(6), 32-35.

Rosen, R.A., & Rosen, H. (1972). Orientation and job stability: The impact of supervisory and employee role-playing groups on turnover of potentially unemployed blacks. Journal of Vocational Behavior, 2, 25-37.

Rosen, H., & Turner, J. 1971). Effectiveness of two orientation approaches in hard core unemployed turnover and absenteeism. Journal of Applied Psychology, 55, 296-301.

Sard, P., & Berni, R. (1974). An incentive contract for nursing home aides. American Journal of Nursing, 74, 475-477.

Seltzer, A., Moskowitz, A.S., Lucas, T., & Moskowitz, J.A. (1975). Welfare recipients as paraprofessionals: Utilization in facility for retarded patients. New York State Journal of Medicine, 75, 751-755.

Sisson, G.R. (1971). Development of training for a new manufacturing process. Training and Development Journal, 26(12), 22-31.

Skuja, A.T., Schneidmuhl, A.M., & Mandell, W. (1975). Alcoholism counselor trainees: Some changes in job related functioning following training. Journal of Drug Education, 5, 151-157.

Smith, P.E. (1976). Management modeling training to improve morale and customer satisfaction. Personnel Psychology, 29, 351-359.

Stude, E.W. (1973). Evaluation of short-term training for rehabilitation counselors: Effectiveness of an institute on epilepsy. Rehabilitation Counseling Bulletin, 16, 146-154.

Wexley, K.N., & Nemeroff, W.F. (1975). Effectiveness of positive reinforcement and goal setting as methods of management development. Journal of Applied Psychology, 60, 446-450.

Appendix C

Glass's Procedures for Calculating Effect Sizes

Glass's Procedures For Calculating
Effect Sizes

Reported by Investigator	Calculation Formula	Implications
Means and standard deviations for experimental and control groups	$\Delta = \frac{\bar{X}_E - \bar{X}_C}{S_C}$ (Where Δ = effect size $\bar{X}_E$ = mean, experimental group $\bar{X}_C$ = mean, control group S_C = SD, control group)	Straightforward calculation; some statisticians argue for corrections or alternatives; needed data not always reported
t-value (or F-value)	$\hat{\Delta} = t\sqrt{1/n_E + 1/n_C}$ (where t and F are test statistics, $F = t^2$ n_E = experimental group size n_C = control group size)	Must assume homogeneous variances; estimator is positively biased for $n < 20$; biased < 10% for $n \geq 20$ (Hedges, 1979, cited in Glass et al., 1981); test statistics not always reported when results are non-significant
Significant findings at an alpha-level	$\hat{\Delta} = Z\sqrt{1/n_E + 1/n_C}$ (where Z = the normal deviate corresponding to the level of significance reported: n_E and n_C as above)	A rough approximation of actual findings; necessarily excludes any non-significant results

Note: Glass also provides procedures for deriving effect sizes from other reported statistics and conditions. These were not required in the present analysis.

Appendix D

Five Meta-Analysis Methods Drawn From Bangert-Drowns

Five Meta-Analysis Methods
Drawn from Bangert-Drowns

Method (proponent)	Purpose	Consideration of Study Features	Representation of Treatment Effect	Unit of Analysis
1. Scientific Criticism (Glass)	(1) Evaluate treatment and effect (2) Study method of research and taxonomy employed by workers in field	(1) Very important (2) No pre-selection on *a priori* quality criteria (3) Multiple representation	Effect size = $\frac{\bar{Y}_E - \bar{Y}_C}{S_C}$	Study
2. Literature Review (Kulik)	Describe most accurate representation of studies' findings	(1) Important (2) Selective - uses only quality studies (3) No multiple representation	(1) Effect size (2) Report separate findings if dependent measures gauge different constructs	Study
3. Combined Probability Method (Rosenthal)	(1) Estimate a reliable treatment effect amd provide some measure of its reliability (2) Focus on wide generalizability	(1) Not important (2) Leaves selection up to reviewer	$Z_{cum} = \frac{\Sigma z}{\sqrt{n}}$, others	Study or Subject depending on statistics used
4. Analysis using tests of homogeneity (Hedges)	Correct Glass and estimate "true" population effects	(1) Diversity accepted (2) Exclude multiple representations	G^u (corrects bias when $N \leq 10$)	Subject
5. Analysis using estimates of population variance (Hunter & Schmidt)	Correct study outcomes for research artifacts to produce best estimates of population treatment effects	Includes all, regardless of methodological quality	(1) Effect size *d* is corrected by reliabilities (2) Measures variance and sampling error	Subject

Appendix E

Coding Sheet

AUTHOR (DATE)	CRITERION	TR METHOD	OBJECTIVE	TRAINER ORG	SIZE ORG	TYPE	WORKER	TRNG TYPE	MGT INVLV.	# WEEKS	WEEKS	COMP GRP
Alpert, 1975	O	IP	S	E	U	O	PROF	TECH	U	11.00	A	NC
Aronoff & Litwin, 1971	O	IP	K	E	L	P	M	EX	U	104.00	D	OT
Beatty, 1975	O	IP,OTJ	K,S,A	U	M	P	C	T&S	U	104.00	D	NC
Cook, Kunce & Sleater, 1974	O	S	S	E	U	G	L	HR	U	6.00	A	OT
Copenhaver, 1973	O	IP,S	S	I-TSP	L	O	M	MM	F	U		NONE
Driscoll, Meyer & Schanie, 1973	O	IP,S	S	U	U	G	PROF	HR	U	26.00	A	NC
Dyer, Monson, & Cope, 1975	O	S,OTJ	S,A	I-LM	U	G	PROF	U	F	26.00	A	NC
										52.00	B	
Ely & Morse, 1974	W	IP,S	K	U	U	P	M	HR	U	52.00	B	SC
Goldstein & Sorcher, 1974	O	S	S	E	U	P	M	MM	F	10.00	A	NC
Hand & Slocum, 1972	O	S	A,S	E	U	P	M	HR	U	78.00	C	TC
Hand, Richards & Slocum, 1973	O	S	A,S	E	U	P	M	HR	F	78.00	C	TC
Holder, 1972	O	IP	S	I-TSP	L	P	M	EX	U	26.00	A	NONE
Karasu, Stein & Charles, 1974	O	OTJ	K,S	I-LM	U	O	PROF	TECH	U	U		NC
Luke, Block, Davey, Averch, 1973	O	OTJ	A	I-LM	L	P	M	EX	D,T	26.00	A	OT
McCormick, 1971	O	IP	S	I-TSP	U	P	M	MM	U	U		U
Mehr, 1971	O	IP	S	I-LM, I-T	U	G	L	TECH	D	0.00	A	NC
Mikesell, Wilson & Lawther, 1975	O	S	S	E	L	G	L	TECH	D	U		NC
Paul, McInnis, & Mariotto, 1973	O	OTJ	S	I-O	U	G	L	TECH	T	18	A	OT
Quilitch, 1975	O	IP	K	I-LM	U	G	PROF	TECH	U	0.00	A	SC
Roberts, 1972	O	S	S	E	U	O	L	T&S	U	52.00	B	NC
Rosen & Rosen, 1972	W	S	A	E	L	G	L	O	D	52.00	B	TC
Rosen & Turner, 1971	W	IP	A	I-O	L	G	L	O	U	26.00	A	NC
Sand & Berni, 1974	O	IP	S	U	U	P	L	TECH	U	U		NC
Seltzer, Moskowitz,												
Lucas, Moskowitz, 1975	W	IP	K	E	U	G	L	O	U	U		NC
Sisson, 1972	O	S	S	I-LM	L	P	L	TECH	D,T,F	U		NC
Skuja, Schneidmuhl, Mandell, 1975	O	IP,OTJ	K,S,A	E	U	P	PROF	TECH	U	7.00	A	SC
										14.00	A	
Stude, 1973	O	IP	K,A	E	L	G	PROF	TECH	U	20.00	A	TC
Wexley & Nemeroff, 1975	W	S	S	E	M	O	M	MM	F	8.00	A	TC
	O	S	S	E	M	O	M	MM	F		A	TC
Brown, Malott, Dillon, Keeps, 1980	O	OTJ	S	I-TSP	M	P	S	S	F	6.00	A	SC
Byham, Adams, Kiggins, 1976	O	S	S	I-TSP	U	U	M	MM	D,F	30.00	B	NC
	W	S	S	I-TSP			M	MM	D,F		B	NC
	D	S	S	I-TSP			M	MM	D,F		B	NC
Crawford, Thomas, & Fink, 1980	O	IP	K,S,A	U	M	G	M	MM	F	34.00	B	SC
	D	IP	K,S,A		M	G	L	HR	F		B	SC
Cullen, Sawzin, Sisson, Swanson, 1976	O	OTJ	S	E	L	P	L	T&S	U	U		OT
Fiedler & Mahar, 1979	O	IP	S	U	U	U	Other	LDSHP	U	0.00	A	TC
Hahne, 1977	O	S	S	I-TSP	L	P	S	SLS	F	8.00	A	SC
Horrigan, 1979	W	U	K,S	U	M	P	L,M,C	TECH	U	52.00	B	NC
Hultman & Cunningham, 1978	O	S	S	I-O	M	G	L	HR	U	0.00	A	SC
Ivancevich & McMahon, 1976	O	IP	K	U	L	P	M	MM	U	52.00	B	OT
Komaki, Heinzmann, Lawson, 1980	D	IP	K	U	M	G	L	S	D,F	45.00	B	SC
Krakhardt, McKenna, Porter, Steers1981	W	IP	S	U	L	P	M	MM	F	16.00	A	NC
Latham & Saari, 1979	O	S	S	U	L	P	M	HR	F	52.00	B	TC
Miron & McClelland, 1979	O	IP	K	E	S	P	M	MM	U	78.00	C	SC
Moses & Ritchie, 1976	O	S	S	U	L	P	M	HR	U	8.00	A	NC
Nemeroff & Cosentino, 1979	O	U	S	E	L	P	M	MM	U	U		TC
	W	U	S	E	L	P	M	MM				TC
Palmer, 1977	O	IP	K	I-LM,E	M	O	OTHER	O	U			NONE
Porras & Anderson, 1981	O	S	S	I-LM	M	P	M	MM	D	24.00	A	SC
	W	S	S	I-LM	M	P	M	MM	D	24.00	A	SC
Smith. 1976	O	S	S	U	L	P	M	MM	U	8.00	A	SC
Ivancevich, 1976	O	IP	U	U	M	P	S	SLS	U	U		SC
Komaki, Berwick & Scott, 1978	D	IP	K	U	U	P	L	S	F	CONTINUOUS		SC

BIBLIOGRAPHY

Adam, E.E., Herschauer, J.C., & Ruch, W.A. (1981). Productivity and quality: Measurement as a basis for improvement. Englewood Cliffs, NJ: Prentice-Hall.

Andrews, F.M. (1979). The international study: Its data sources and measurement procedures. In F.M. Andrews (Ed.), Scientific productivity: The effectiveness of research groups in six countries. Cambridge: Cambridge University Press.

Bangert-Drowns, R.L. (1984). Developments in meta-analysis: A review of five methods. Ann Arbor, MI: University of Michigan, Center for Research on Learning and Teaching. (ERIC Document Reproduction Service No. ED 248 262)

Bass, B. (1982). Individual capability, team performance, and team productivity. In M.D. Dunnette & E.A. Fleishman (Eds.), Human performance and productivity: Vol. 1. Human capability assessment (pp. 179-232). Hillsdale, NJ: Erlbaum.

Bloom, B.S. (Ed.). (1956). The taxonomy of educational objectives: Handbook I. Cognitive domain. New York: McKay.

Bove, R. (1986). New-wave HRD accounting. Training and Development Journal, 40(12), 36-37.

Brand, H., & Belitsky, H.A. (Comp.). (1980). Productivity: A selected annotated bibliography. Washington, DC: Department of Labor, Bureau of Labor Statistics. (ERIC Document Reproduction Service No. ED 193 492)

Brinkerhoff, R.O. (1981). Making training evaluation more useful. Training and Development Journal, 35 (12), 35-39.

Burke, M.J., & Day, R.R. (in press). A cumulative study of the effectiveness of managerial training. Journal of Applied Psychology.

Burke, W.W. (1978). The role of training in organization development. In M. Knowles, The adult learner: A neglected species (3rd ed., pp. 169-175). Houston, TX: Gulf. (Reprinted from Training and Development Journal, 26(9), 30-34)

Carroll, S.J., Jr., Paine, F.T., & Ivancevich, J.J. (1972). The relative effectiveness of training methods--expert opinion and research. Personnel Psychology, 25, 495-509.

Cooper, H.M. (1982). Scientific guidelines for integrating research reviews. Review of Educational Research, 52(2), 291-302.

Davies, I.K. (1981). Instructional technique. New York: McGraw-Hill.

Delamontagne, R.P. (1984). Personnel's role in productivity improvement. Performance & Instruction Journal, 23(3), 8-9.

Donaldson, L., & Scannell, E. (1978). Human Resource Development. Reading, MA: Addison-Wesley.

Drucker, P.F. (1964). Managing for results. New York: Harper & Row.

Dunn, S., & Thomas, K. (1985). Surpassing the "smile sheet" approach to evaluation. Training, 22(4), 65-67, 71.

Eurich, N.P. (1985). Corporate classrooms: The learning business. Princeton, NJ: The Carnegie Foundation for the Advancement of Teaching.

Feuer, D. (1985). Where the dollars go. Training, 22 (10), 45-46, 48, 50-51, 53.

Friedman, P.G., & Yarbrough, E.A. (1985). Training strategies from start to finish. Englewood Cliffs, NJ: Prentice-Hall.

Gallucci, S. (1983). Training for productivity. Training, 20(8), 49-52.

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

Goldstein, I.L., & Buxton, V.M. (1982). Training and human performance. In M.D. Dunnette & E.A. Fleishman (Eds.), Human performance and productivity: Vol. 1.

Human capability assessment (pp. 135-177). Hillsdale, NJ: Erlbaum.

Guzzo, R.A., & Bondy, J.S. (1983). A guide to worker productivity experiments in the United States, 1976-81. New York: Pergamon.

Guzzo, R.A., Jette, R.D., & Katzell, R.A. (1985). The effects of psychologically based intervention programs on worker productivity: A meta-analysis. Personnel Psychology, 38(2), 275-291.

Hersey, P., & Blanchard, K.H. (1982). Management of organizational behavior: Utilizing human resources (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Holoviak, S.J. (1982). The impact of training on productivity levels. Performance & Instruction Journal, 21(5), 6-8.

Juran, J.M. (1974). Quality policies and objectives. In J.M. Juran, F.M. Gryna, Jr., & R.S. Bingham, Jr. (Eds.). Quality Control Handbook (3rd ed.), (pp. 3-1 - 3-12). New York: McGraw-Hill.

Katzell, R.A., Bienstock, P., & Faerstein, P.H. (1977). A guide to worker productivity experiments in the United States, 1971-75. New York: New York Press.

Kendrick, J.W. (1984). Improving company productivity: Handbook with case studies. Baltimore, MD: Johns Hopkins University Press.

Kulik, J.A. (1984, April). Uses and misuses of meta-analysis. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED 248 618)

Laird, D. (1978). Approaches to training and development. Reading, MA: Addison-Wesley.

Landman, J.T., & Dawes, R.M. (1982). Psychotherapy outcome: Smith and Glass' conclusions stand up under scrutiny. American Psychologist, 37, 504-516.

Medoff, J.L. (1982). Formal training and labor productivity. Washington, DC: National Institute of Education. (ERIC Document Reproduction Service No. ED 227 360)

Michalak, D.F. (1981). The neglected half of training. Training and Development Journal, 35, 22-28.

Miles, R.H. (1980). Organizational effectiveness. In R.H. Miles, Macro organizational behavior (pp. 353-385). Santa Monica, CA: Goodyear.

Muckler, F.A. (1982). Evaluating productivity. In M.D. Dunnette & E.A. Fleishman (Eds.), Human performance and productivity: Vol. 1. Human capability assessment (pp. 135-177). Hillsdale, NJ: Erlbaum.

Nadler, L. (Ed.). (1984). The handbook of human resource development. New York: Wiley.

Newstrom, J.W. (1975). Selecting training methodologies, a contingency approach. Training and Development Journal, 23(10), 12-16.

Newstrom, J.W. (1980). Evaluating the effectiveness of training methods. Personnel Administrator, 25(1), 55-60.

People and productivity: Excerpts from a study by the New York Stock Exchange. (1983). Productivity Brief 21. Houston, TX: American Productivity Center.

Phillips, J.J. (1983). Handbook of training evaluation and measurement methods. Houston, TX: Gulf.

Pritchard, R.D., Montagno, R.V., & Moore, J.R. (1978). Enhancing productivity through feedback and job design. (AFHRL-TR-78-44). Brooks Air Force Base, TX: Air Force Systems Command, Air Forces Human Resources Laboratory.

Robinson, D.G. (1984). Training for impact (how to stop spinning your wheels and get into the race). Training, 21(2), 42-47.

Russell, J.S., Terborg, J.R., & Powers, M.L. (1985). Organizational performance and organizational level training and support. Personnel Psychology, 38, 849-863.

Schonberger, R.J. (1986). World class manufacturing: The lessons of simplicity applied. New York: Free Press.

Spencer, L.M., Jr. (1984). How to calculate the costs and benefits of an HRD program. Training, 21(7), 40-44, 48-51.

Stoner, J.A.F. (1978). Management. Englewood Cliffs, NJ: Prentice-Hall.

Sutermeister, R.A. (1976). People and productivity (3rd ed.). New York: McGraw-Hill.

Tracey, W.R. (1971). Designing training and development systems. New York: American Management Association.

Trainers rank tools of their trade. (1983). Training, 20(3), 96.